JMIR Research Protocols

Contents

Protocols

Proposal

Protocol

Web-Based Skin Cancer Prevention Training for Massage Therapists: Protocol for the Massage Therapists Skin Health Awareness, Referral, and Education Study

Lois J Loescher[1,2*], FAAN, PhD; Kelly M Heslin[1*], MPH, CHES; Laura A Szalacha[3,4*], EdD; Graciela E Silva[1*], MPH, PhD; Myra L Muramoto[2,5*], MPH, MD

[1]College of Nursing, University of Arizona, Tucson, AZ, United States

[2]College of Public Health, University of Arizona, Tucson, AZ, United States

[3]College of Medicine, University of South Florida, Tampa, FL, United States

[4]College of Nursing, University of South Florida, Tampa, FL, United States

[5]College of Medicine, University of Arizona, Tucson, AZ, United States

[*]all authors contributed equally

Corresponding Author:
Lois J Loescher, FAAN, PhD
College of Nursing
University of Arizona
1305 N Martin Ave
Tucson, AZ,
United States
Phone: 1 (520) 626 6196
Email: loescher@email.arizona.edu

Abstract

Background: Skin cancer, the most common cancer in the United States, is costly and potentially deadly. Its burden can be reduced by early detection and prevention activities. The scope of skin cancer requires going beyond traditional health care providers to promote risk reduction. Partnering with the nonbiomedical workforce, such as massage therapists (MTs), may reach more individuals at risk. MTs see much of their clients' skin and are amenable to performing skin cancer risk reduction activities during massage appointments.

Objective: The objective of this study is to describe the Massage Therapists Skin Health Awareness, Referral, and Education protocol, presenting an overview of our systematic approach to developing rigorous e-training for MTs to enable them to be partners in skin cancer risk reduction. We also describe procedures for usability and feasibility testing of the training.

Methods: We developed an integrated electronic learning system that includes electronic training (e-training) technology, simulated client interactions, online data collection instruments, and in-person assessment of MTs' application of their training.

Results: A total of 20 participants nationally scored the e-training as high for usability and satisfaction. We have screened an additional 77 MTs in Arizona for interest and eligibility, and currently have 37 enrolled participants, of whom 32 have completed the Web-based training.

Conclusions: The structured and rigorous development approach for this skin cancer risk reduction and brief behavioral intervention e-training for MTs begins to fill a gap in skin cancer risk reduction research. Iterative usability testing of our asynchronous Web-based training resulted in positive participant response. Our e-training approach offers greater learner accessibility, increased convenience, and greater scalability than the few existing programs and has the potential to reach many MTs nationally.

International Registered Report Identifier (IRRID):   DERR1-10.2196/13480

XSL·FO
RenderX

Introduction

Background

Skin cancer, the most common cancer in the United States, constitutes a serious public health burden [1-4]. Skin cancer may be deadly or disfiguring. The most serious skin cancer, melanoma, resulted in an estimated 9000 deaths in 2018 [4]. Skin cancer treatment costs approximated US $8.1 billion in 2011 [1-3]. Most skin cancers can be prevented by simple behaviors to protect the skin from ultraviolet radiation (UVR), such as staying in shade, wearing sun protective clothing, applying sunscreen, and avoiding indoor tanning [4,5]. Early detection of skin cancer greatly decreases its potential morbidity, mortality, and cost [6-8]. The probability of early skin cancer detection increases with full body visual skin assessment (VSA) [9]. Despite these effective prevention and early detection strategies, over 5 million skin cancer cases are diagnosed or treated annually [10]. Thus, decreasing the burden of skin cancer depends on concerted and innovative public health efforts that extend beyond the conventional biomedical practitioners to complementary and integrative health care practitioners. These efforts also involve other community-wide sectors and could incorporate electronic learning (e-learning) technology to allow widespread and easy dissemination of knowledge.

In 2014, the Surgeon General issued a *Call to Action to Prevent Skin Cancer*, endorsing comprehensive community-wide efforts to prevent skin cancer by diverse partners and sectors, including business, health care, and education [5]. Massage therapists (MTs) are community members typically practicing outside of conventional health care settings, yet are professionals involved in promoting health and wellness. Despite their interface with health and wellness, MTs have been overlooked as a community-based resource to (1) help promote skin cancer risk reduction and (2) reinforce consumer-targeted public heath skin cancer awareness messages.

MTs are uniquely positioned to promote skin cancer risk reduction through *eyes on the skin* observation and client-centered communication. During a typical full body massage, the client is unclothed under a drape. MTs systematically undrape and view each anatomical area, allowing the opportunity to visualize skin cancer risk factors such as sunburn, tanning lines, high mole counts, or suspicious lesions. Clients typically see their MTs more often and for longer durations than their primary care provider and are more likely to discuss health promotion [11-13], thereby providing greater opportunities for successful client-centered communication and encouragement of effective skin cancer risk reduction behaviors such as reducing UVR exposure [14].

In our prior work with MTs, we conducted in-person and Web-based tobacco cessation brief behavioral intervention (BBI) training for nonbiomedical health care practitioners (including MTs) in private practice contexts [15,16]. This electronic training (e-training) significantly increased practitioners' use of client-centered communication, BBI, and referral skills in the form of offering clients a *helping conversation*. The helping conversation is a BBI that emphasizes active listening skills and motivational communication strategies to encourage and support clients' healthy behavior change [15].

Skin cancer education and training for MTs has been inconsistent and not rigorously evaluated. Although many MTs receive some skin cancer education, the format, content, duration, source, and depth of this education varies [17]. The few skin cancer–focused in-person workshops and 1 Web-based course available to MTs [18] have not been systematically evaluated and, to our knowledge, do not include training for VSA, client risk assessment, client-centered communication, BBI, and referral skills [17].

Objective

There is a need to develop more comprehensive, accessible skin cancer risk reduction training for MTs. Here, we describe the development of the Massage Therapists Skin Health Awareness, Referral, and Education (MTsSHARE) protocol, including the development of e-training technology, simulated client interactions, online data collection instruments, and in-person assessment of MTs' application of their training. We will describe procedures for usability and feasibility testing of the training.

Methods

Phase 1 (Complete): Adapting Existing Programs and Development of Training and Assessment Technology

Conceptual Framework

Social cognitive theory (SCT) guided the overall study. Individuals learn and maintain new behaviors in a social context through reciprocal interaction of person, environment, and behavior. In total, 4 SCT constructs guided the overall training: (1) reciprocal determinism, or the dynamic and reciprocal interaction of MTs, their external social context, and behavioral responses to the training; (2) behavioral capability to have a helping conversation; (3) observational learning from e-training vignettes; and (4) self-efficacy, affecting behavior choices, efforts to overcome barriers to behaviors, and mastery of the behaviors [19]. According to SCT, observations of a behavior, in this case conversing with massage clients about skin health, influence observers' perceived ability to perform the behavior (self-efficacy) and their perceived expected outcomes of the behavior, including strategies for effective performance.

To frame the BBI, we used the 4 steps of a helping conversation (awareness, understanding, helping, and relating), client-centered communication skills, client education and referral skills, and strategies for practice system involvement developed in our prior work [15]. The helping conversation framework emphasizes a brief motivational, client-centered approach that allows a range of MT behaviors in response to the situational context (eg, new, returning, or long-term client; massage routine; practice workflow) and the client's readiness to change behavior—an approach more acceptable to MTs than proscriptive approaches to BBI used often in conventional health care contexts [15]. This framework is also easily adaptable for e-learning dissemination.

Formative Data Collection

To initiate training development, we conducted 5 key informant telephone interviews with subject matter experts (SMEs) who were licensed MTs in Arizona and had current or previous experience in MT education or online training. The interview responses illuminated strategies to engage MTs, assets to include in the training, and approaches for discussing health issues within MT practice. Specifically, the informants stressed the importance of considering the scope of practice throughout the training development (*don't diagnose*) and the role MTs play in the health of their clients. They also helped establish the desired level of information throughout the training, for example, suggesting the inclusion of more detailed information regarding skin anatomy and the effects of UVR on the skin.

We then conducted 1 focus group with 5 additional locally practicing licensed MTs. The focus groups reviewed the themes that emerged during key informant interviews and generated data to further support the training. The key results highlighted the importance of the following:

- discussion of skin cancer risk reduction during appointments and why this activity is within the scope of practice;
- myriad ways to begin a conversation about skin cancer risk reduction with clients, including personal experience and nonjudgmental comments and questions;
- major barriers to conversing with clients about skin cancer risk reduction, such as lack of confidence and knowledge about skin cancer, and how to address these barriers;
- recommendations for how to teach MTs to have conversations with their clients about skin cancer risk reduction.

The informants focused on MTs' ethical responsibility to share important health-related information with their clients, ask permission to chart any new or changing lesions noticed, and provide a list of local dermatologists for referral purposes.

E-Learning Module Development

Guided by our conceptual framework and formative data collection, we adapted content from 2 existing Web-based training programs: (1) a multimedia skin cancer risk reduction academic course, currently tailored for university students in the health sciences [20] and (2) MT client-centered communication and referral skills modules used in a BBI training for tobacco cessation [16].

We adapted skin cancer risk reduction content from the university's academic course and the *Surgeon General's Call to Action to Prevent Skin Cancer* [5] to include skin cancer risk factors, sun safety, VSA, and skin lesion assessment. We endeavored to provide MTs with a refresher of the information some may have received during their professional training, while focusing on content for the expected MT-client interaction and helping conversation. We adapted client-centered communication and referral skills content from previous studies that trained MTs to offer their clients helping conversations and referrals addressing tobacco cessation [15,16]. The 4 steps of a helping conversation as applied to skin cancer, awareness, understanding, helping, and relating, are described in Table 1.

Module development included (1) creating overall competencies and module-specific learning objectives, (2) reviewing existing curricula for structure, (3) adapting existing curricula resources or creating new multimedia content for e-learning, and (4) reviewing draft modules by SME, revising as needed. The e-training was asynchronous, interactive, and less than 2 hours in length, including accessing the modules via a Web-based learning management system, viewing and completing the modules, and completing study assessments. We chose Articulate Storyline for our e-learning course authoring software. Articulate Storyline provides the ability to create responsive modules that integrate audio, video, quiz, and activity components, allowing for a streamlined development and user interface experience.

The final training is based on 22 core competencies (see Table 1, column 2), adapted from the learning objectives of previous helping conversation-oriented Web-based training modules [15], which integrate the skin cancer content across the 4 steps of a helping conversation (see Table 1, column 1). The training comprises 6 modules (1) introduction, (2) awareness, (3) understanding, (4) helping, (5) relating, and (6) closing; each module contains photo and video media produced specifically for this project, as well as interactive activities that serve as *knowledge checks* focusing on specific content and skills.

SMEs in MT education, skin cancer, BBI training, online learning, public health, and information technology critiqued the content using an iterative process of review and structured/open-ended feedback proven successful in prior training projects [15]. The massage therapy SMEs were local and national opinion leaders, respected practitioners, and educators. Multimedia Appendix 1 contains screenshots from the Understanding module of the e-training.

Table 1. Massage Therapists Skin Health Awareness, Referral, and Education electronic training modules and competencies.

Helping conversation step and module name	Competencies	Content examples
Awareness (asking about/awareness of skin cancer risk/risk behaviors and opportunities to help)	Describe benefit of MTs as partners in skin cancer risk reduction; begin a helping conversation in a nonconfrontational and supportive way	Multimedia Appendix 1: Opportunity to help
Understanding (assessing readiness to change behaviors to reduce skin cancer risk and/or seek medical evaluation of a suspicious skin lesion and seeking understanding of the client's motivations for/against behavior change)	Apply active listening skills: open-ended questions, clarifying questions, reflective questions/statements; use positive communication skills: express empathy, avoid problem solving, avoid lecturing, avoid arguing; assess and acknowledge major barriers to skin cancer risk reduction; elicit motivators that inspire risk-reducing behaviors; reinforce motivators that inspire risk-reducing behaviors; assess and acknowledge manageable risks for skin cancer; assess and acknowledge the client's willingness to take action; assess suspicious skin lesions; set realistic goals for the outcome of helping conversations	Multimedia Appendix 2: Understanding module screenshots
Helping (offering information about skin cancer risk reduction and referrals for medical evaluation)	Recognize how to offer support and encouragement based on the client's risk profile and willingness to take action; identify different types of referral resources for professional help with skin health; provide information about professional skin health services; explain how the skills learned in this training can be applied in different situations	Multimedia Appendix 3: Returning client with suspicious lesion
Relating (arranging client follow-up on skin cancer risk reduction behaviors and referrals and offering ongoing encouragement for behavior change)	Seek permission to follow-up in a respectful manner; facilitate probability of follow-up by finishing the helping conversation on a positive note	Multimedia Appendix 4: Relating

Media Asset Development

We developed 5 brief MT-client interaction scenario example videos (average time of 30 seconds), using procedures from our previous consumer studies [21]. Specifically, we asked our SMEs in MT education to (1) review training content on helping conversation goals and skills, (2) reach consensus on the purpose of the video, and (3) review video scripts and storyboards. We recruited a convenience sample of consumers to read scripts and act as massage clients; our MT consultant acted in the MT role during video recording (see Multimedia Appendices 2 and 3 for example scenarios). We also recorded 5 testimonial videos (average time of 1 min) wherein the MTs described their experiences with providing skin cancer risk reduction information in their practices (see Multimedia Appendix 4 for example testimonials). We used the scenario and testimonial videos to enhance and reinforce topics discussed throughout the training by embedding the videos into the Web-based modules.

Development of Simulated Decision-Making Cases

Using procedures and cases from our previous BBI research [16], we developed electronic simulations of case-based practice of communication skills and application of skin cancer risk reduction knowledge using the Kynectiv DecisionSim [22] platform. Each case comprises skin cancer–focused scenarios that simulate MT decision making during a helping conversation. Furthermore, 3 MT-client interaction decision paths *optimal*, *feedback required*, and *suboptimal* are accompanied by a rubric for scoring each decision (see Figure 1). The rubric was informed by the competencies for the training modules (Table 1). Each response option is associated with a tag (in the back-end database) for each decision path (see Figure 2). To successfully complete the training, MTs must meet a minimum level of competence—selection of a response path that is within a specified range of the *ideal* interaction path (eg, appropriate MT response during a helping conversation when a client is open to a discussion about skin cancer risk reduction behavior vs a different MT response if the client is resistant to discussion). We developed 5 case simulations for participants to complete following the 6 training modules.

Figure 1. Decision path template.
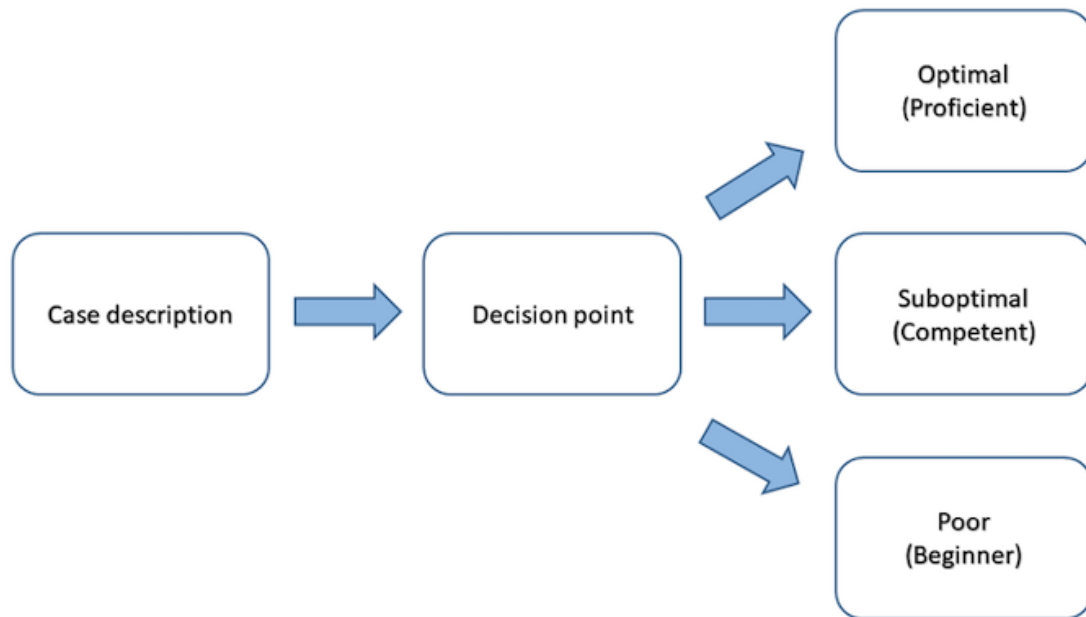
Figure 2. Screenshot of decision path tags.
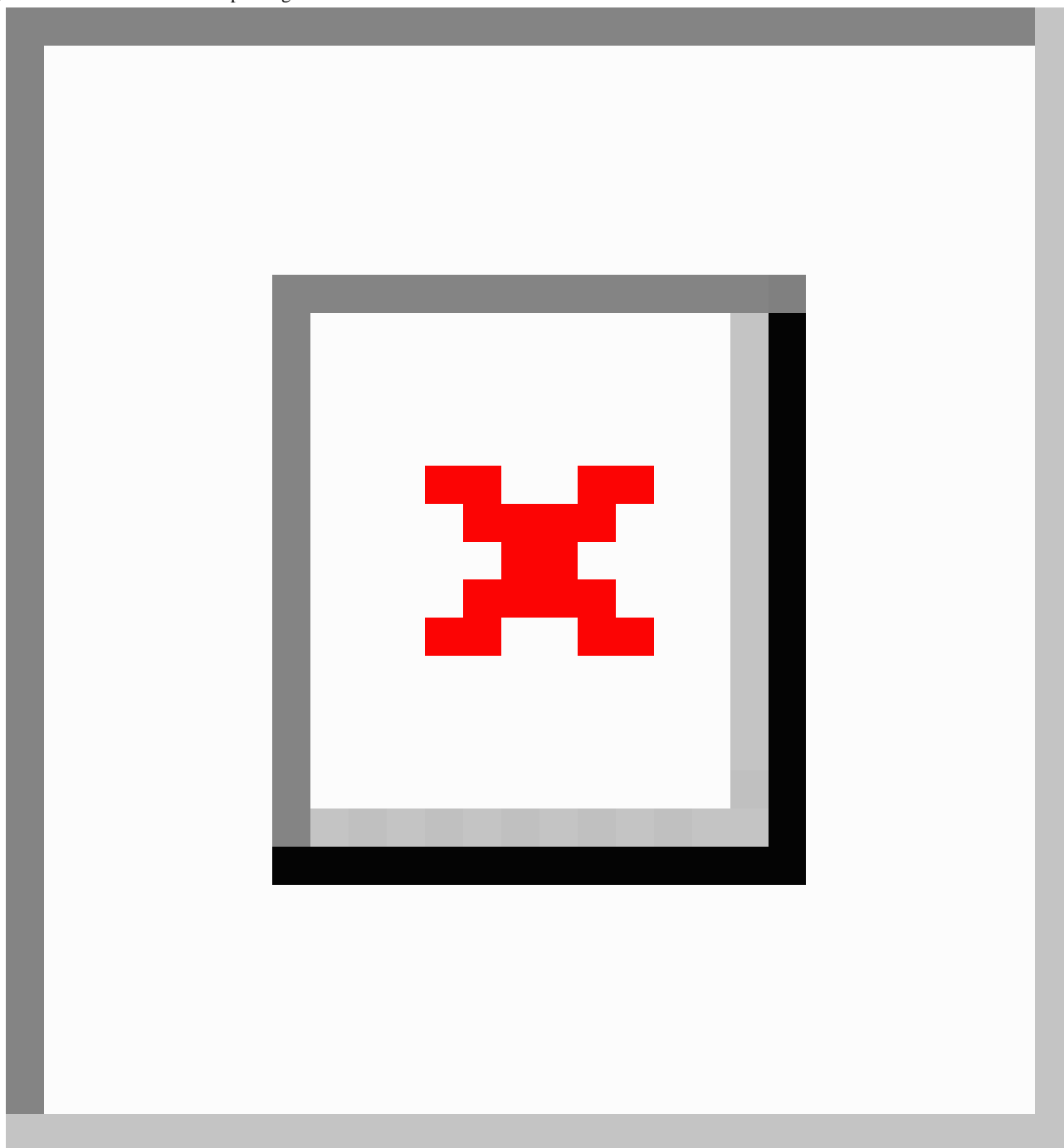
Development of Learning Management System

The learning management platform chosen to host the training modules was Desire2Learn (D2L; Desire2Learn Inc), which could seamlessly link to the case simulations. D2L allowed us to design the course components and navigation to facilitate engagement and ease of use. We designed course content to be accessed sequentially, requiring the completion of a module before accessing the next one; this ensures that participants complete the modules in the intended order, but at their own pace. The course home page features a resources section with downloadable PDFs, intended for both MT and client use. These provide skin cancer–related information, as well as tips for offering helping conversations.

Development and Adaptation of E-Learning Data Collection Methods and Assessments

The study personnel collected and managed study data using the Research Electronic Data Capture (REDCap) electronic data capture tool hosted at the University of Arizona [23]. REDCap is a secure, Web-based application designed to support data capture for research studies, which provides (1) an intuitive interface for validated data entry, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) automatic triggering of surveys and email correspondence.

We selected and modified our training assessments from the literature and those used in previous research. Participant assessments are timepoint-specific versions of 2 surveys: (1) baseline survey to assess participants' sociodemographic and practice-related data, as well as skin cancer risk reduction knowledge and (2) a case-based skin lesion image assessment that allowed participants the opportunity to view images of skin lesions and determine whether they were suspicious and prompted referral to a physician, or nonsuspicious. The baseline survey was modified into a posttest without sociodemographic data to assess knowledge and practice-related behaviors immediately posttraining and at 3 and 6 months.

We also developed a 5-item client survey to be advertised in participating MT offices and lobbies, inviting all clients of participating MTs to anonymously share whether their MT engaged in skin cancer risk reduction conversations during their massage visit. To further validate the MTs' application of the training, we asked super clients to conduct an immediate postmassage assessment. A super client is a study participant who participates in the study by receiving a massage and assesses MT's use of helping conversations pertaining to skin cancer risk reduction. This in-person observational assessment was adapted from the concept of unannounced standardized patients commonly used in medical and clinical education [24].

Iterative Usability Testing

The University's Institutional Review Board approved all human subject procedures for iterative user testing of the MTsSHARE e-training. To assess the usability of the training modules, assessments, and procedures before implementing a larger quasi-experimental longitudinal study with Arizona MTs, we enrolled a convenience sample of 20 licensed MTs from throughout the United States (except Arizona). We used the predetermined feasibility study eligibility criteria to determine MT eligibility (see below), enrolling eligible MTs in 4 waves of 5. Following all training and assessment components, participants completed a 27-item usability survey adapted from the feedback form used by SMEs during module development. The usability survey questions focused on course content (the content of the modules is at the appropriate level for MTs), accessibility (it is easy to access the helping conversation simulations), and relevance (the content in the modules is relevant to my (MT's) practice). We scored all items on a 5-point scale from strongly disagree to strongly agree. After each wave of 5, we analyzed responses and made appropriate changes to study protocol, assessment, and training components as necessary. We coded open-ended responses for major categories. The massage therapy SMEs were local and national opinion leaders, respected practitioners, and educators categories using established methods for qualitative content analysis [25]. The usability assessment was conducted from March 3, 2018 to July 31, 2018.

The overall mean scores for usability slightly improved with each iteration, increasing from a low score of 3.5 to a 5 (moderately agree to strongly agree), with an overall usability mean score of 4.96. However, the key findings from usability testing were the appropriateness of simple, seamless technology, the progression and relevance of the information presented, suggestions for additional content and general instruction, and the utility of including interactive assessments and client simulation exercises. Making changes after each wave resulted in the progressive improvement of the modules. The final version of the e-training tested well for usability and satisfaction.

Phase 2 (In Progress)

Feasibility Study Design

The feasibility study is a single cohort design (see Figure 3) with participant assessments at 4 time points: (1) immediately upon study enrollment (baseline survey and image assessment 1), (2) posttest 1 occurring immediately after training completion and image assessment, (3) posttest 2 occurring 3 months after training completion and image assessment, and (4) posttest 3 occurring 6 months after training completion and image assessment. After completing the e-training, participants receive an electronic gift card and a certificate for 1 hour of continuing education (CE), approved by the National Certification Board for Therapeutic Massage and Bodywork.

A subset of 20 Tucson-based MTs will receive a visit from a trained super client at least 3 months after completing the e-training.

The phase 2 of the study, participant enrollment and data collection, is ongoing. All survey invitations are delivered via automated email from the REDCap system, triggered by items completed in an administrative survey by study staff or timepoints based on completion of the e-training.

Figure 3. Feasibility study protocol.
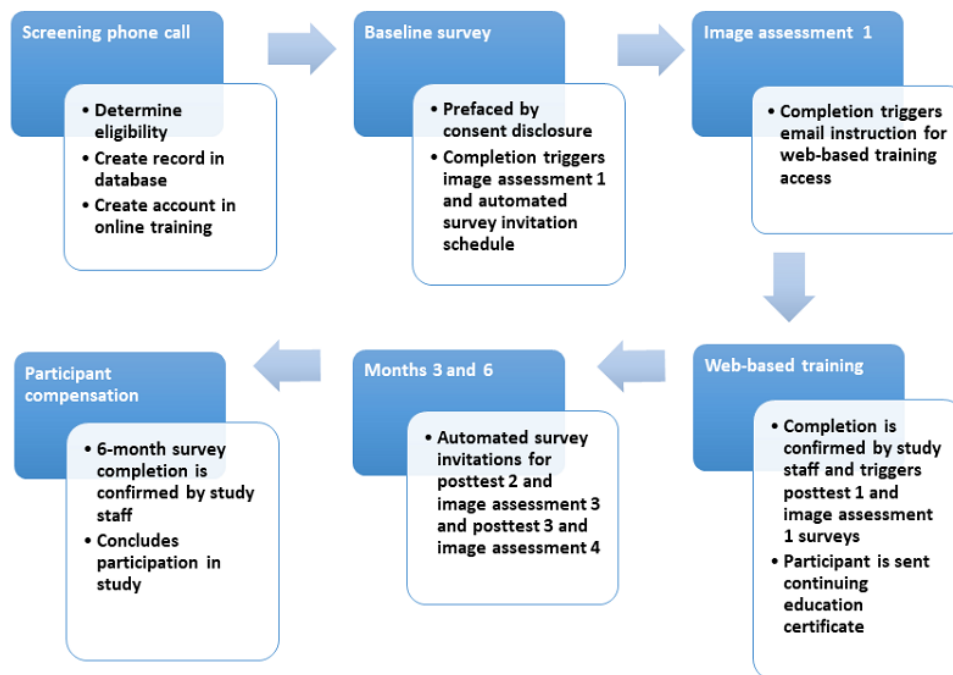
Participant Recruitment, Eligibility, and Enrollment

Our goal for enrollment is 80 MTs practicing in the state of Arizona. We based our sample size estimation on the published literature on skin cancer training for medical students [26] and our own research on tobacco cessation BBI with licensed MTs. Given the potential for a high attrition rate experienced in Web-based trainings [27], we concluded that a sample of 80 participants is sufficiently powered and allows for possible attrition. The sample size analyses were conducted using PASS (V.12). This sample size is large enough to reasonably estimate, in conjunction with sensitivity analysis, relevant variance components, recruitment, and dropout rates for use in a future definitive trial [28].

To be eligible, MTs must be aged at least 21 years, be a licensed MT in the state of Arizona, have practiced for at least 3 years, provide mainly full body massages, have access to computer with broadband internet connection and audio, and agree to forego continuing education on skin cancer for the duration of study participation. Excluded are MTs who have received continuing education on skin cancer, sun safety or client communication skills training in the past 2 years, and those who perform only partial body massages.

The recruitment began with Arizona-based MTs who initially responded to the recruitment efforts for usability testing. We then contacted state and national MT organizations, as well as Arizona-based leaders in the massage therapy business, previously known to study staff, to share recruitment materials via social media accounts.

Interested MTs email or call the designated initial contact study staff member, who enters the MTs' information into a shared recruitment database in REDCap. The study staff then schedule and conduct a screening phone call. During this call, if the MT assents verbally to study participation, study staff enroll the MT into the REDCap system, which immediately sends the baseline survey link containing the electronic consent disclosure.

Participant Training and Follow-Up (In Progress)

We designed the REDCap database to send an automated notification of completion of image assessment 1 to study staff; this instructs study staff to send the e-training login instructions to the participant via email. The participants have 2 weeks to complete the 6 Web-based training modules and 5 DecisionSim cases. The study personnel check daily for training completion, emailing the certificate for 1 hour of continuing education, and logging the date of completion in the administrative survey. Completion of the training triggers posttest 1 and image assessment 2. When a participant completes the training, study personnel distribute client survey flyers to the MT in person, electronically or via US mail.

Super Client In-Person Observational Assessment

Super clients will visit Tucson-based participating MTs for an average 60-min full body massage. Super clients will receive a simple henna tattoo by study personnel, imitating a suspicious lesion, placed on their foot or ankle region before their first massage; this tattoo will be identical on each super client and will serve as a standardizing feature. Following their massage, the super client will complete a brief electronic survey about their massage experience, focusing on whether their MT engaged in a conversation about skin cancer risk reduction and whether they mentioned the suspicious lesion (henna tattoo).

To date, we have enrolled and trained 5 super clients (4+1 alternate), who will each visit 5 MTs, for a total of 20 MTs visited. We selected a convenience sample of super clients to represent a variety of demographic characteristics (age, gender, phenotype, health history, and sun protective behaviors).

Data Analysis

Feasibility outcomes, including recruitment and dropout, training completion, overall client feedback, and MT satisfaction, will be described using frequencies and percentages and 95% CIs.

The longitudinal measures will use appropriate mixed models (linear for continuous outcomes and generalized linear with a logistic link for binary) using time categorically to protect against model misspecification. Comparisons of baseline with 3- and 6-month measures will be carried out using contrasts within these models. The mixed-models are robust to missing outcome data (including dropout) and model misspecification [29,30].

Sensitivity/specificity across 4 timepoints will be compared for image assessment scores. For each timepoint we will assess the following parameters for image assessment: sensitivity, specificity, the likelihood ratio for a positive result, and the likelihood ratio for a negative test result. We will evaluate separate bivariate logistic regression models for each set of image assessments to determine the odds ratio in predicting the correct image. We will evaluate separate models including the scores for each timepoint to determine the areas under the receiver operating characteristic (ROC) curves for image assessments. The area under the ROC curve measures the probability of correctly identifying a true negative (not suspicious) or true positive (suspicious) image.

Client survey data and super client data will be analyzed with descriptive statistics. We will correlate scores from the super client assessment with the DecisionSim scores to further validate MTs' application of helping conversation skills learned in the training. We will conduct an optional debriefing webinar for the 80 MTs in the third year to gain further information about their experience with the curriculum.

Mixed-effects linear regression models for longitudinal data will be fitted to evaluate intervention outcomes adjusted for participant characteristics, for example, age, gender, years in practice, geographical area, and client workload. The mean differences in each of the primary outcomes will be evaluated in separate models, including the covariates as fixed effects and subjects as random effects. We may also consider geographical area as random effect (urban vs rural). In this case, geographical area and subjects will be fitted as random effects to account for the correlation within geographical area and serial intrasubject correlations. Predictor variables with multiple categories will be entered as indicator variables. For dichotomized intervention effects, we will use mixed-effects logistic regression models.

Results

For Phase 2, we have screened 77 MTs who have expressed interest in participating. Of those, 14 were either not interested or not eligible (either lacked time to participate or did not see an average of 10 clients per week) and 15 did not follow up after contact attempts were made. We enrolled and consented the remaining 48 MTs. At the time of the paper submission, 11 enrolled MTs had dropped from the study, owing to lack of time to participate. Of the 37 MTs still enrolled, 32 have completed the training, with the remaining 5 having begun but not yet completed the training. We will close recruitment in August 2019.

Discussion

Principal Findings

The current prevention and early detection strategies have not had a significant impact on reducing the public health burden of skin cancer [5]. We used a rigorous strategy to enlist MTs as partners in skin cancer prevention and detection, developing innovative e-training and assessment protocols. The American Massage Therapy Association estimates that there are 335,000 to 385,000 licensed or certified MTs in the United States, who see about 39.1 million clients annually [31]. Thus, MTs are a largely untapped resource for reducing skin cancer risk. A search of MT training on PubMed and Google Scholar revealed that most articles focus on massage therapy as an intervention and the health outcomes of massage. Few scholarly articles addressed training MTs for a specific skill following their primary professional education [15,16]. Muramoto et al [16] were the first to successfully develop and implement e-training (BBI training and competency evaluation) of complementary and alternative medicine providers (including MTs) for screening clients for tobacco use and encouraging tobacco cessation. These authors also were the first to develop e-training for these specialized providers.

The structured and rigorous development approach for this skin cancer risk reduction and BBI e-training for MTs begins to fill a gap in skin cancer risk reduction research. We surveyed 100 MTs in an elecronic, national survey where we asked for the MTs' perceptions of conversations with clients related to skin cancer prevention, as well as detection [32]. The 2 published studies have targeted skin cancer risk reduction in convenience samples of MTs who were attending national MT conferences. One study surveyed 262 MTs to assess their comfort level regarding potential assessment of suspicious skin lesions [17]. The other study reported findings from a face-to-face, 4-hour education session that provided information only to 114 MTs [33]. No previous studies have addressed how MTs could integrate this information into the context of a client visit via client education, or communication skills, such as a BBI (helping conversation) to encourage skin cancer risk reduction. Our e-training approach offers greater learner accessibility, increased convenience, and greater scalability [34]. Thus, the e-training format has the potential to reach many more MTs, nationally.

We found few other e-learning opportunities pertaining to skin cancer, most of which targeted conventional health care providers with a goal of increasing competencies in diagnostic knowledge and skills competency [35-37]. These ranged from several Web-based modules to video training delivered by electronic links [37]. In these studies, providers had a positive impression of the Web-based curriculum, and in one case, increased the likelihood of discussion with patients about skin cancer. The accessibility, effectiveness, and popularity of the curriculum indicated potential for implementation in the primary care setting. Our e-training is designed to be brief, yet engaging, informative, and integrated into the context of a typical client visit to an MT. MTs can access the training when convenient

XSL•FO

RenderX

and move from one module to the next at their own pace, both of which are important for learner control and engagement [38].

Previous training targeting MTs did not appear to be pilot tested or assessed for usability. Our use of formative and summative evaluations along with predesignated *stopping rules* (ie, 4 iterations) represented the ideal conceptualization of usability [39]. The ease of navigation of the training modules and available resources made this training appealing to the participating MTs. The training incorporated highly interactive, scenario-based, simulated helping conversations focused on skin cancer risk reduction, and the simulations provided participants with opportunity to interact with the training, apply knowledge gained, and practice skills learned, reflecting the SCT theory. These features also are important to enhancing e-learning [34].

Barriers and Opportunities

The preliminary results reveal anticipated difficulties with recruitment and retention within the MT population. No previously published studies of MTs as participants have addressed recruitment challenges. For the in-progress feasibility study, barriers to enrollment have related to practice-related eligibility conditions, such as number of years in practice and number of clients seen per week. The primary barrier to retention following enrollment has been a self-professed lack of time to participate. It is encouraging that, of the eligible and enrolled MTs, 67% (32/48) have completed the training and progressed to follow-up assessments. Offering incentives in the form of monetary compensation, as well as continuing education credit, has been a useful approach to address both recruitment and retention.

This paper provides an overview of our systematic approach to developing rigorous e-training for MTs to enable them to be partners in skin cancer risk reduction. The phase 2 results will explicate the feasibility of the e-training approach for further efficacy testing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Opportunity to help.

[MP4 File (MP4 Video), 87MB - resprot_v8i5e13480_app1.mp4]

Multimedia Appendix 2

Understanding module screenshots.

[PDF File (Adobe PDF File), 427KB - resprot_v8i5e13480_app2.pdf]

Multimedia Appendix 3

Returning client with suspicious lesion.

[MP4 File (MP4 Video), 59MB - resprot_v8i5e13480_app3.mp4]

Multimedia Appendix 4

Relating.

[MP4 File (MP4 Video), 16MB - resprot_v8i5e13480_app4.mp4]

References

1. Guy Jr GP, Machlin SR, Ekwueme DU, Yabroff KR. Prevalence and costs of skin cancer treatment in the U.S., 2002-2006 and 2007-2011. Am J Prev Med 2015 Feb;48(2):183-187 [FREE Full text] [doi: 10.1016/j.amepre.2014.08.036] [Medline: 25442229]

XSL•FO

RenderX

2. Schmults CD, Karia PS, Carter JB, Han J, Qureshi AA. Factors predictive of recurrence and death from cutaneous squamous cell carcinoma: a 10-year, single-institution cohort study. JAMA Dermatol 2013 May;149(5):541-547. [doi: 10.1001/jamadermatol.2013.2139] [Medline: 23677079]

3. Jensen AO, Lamberg AL, Jacobsen JB, Braae Olesen A, Sørensen HT. Non-melanoma skin cancer and ten-year all-cause mortality: a population-based cohort study. Acta Derm Venereol 2010 Jul;90(4):362-367 [FREE Full text] [doi: 10.2340/00015555-0899] [Medline: 20574600]

4. American Cancer Society. Atlanta; 2018. Cancer facts and figures 2018 URL: https://tinyurl.com/yb54a93t [accessed 2018-10-02] [WebCite Cache ID 786CYQLvT]

5. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. The Surgeon General's Call to Action to Prevent Skin Cancer URL: https://www.surgeongeneral.gov/library/calls/prevent-skin-cancer/call-to-action-prevent-skin-cancer.pdf [accessed 2019-05-08] [WebCite Cache ID 78DEyrlVD]

6. Curiel-Lewandrowski C, Chen SC, Swetter SM, Melanoma Prevention Working Group-Pigmented Skin Lesion Sub-Committee. Screening and prevention measures for melanoma: is there a survival advantage? Curr Oncol Rep 2012 Oct;14(5):458-467 [FREE Full text] [doi: 10.1007/s11912-012-0256-6] [Medline: 22907282]

7. Katalinic A, Waldmann A, Weinstock MA, Geller AC, Eisemann N, Greinert R, et al. Does skin cancer screening save lives?: an observational study comparing trends in melanoma mortality in regions with and without screening. Cancer 2012 Nov 01;118(21):5395-5402 [FREE Full text] [doi: 10.1002/cncr.27566] [Medline: 22517033]

8. Waldmann A, Nolte S, Weinstock MA, Breitbart EW, Eisemann N, Geller AC, et al. Skin cancer screening participation and impact on melanoma incidence in Germany--an observational study on incidence trends in regions with and without population-based screening. Br J Cancer 2012 Feb 28;106(5):970-974 [FREE Full text] [doi: 10.1038/bjc.2012.22] [Medline: 22294187]

9. Hoorens I, Vossaert K, Pil L, Boone B, De Schepper S, Ongenae K, et al. Total-body examination vs lesion-directed skin cancer screening. JAMA Dermatol 2016 Jan;152(1):27-34. [doi: 10.1001/jamadermatol.2015.2680] [Medline: 26466155]

10. Rogers HW, Weinstock MA, Feldman SR, Coldiron BM. Incidence estimate of nonmelanoma skin cancer (Keratinocyte Carcinomas) in the U.S. population, 2012. JAMA Dermatol 2015 Oct;151(10):1081-1086. [doi: 10.1001/jamadermatol.2015.1187] [Medline: 25928283]

11. Nahin RL, Barnes PM, Stussman BJ, Bloom B. Costs of complementary and alternative medicine (CAM) and frequency of visits to CAM practitioners: United States, 2007. Natl Health Stat Report 2009 Jul 30(18):1-14 [FREE Full text] [Medline: 19771719]

12. Cherkin DC, Deyo RA, Sherman KJ, Hart LG, Street JH, Hrbek A, et al. Characteristics of visits to licensed acupuncturists, chiropractors, massage therapists, and naturopathic physicians. J Am Board Fam Pract 2002;15(6):463-472 [FREE Full text] [Medline: 12463292]

13. Heiligers PJ, de Groot J, Koster D, van Dulmen S. Diagnoses and visit length in complementary and mainstream medicine. BMC Complement Altern Med 2010 Jan 25;10:3 [FREE Full text] [doi: 10.1186/1472-6882-10-3] [Medline: 20100343]

14. Grossman DC, Curry SJ, Owens DK, Barry MJ, Caughey AB, Davidson KW, et al. Behavioral counseling to prevent skin cancer: US Preventive Services Task Force recommendation statement. J Am Med Assoc 2018 Dec 20;319(11):1134-1142. [doi: 10.1001/jama.2018.1623] [Medline: 29558558]

15. Muramoto ML, Matthews E, Ritenbaugh CK, Nichter MA. Intervention development for integration of conventional tobacco cessation interventions into routine CAM practice. BMC Complement Altern Med 2015 Mar 29;15:96 [FREE Full text] [doi: 10.1186/s12906-015-0604-9] [Medline: 25887742]

16. Muramoto ML, Howerter A, Eaves ER, Hall JR, Buller DB, Gordon JS. Online tobacco cessation training and competency assessment for complementary and alternative medicine (CAM) practitioners: protocol for the CAM Reach Web Study. JMIR Res Protoc 2016 Jan 06;5(1):e2 [FREE Full text] [doi: 10.2196/resprot.5061] [Medline: 26740468]

17. Campbell SM, Louie-Gao Q, Hession ML, Bailey E, Geller AC, Cummins D. Skin cancer education among massage therapists: a survey at the 2010 meeting of the American Massage Therapy Association. J Cancer Educ 2013 Mar;28(1):158-164. [doi: 10.1007/s13187-012-0403-7] [Medline: 22915212]

18. Pears S, Morton K, Bijker M, Sutton S, Hardeman W, VBI Programme Team.. Development and feasibility study of very brief interventions for physical activity in primary care. BMC Public Health 2015 Apr 08;15:333 [FREE Full text] [doi: 10.1186/s12889-015-1703-8] [Medline: 25887643]

19. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]

20. Davis R, Loescher LJ, Rogers J, Spartonos D, Snyder A, Koch S, et al. Evaluation of project Students are Sun Safe (SASS): a university student-delivered skin cancer prevention program for schools. J Cancer Educ 2015 Dec;30(4):736-742. [doi: 10.1007/s13187-014-0742-7] [Medline: 25417824]

21. Loescher LJ, Hansen C, Hepworth JT, Quale L, Sligh J. A preliminary study of a video intervention to inform solid organ transplant recipients about skin cancer. Transplant Proc 2013 Nov;45(9):3187-3189 [FREE Full text] [doi: 10.1016/j.transproceed.2012.09.118] [Medline: 23747188]

22. DecisionSim. kynectiv. Decision Simulation Software URL: https://www.kynectiv.com/ [accessed 2019-05-03] [WebCite Cache ID 786DaKkYE]

23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

24. Taormina DP, Zuckerman JD, Karia R, Zabar S, Egol KA, Phillips DP. Clinical skills and professionalism: assessing orthopaedic residents with unannounced standardized patients. J Surg Educ 2018;75(2):427-433. [doi: 10.1016/j.jsurg.2017.08.001] [Medline: 28888419]

25. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]

26. Cliff S, Bedlow AJ, Melia J, Moss S, Harland CC. Impact of skin cancer education on medical students' diagnostic skills. Clin Exp Dermatol 2003 Mar;28(2):214-217. [Medline: 12653717]

27. Tyler-Smith K. Early attrition among first time eLearners: a review of factors that contribute to drop-out, withdrawal and non-completion rates of adult learners undertaking eLearning programs. JOLT 2006;2(2):73-85.

28. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceut Statist 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]

29. Mallinckrodt CH, Watkin JG, Molenberghs G, Carroll RJ. Choice of the primary analysis in longitudinal clinical trials. Pharmaceut Statist 2004 Jul;3(3):161-169. [doi: 10.1002/pst.124] [Medline: 25855820]

30. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. Stat Methods Med Res 2014 Oct;23(5):440-459. [doi: 10.1177/0962280213476378] [Medline: 23427225]

31. American Massage Therapy Association. 2017. Massage Therapy Industry Fact Sheet Online URL: https://www.amtamassage.org/infocenter/economic_industry-fact-sheet.html [accessed 2019-05-03] [WebCite Cache ID 786DwdxHh]

32. Loescher LJ, Howerter AL, Heslin KM, Azzolina CM, Muramoto ML. A survey of licensed massage therapists' perceptions of skin cancer prevention and detection activities. Int J Ther Massage Bodywork 2018;11(4):4-10. [Medline: 30524632]

33. Trotter SC, Louie-Gao Q, Hession MT, Cummins D. Skin cancer education for massage therapists: a novel approach to the early detection of suspicious lesions. J Cancer Educ 2014;29(2):266-269. [Medline: 24407881]

34. Noesgaard S, Orngreen R. The effectiveness of e-learning: an explorative and integrative review of the definitions, methodologies and factors that promote e-learning effectiveness. EJEL 2015;13(4):278-290.

35. Harris JM, Salasche SJ, Harris RB. Can internet-based continuing medical education improve physicians' skin cancer knowledge and skills? J Gen Intern Med 2001 Jan;16(1):50-56 [FREE Full text] [Medline: 11251750]

36. Jiang AJ, Eide MJ, Alexander GL, Altschuler A, Asgari AM, Geller AC, et al. Providers' experiences with a melanoma web-based course: a discussion on barriers and intentions. J Cancer Educ 2017 Jun;32(2):272-279 [FREE Full text] [doi: 10.1007/s13187-015-0910-4] [Medline: 26391994]

37. Lofgreen S, Lehrer M, Bennett P, Garg A, Dunnick CA. Integrating the integrated skin exam film into medical education. Dermatol Online J 2016 Nov 15;22(11). [Medline: 28329558]

38. Karich AC, Burns MK, Maki KE. Updated meta-analysis of learner control within educational technology. Rev Educ Res 2014 Sep;84(3):392-410. [doi: 10.3102/0034654314526064]

39. Lewis JR. Usability: lessons learned … and yet to be learned. Int J Hum Comput Stud 2014 Jun 18;30(9):663-684. [doi: 10.1080/10447318.2014.930311]

Abbreviations

BBI: brief behavioral intervention
D2L: Desire2Learn
e-learning: electronic learning
e-training: electronic training
MT: massage therapist
MTsSHARE: Massage Therapists Skin Health Awareness, Referral, and Education
REDCap: Research Electronic Data Capture
ROC: receiver operating characteristic
SCT: social cognitive theory
SME: subject matter expert
UVR: ultraviolet radiation
VSA: visual skin assessment

XSL•FO
RenderX

<u>Protocol</u>

Evaluating a Web-Based Mental Health Service for Secondary School Students in Australia: Protocol for a Cluster Randomized Controlled Trial

Bridianne O'Dea[1,2], BHlthSci (Hons), PhD; Catherine King[1,2], BLibSt (Hons), MClinPsych; Mirjana Subotic-Kerry[1], GradDipPsych, BSc (Hons), PhD; Melissa Anderson[1], BCA, BA Psych (Hons), MCouns&PsychTh; Melinda Rose Achilles[1], BBehavSc (Psych); Belinda Parker[1], GradDipSocSc, BSocSc (Hons), PhD; Andrew Mackinnon[1,3], PhD; Josey Anderson[1,2], MMed; Nicole Cockayne[1,2], BSc (Hons), MPH; Helen Christensen[1,2], MPsych, PhD

[1]Black Dog Institute, Sydney, Australia
[2]Faculty of Medicine, University of New South Wales, Sydney, Australia
[3]Centre for Mental Health, University of Melbourne, Melbourne, Australia

Corresponding Author:
Bridianne O'Dea, BHlthSci (Hons), PhD
Black Dog Institute
Prince of Wales Hospital, Hospital Road, Randwick
Sydney, 2031
Australia
Phone: 61 0293828509
Email: b.odea@blackdog.org.au

Abstract

Background: Mental health problems are prevalent among Australian secondary school youth; however, help-seeking is low. Schools offer an ideal setting to address these concerns. The Black Dog Institute has developed a Web-based mental health service for secondary schools that is modeled on the principles of stepped care. The Smooth Sailing service aims to improve help-seeking and reduce anxiety and depressive symptoms in secondary school students. The acceptability of this service has been demonstrated in a pilot study. A full trial is now warranted.

Objective: This study protocol for a cluster randomized controlled trial (RCT) aims to evaluate the effectiveness of the Smooth Sailing Web-based service for improving help-seeking intentions and behavior, and reducing depressive and anxiety symptoms, alongside other mental health outcomes, when compared with a school-as-usual control condition in secondary school youth.

Methods: This RCT aims to recruit 1600 students from 16 secondary schools in regional and urban locations throughout New South Wales, Australia. Schools are randomly assigned to the intervention or school-as-usual control condition at the school level. Approximately 100 students from 1 or multiple grades are recruited from each participating school. Participants complete measures at 3 timepoints: baseline, 6 weeks post, and 12 weeks post, with the primary outcome assessed at 12 weeks posttest. Participants assigned to the intervention condition register to the Web-based service at baseline and receive care in accordance with the service model. Participants in the control condition receive school-as-usual.

Results: The first baseline assessment occurred on February 22, 2018, with the 12-week endpoint assessments completed on Friday, June 29, 2018. Control schools are currently receiving the service, due for completion by June 30, 2019. The trial results are expected to demonstrate improved help-seeking intentions and behavior among students assigned to the intervention condition, alongside improvements in symptoms of depression, anxiety, distress, and other mental health outcomes when compared with students assigned to the control condition.

Conclusions: To our knowledge, this is the first time that a Web-based mental health service based on the principles of stepped care will have been integrated into, and evaluated in, the Australian school context. The findings of this trial will have implications for the suitability of this type of service model in Australian schools and for the delivery of school-based mental health services more broadly.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12618001539224 https://anzctr.org.au/Trial/Registration/TrialReview.aspx?id=375821&isReview=true (Archived by WebCite at http://www.webcitation.org/77N3MDGS6)

Introduction

Background

Adolescence is a key developmental period for mental illness with 50% of all mental disorders emerging before the age of 18 years [1]. As most young people spend these years in secondary education, schools now play an increasing role in addressing the mental health needs of their students. Schools are ideal settings for recognizing the early behavioral and emotional signs of mental illness among students and for trusted adults to initiate help-seeking [2]. This is important as youth are reluctant to seek formal care [2,3], despite the negative impacts of poor mental health on social and educational functioning [4-7]. Schools have initiated a range of programs aimed at reducing mental illness and improving mental health literacy among students, as well as increasing their likelihood of seeking help [8-11]. However, many of these programs lack evidence or have not had substantial uptake because of implementation barriers [12]. A total of 2 major challenges remain: identifying the students who are experiencing mental health problems and delivering care to those who require it.

Currently, most schools utilize a wait-to-act model in which the staff refer students to school counselling services, learning support or well-being teams after observing concerning behaviors or from students' self-disclosures [13]. A more proactive approach, one that incorporates mental health screening and automatic stratification and response based on symptom severity, may improve schools' capacity to address students' mental health needs. Web-based stepped care presents a viable option for providing this type of care. This is based on the premise that simple, cost-effective Web-based psychotherapy is offered to youth with mild-moderate symptoms, with more costly, intensive face-to-face interventions reserved for those with more severe and persistent symptoms [14]. Although complex, this approach can be efficient, as it provides tailored help as required and aims to prevent serious mental illness by detecting symptoms early [15,16]. This type of care is also equitable, providing all students with an opportunity to have their mental health needs addressed. Internet screening and Web-based cognitive behavioral therapy (CBT) programs can be readily integrated into such a model as these require minimal human input, are acceptable to youth, preserve fidelity of care, and allow for ongoing monitoring and automated feedback [17]. Notably, this type of care system is engineered to reach out to youth rather than wait for them to approach.

Codesigned with students, school counselors [18], general practitioners [19], and parents [20], the Black Dog Institute has developed a Web-based mental health service called Smooth Sailing. On the basis of the principles of stepped care, Smooth Sailing uses a website to screen, assess, allocate, and deliver psychological interventions to improve help-seeking for mental

health problems and reduce depressive and anxiety symptoms among secondary school youth. This service uses brief, validated self-report measures for anxiety and depression [21,22] to accurately determine young people's symptom severity and their required level of care. The service has 3 degrees of treatment intensity that are consistent with Australian Clinical Practice Guidelines [23]: self-directed Web-based psychoeducation for students with nil to mild symptoms, self-directed Web-based CBT for students with moderate symptoms [9,10], and a direct link to face-to-face care with a school counselor for students with moderately severe to severe symptoms. Students are monitored fortnightly using a brief survey delivered via short message service (SMS) text messaging or email. Every 6 weeks, students complete a step assessment from which care is reallocated based on their results. If a student reports having thoughts of self-harm or death during any of the assessments, the school counselor receives an automatic notification.

The Smooth Sailing service is designed as a universal intervention to improve young people's attitudes toward seeking help for mental health problems. School-based screening has increased referral rates to health care services, demonstrating the positive impact that schools can have on initiating access to care [24,25]. Establishing healthy attitudes to help-seeking for mental health in adolescence, particularly from adults and professionals, is key to supporting lifelong mental health. A number of studies have revealed an association between positive attitudes to professional help-seeking and increased help-seeking intentions and behaviors in adolescents [26,27], university students [28], and adults [29]. The Smooth Sailing service is based on Rickwood et al's [30] model in which help-seeking is defined as a process with 4 key stages. Outlined in Figure 1, Smooth Sailing attempts to target each of these stages through a range of different functions and content.

A 6-week single-arm pilot study of the Smooth Sailing service was undertaken in 4 New South Wales (NSW) secondary schools in 2017 with 59 students taking part. At posttest, 93% (55/59) of students remained enrolled in the service. The service was found to be acceptable and feasible to students, parents, and school counselors. Students followed up by the school counselor reported significant symptom improvements at posttest. Importantly, most of the participants found the service easy to understand (96%, 53/55), easy to use (95%, 52/55), and enjoyable (89%, 49/55). The majority were also comfortable being followed up by the counselor (82%, 45/55), were likely to use the service again (73%, 40/55) and would tell a friend to use the service (85%, 47/55). However, key questions remain because of the small sample size and lack of a control group. Despite the promise and practicality of the proposed service model, the effectiveness of universal screening, stepped care, and active follow-up from school counselors for improving

young people's help-seeking and reducing symptoms is still unknown. It is now timely to examine the primary effects of this type of service in comparison with a control group.

Figure 1. Applying Rickwood et al's help-seeking model to the Smooth Sailing service.
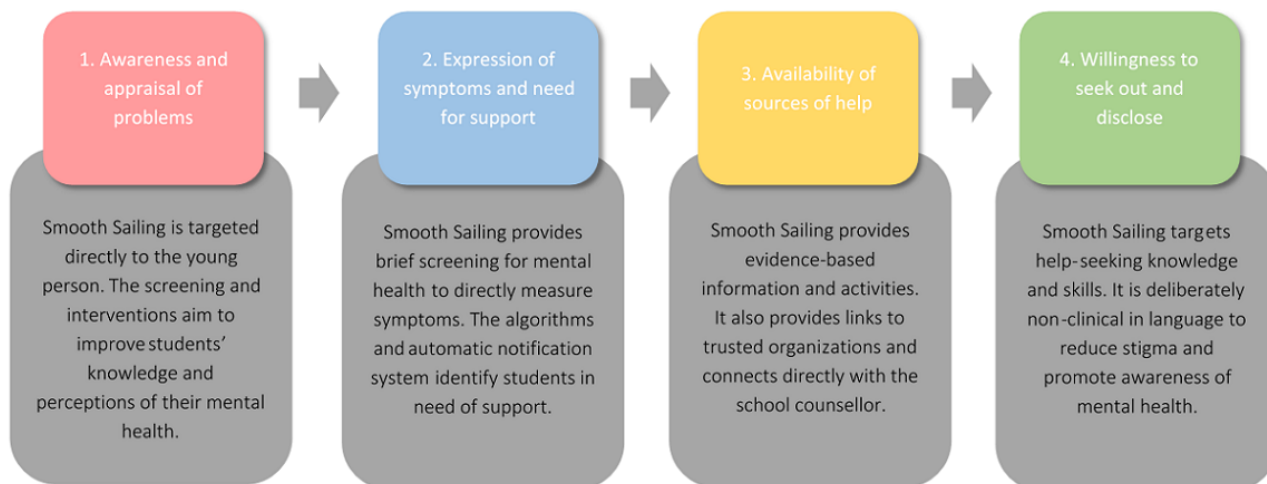
Objectives

The main objective of this trial is to evaluate the effectiveness of the Smooth Sailing service for improving help-seeking intentions (primary outcome), as well as for improving help-seeking behavior, reducing symptoms of depression, anxiety, distress, and barriers to care, and improving mental health literacy/stigma (secondary outcomes) among secondary school students, in comparison with a school-as-usual control group.

Hypotheses

The primary hypothesis is that those who receive the Smooth Sailing service will report higher scores of overall help-seeking intentions at 12 weeks posttest compared with baseline. The secondary hypothesis is that compared with students in the control condition, those assigned to intervention will report improved help-seeking behavior and reduced levels of depression, anxiety, distress, and barriers to care, as well as improved mental health literacy and stigma at 12 weeks posttest.

Methods

Trial Design

This clinical trial protocol adheres to the Standard Protocol Items: Recommendations for Interventional Trials 2013 guidelines [31] (see Table 1). This study is a 2-arm 12-week cluster randomized controlled trial (RCT). Data are collected at 3 intervals: baseline, 6 weeks, and 12 weeks posttest with the primary outcome measure assessed at 12 weeks posttest.

Randomization

Assignment of schools to the control or intervention condition is carried out according to the International Council for Harmonisation guidelines [32] and performed by a researcher not involved in the day-to-day conduct of the trial. Schools are allocated to a single condition (cluster design) to avoid contamination and for administrative convenience [33]. A minimization approach [34,35] is used to ensure balance across

conditions in terms of the Index of Community Socio-Educational Advantage level (<1000 vs ≥1000), gender mix (coeducational vs single sex), and year level involved (year 9 students only vs multiple or other years). Minimization is undertaken in StataCorp LLC Stata statistical software version 14.2 using the rct_minim procedure [36]. The pool of already available schools is sorted in random order using the Excel 2003 data analysis random number generator and entered into the minimization routine in ascending order using the factors specified above. Subsequent schools are assigned to an intervention arm using rct_minim in the order that they join the trial and provide complete information. As this trial uses a no-treatment comparator control group, participants and researchers are not blinded to the allocation assignment.

Ethics Approval

Ethics approvals are obtained from the University of New South Wales (UNSW) Human Research Ethics Committee (HREC; HC17910), the State Education Research Applications Process (SERAP) for the NSW Department of Education (SERAP 2016471), the Sydney Catholic Schools (SCS) Research Centre (20186), and the Catholic Schools Office Diocese of Maitland-Newcastle.

Setting

This trial is conducted in government, independent, and Catholic secondary schools located throughout NSW, Australia.

Participants

Inclusion Criteria

All secondary students from the participating year groups, aged between 11 and 19 years, who attend one of the participating schools, are invited to participate. Both males and females are eligible. Participants are required to have an active email address for the duration of the trial. Only those students who can read and understand English and provide their signed written consent are able to participate in the research study and use the service.

All participants can access their own mental health support or treatment throughout the trial.

Exclusion Criteria

Schools are required to have a school counsellor available onsite for the school visits and for the duration of the study.

Table 1. Standard Protocol Items: Recommendations for Interventional Trials Compliance: Items from the World Health Organization dataset.

Data category	Information
Primary registry and trial identifying number	Australian New Zealand Clinical Trials Registry (ANZCTR): ACTRN12618001539224
Date of registration in primary registry	September 14, 2018
Secondary identifying numbers	__a
Source(s) of monetary or material support	Hong Kong and Shanghai Banking Corporation
Primary sponsor	Black Dog Institute, University of New South Wales
Secondary sponsor(s)	—
Contact for public queries	Dr Bridianne O'Dea b.odea@blackdog.org.au
Contact for scientific queries	Dr Bridianne O'Dea b.odea@blackdog.org.au
Public title	Smooth Sailing: Evaluating an online service for student well-being
Scientific title	Smooth Sailing: Evaluating an online service for student well-being
Countries of recruitment	Australia
Health condition(s) or problem(s) studied	Depression, anxiety, help-seeking for mental health
Intervention(s)	Intervention: A Web-based school-based mental health service for secondary students (Smooth Sailing); based on the principles of stepped care; 12-week duration
Control	Control: school-as-usual; waitlist
Key inclusion and exclusion criteria	Inclusion: Male and female secondary students; currently attending high school at 1 of the participating schools; have an active email address for the duration of the trial; Exclusion criteria: nil
Study type	Study type: Interventional; Allocation: Randomized; Intervention model: Parallel assignment; Primary purpose: Prevention
Date of first enrolment	February 22, 2018
Target sample size	1600
Recruitment status	Completed
Primary outcome(s)	Help-seeking intentions (measured using the General Help-Seeking Questionnaire). Primary timepoint: 12-week endpoint
Key secondary outcomes	Help-seeking behavior (measured using the Actual Help-Seeking Questionnaire); Symptom levels of: Distress (measured using the Distress Questionnaire-5); Depression (Center for Epidemiologic Studies Depression Scale-Child Version); Anxiety (measured using the Generalized Anxiety Disorder Questionnaire 7-item); Mental health literacy and stigma (measured using the Mental Health Literacy and Stigma Scale); Barriers to seeking help (measured by the Barriers to Seeking Help-Brief Scale); Functioning (measured by 1-tem scale). Primary timepoint: 12-week endpoint
Ethics review	Approved by University of New South Wales Human Research Ethics Committee on November 24, 2017
Completion date	June 29, 2018
Summary results	Data currently under analysis
Individual participant data sharing statement	Plan to share individual participant data (No)

[a]Not applicable.

Recruitment

The flow chart used to outline recruitment, randomization, and participation for this trial is provided in Figure 2.

To recruit schools, the study advertisement is circulated by email to NSW school counsellors who subscribe to the Black Dog Institute's mailing list. Counsellors are invited to express an interest in the study by contacting the chief investigator (CI)

by email. The CI responds to these emails with a study information pack and schedules a phone call to explain the study process and answer additional questions. Schools are informed that the study is available to students in any year group but is considered by the research team to be well suited to students in Grade 9 (ages 14 to 15 years), as half of all lifetime mental disorders emerge during the mid-teens [1]. To recruit students, a set of Student Information Forms are mailed to each school

and distributed to students by a school staff member before the baseline assessment. Interested students are asked to review the material, discuss with their parents/guardians, and consider study participation.

Figure 2. Consolidated Standards of Reporting Trials flow diagram that will be used to illustrate participation throughout the phases of the Smooth Sailing Cluster Randomized Controlled Trial.
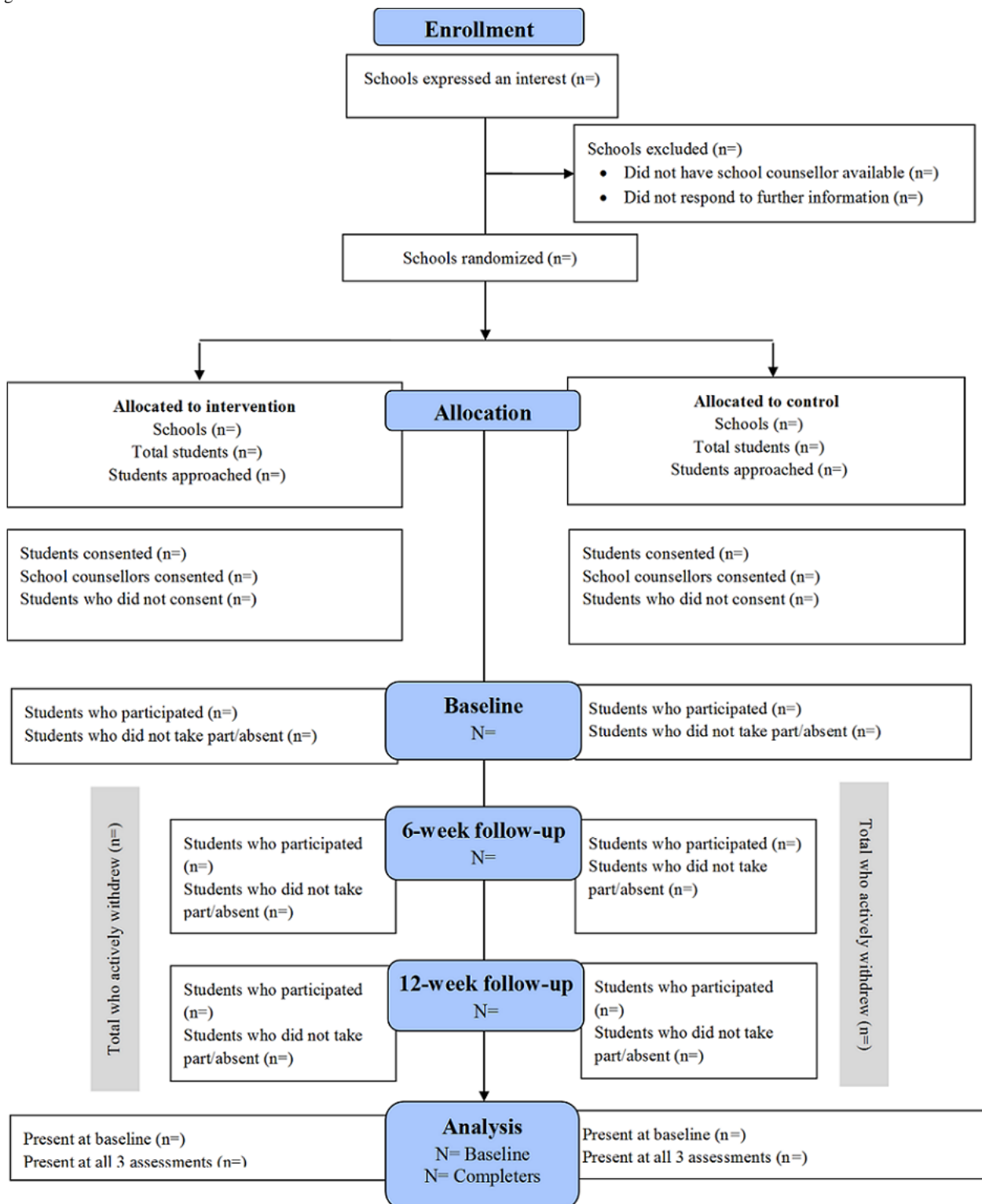
Consent

For school consent, schools are asked to provide a signed letter of support from the school principal. This letter is then forwarded to the governing ethics bodies to confirm school participation. Individual student consent is provided by a signed copy of the Participant Information and Consent Form (PICF) that is obtained on the day of the baseline assessment. For all schools other than those in the Sydney Catholic Diocese, an opt-out consent process is employed for parental consent. In these schools, parents are notified of the study using the school's usual methods of communication (eg, school newsletter). Parents

are given 14 days to inform the school if they do not wish for their child to participate and can withdraw their child at any time. Students from schools in the Sydney Catholic Diocese are required to obtain signed parental consent. The PICFs for SCS are distributed to students 2 weeks before the baseline assessment. On the day of the baseline visit, PICFs are checked by the researchers to ensure student and parent consent is provided.

Withdrawal of Consent

The PICF informs participants that taking part is completely voluntary and that they are free to withdraw from the study at any time without penalty and without having to give a reason. Participants can withdraw by emailing the research team or notifying the researchers at the school visits. Parents can also withdraw their child at any time using the same methods or by contacting their child's school.

The Intervention

The Smooth Sailing service is accessed using a website that requires students to register using a unique code. At registration, students complete an additional consent measure (a 6-item Gillick Competency Test) to ensure they understand the terms and conditions of the service. After correctly completing this measure, students are asked to input their code, provide their email address, mobile phone number (optional) and create a password. Once registered, the students complete a step assessment. This consists of reliable and valid measures of depression and anxiety symptoms [21]. The 9-item Patient Health Questionnaire-9 (PHQ-9) [21] is used to measure the presence of depressive symptoms over the past 2 weeks, with

items rated on a scale from not at all (0) to nearly every day (4). The service also uses the Generalized Anxiety Disorder Scale-7 (GAD-7) [22] to measure the presence of anxiety symptoms in the past 2 weeks using the same response options as the PHQ-9. Students' total scores on the PHQ-9 and GAD-7 scales are used to allocate them to a step of care (see Table 2).

After the step assessment, the Smooth Sailing service produces a personalized dashboard that provides students with an overview of their recommended activity. The Web-based psychoeducation consists of 5 10-min modules that provide information about anxiety, depression, and help-seeking (see Table 3). The modules were created specifically for the Smooth Sailing service and were reviewed in the co-design process by young people as well as a clinical psychologist. The content was also edited by a copywriter to ensure it was written at an appropriate reading level. The modules are complemented by animations and illustrations as well as hyperlinks to other credible youth mental health services and websites. All modules are self-paced and can be completed in any order. Module 6 includes referral to 2 Web-based, publicly available, free evidence-based CBT programs for depression and anxiety [37-39]. MoodGym [40] comprises 5 modules in which young people learn a range of strategies to identify and manage unhelpful patterns of thinking, connect their thoughts to their feelings, reduce worry, and improve self-esteem and interpersonal relationships. The BRAVE Program [41] includes 10 1-hour self-directed sessions that are usually completed over 10 weeks that teach young people to identify anxiety and stress, develop relaxation and problem-solving skills, and reframe negative thinking. These programs are provided to any student who is allocated to Steps 2, 3, and 4 throughout the study period.

Table 2. Smooth Sailing service model: step criteria and intervention provided.

Criteria	Step				
	0	1	2	3	4
Score range	0-4	5-9	10-14	15-19	20+
Symptom severity	Nil	Mild	Moderate	Moderately severe	Severe
Interventions provided					
Web-based psychoeducation	Yes	Yes	Yes	Yes	Yes
Self-directed Web-based cognitive behavioral therapy	No	No	Yes	Yes	Yes
Face-to-face sessions with a school counsellor to provide counselling and/or referral to external services	No	No	No	Yes	Yes

Table 3. Module overviews.

Title	Content description
What is mental health?	Information about mental health issues that are common among youth and when it might be time to seek help
Feeling on edge	Information on anxiety, how to identify it, potential causes, where to seek help, and practical tips for managing it
Waves of sadness	Information on depression, differences between sadness and depression, potential causes, how and where to seek help, and practical tips to cope
When it's time to tell someone	Information about when to seek help, how to talk to friends and parents, seek help from a general practitioner, and the roles of different health professionals
When a mate needs a hand	Ways to help others including having a private chat, seeking help together, respecting the treatment process, and the importance of looking after yourself
Don't fret, help is here	Offers access to 2 evidence-based, free Web-based cognitive behavioral therapy programs, produced by Australian universities. Young people can select which program they prefer

In the Smooth Sailing service, school counsellor follow-up notifications are initiated if a student is allocated to Steps 3 or 4 or if a student reports having thoughts that they would be better off dead or thoughts of hurting themselves in the past 2 weeks (ie, score >0 on item 9 of the PHQ-9). School counsellors conduct the follow-up sessions during school time but after completion of the researcher visit. School counsellors follow their normal school protocols and duty of care procedures when attending to the notified students, initiating external referrals and parental contact when necessary.

For the duration of the study period, the Smooth Sailing service sends a fortnightly check-in survey to participants via email or SMS text messaging. This consists of the PHQ-2 and GAD-2 and is designed to test participants' engagement and adherence to monitoring. These are ultrabrief versions of the longer counterparts and measure the presence of depression and anxiety symptoms in the past 2 weeks, with responses scored the same as the full scale [42]. Total scores range from 0 to 12. Users are reminded to use the Smooth Sailing program through email and SMS text messaging notifications that are sent fortnightly, on alternate weeks to the check-in survey. Every 6 weeks, a step assessment is completed which reallocates the level of care, depending on participants' results. In accordance with the Australian Clinical Practice Guidelines for the Treatment of Depression in Adolescents and Young Adults [23], if a participant has not responded (ie, symptoms remain elevated or have worsened) within 6 weeks, they are stepped up to the next level of care. The combined step allocation accounts for the participants' previous step, their current step, and whether there has been any change (See Table 4 for stepping decisions). Given that the service is in its infancy and its effectiveness is yet to be established, there is no stepping down in the current model. As such, no care is removed from participants, and they are instructed to continue using the service as advised by their updated personalized dashboard. Participants who show improvements can continue to use the Web-based CBT programs and all other modules as they wish.

The data collected by the service are stored securely on the Black Dog Institute research platform, which is supported by the UNSW servers. The school counsellor receives real-time notifications for students who require follow-up via a secure, purpose-built Web-based portal. The school counsellor can log into the portal to identify the students requiring follow-up and utilize their usual duty of care and referral pathways to treat students. Using the portal functions, school counsellors can notify the research team that follow-ups have been completed. There is also the option for counsellors to provide case notes, detailing treatment action and any adverse events. School counsellors are also able to view these participants' use of the Smooth Sailing program, including module completion. School counsellors receive an automated email reminder to check the portal after every assessment point. Researchers also receive email notifications for any student who triggers a notification. No identifiable information is contained in these emails.

Control

This is a school-as-usual condition. As such, students can access the school counsellor as needed and are permitted to participate in any mental health education or activities initiated by the school or external mental health professionals. Schools allocated to the control condition are placed on a wait list to receive the Smooth Sailing service after the trial data collection period has been completed.

Procedure

At baseline, researchers from the Black Dog Institute visit the school to conduct the study during class time. After providing a verbal explanation of the study and confirming consent, students are asked to use any internet device to visit the study website. Instructions for registering and completion of study questionnaires are provided to students on paper. The researchers remain with students to provide assistance and answer any questions. At 6 weeks and 12 weeks, researchers revisit the schools and repeat this process. Students who are unable to access the website at 6 weeks or 12 weeks because of technical issues are given a paper copy of the questionnaires to complete. If these students are the intervention arm, researchers use a paper scoring method to calculate participants' step at these timepoints and inform the school counsellor verbally if the student requires follow-up. The research team then enters these responses into the system on return to the office resulting in the immediate update of the Web-based portal. Researchers meet with the school counsellor after each visit to review the students who require follow-up and to ensure that counsellors feel adequately supported to conduct these. The researchers contact the school counsellors 2 working days after each visit to review student follow-ups and monitor any adverse events.

XSL•FO

RenderX

Table 4. Stepping decisions at 6 weeks and 12 weeks.

Current step (symptom severity: score range)	Previous step				
	0	1	2	3	4
0 (Minimal: 0-4)	0	1	2	3	4
1 (Mild: 5-9)	1	1	2	3	4
2 (Moderate: 10-14)	2	2	3	3	4
3 (Moderately severe: 15-19)	3	3	3	4	4
4 (Severe: 20+)	4	4	4	4	4

Sample Size

The target sample size for this trial is 1600. This calculation is based on detecting an effect size of 0.20, which is similar to that obtained in previous school-based depression intervention research [9]. The statistical power level is set at 0.8, alpha=.05 (2-tailed), whereas a correlation of 0.5 is assumed between the scores at baseline and endpoint. A design effect is calculated assuming an intraclass correlation of 0.02 and an average school sample of 300 students to allow for possible clustering effects. This estimate was derived from previous Australian school–based studies [9]. The estimated sample size also accommodates for a 20% attrition rate based on previous school-based trials [43]. To achieve the desired sample size and ensure representation from a range of different types of schools, the target is set at 16 schools with approximately 100 students participating per school.

Outcome Measures

The administration schedule for the outcome measures (Multimedia Appendix 1) is outlined in Table 5.

Demographics

At baseline, students provide their name, email, mobile phone number (optional), gender, and date of birth. Students are asked to report their current employment status (answered part-time, casual, or nil) and whether they identify as Aboriginal and or Torres Strait Islander (answered yes, no, I'd rather not say) or Lesbian, Gay, Bisexual, Trans or Intersex (answered yes, no, I'd rather not say).

Mental Health History

At baseline, participants are asked to report whether they have ever known someone with a mental illness (answered yes, no, I'm not sure); cared for someone with a mental illness (answered yes, no, I'm not sure); or if they have previously had a session with the school counsellor at their school (answered yes, no, I'd rather not say). Participants are also asked if they have ever used the internet to find information about a mental health problem (answered yes, no, I'm not sure) and whether they think doing a Web-based program will help their mental health (answered yes, no, I'd rather not say). At baseline, 6 weeks, and 12 weeks, participants are also asked if they have been diagnosed with a mental health problem or mental illness, if they have received any treatment for a mental health problem or mental illness, or if they have taken any prescribed medication (eg, antidepressants) for a mental health problem or mental illness (answered yes, no, I'd rather not say).

Primary Outcome Measure

General Help-Seeking Questionnaire

The General Help-Seeking Questionnaire (GHSQ) is used to measure help-seeking intentions. Participants are asked to rate how likely they are to seek help from 13 different sources (friend, partner, parent, other family member, teacher, other adult, school counsellor, doctor/general practitioner, mental health professional, telephone helpline, mental health website, a Web-based mental health program, and other internet activity) if faced with a mental health problem, with each item answered on a 5-point scale ranging from extremely unlikely to extremely likely [27]. Participants can also indicate if they would not seek help from anyone at all or if they would seek help from someone not listed. The GHSQ has demonstrated satisfactory psychometric properties for measuring the likelihood of seeking help among adolescents [27]. In this study, a *total help-seeking intentions score* is calculated for each participant consisting of their responses to the 13 sources, with total possible scores ranging from 14 to 70.

Table 5. Outcome measures administration schedule.

Assessment	Construct	Baseline	6 weeks	12 weeks
Primary outcome				
General Help-Seeking Questionnaire	Intentions to seek help	Yes	No	Yes
Secondary outcomes				
Actual Help-Seeking Questionnaire	Actual help-seeking behavior	Yes	Yes	Yes
Generalized Anxiety Disorder Questionnaire	Generalized anxiety	Yes	Yes	Yes
Center for Epidemiologic Studies Depression Scale—Child Version	Depressive symptoms	Yes	Yes	Yes
Functioning	Functioning	Yes	Yes	Yes
Distress Questionnaire-5	Psychological distress	Yes	Yes	Yes
Barriers to Adolescents Seeking Help-Brief	Barriers to seeking help	Yes	No	Yes
Mental Health Literacy and Stigma Scale	Mental health literacy and stigma	Yes	No	Yes
Tertiary outcome				
Satisfaction questionnaire[a]	Service adherence and satisfaction	No	No	Yes

[a]Delivered to Intervention schools only.

Secondary Outcome Measures

Actual Help-Seeking Questionnaire

The Actual Help-Seeking Questionnaire is used to assess help-seeking behavior [44]. Using the same sources presented in the GHSQ, participants are asked to report whether they have sought help for a mental health problem from these sources in the past 6 weeks (answered yes or no). Participants can also indicate if they have sought help from someone not listed. An additional question asks participants whether they have needed support for their mental health but did not seek help. The use of this questionnaire in youth has been well supported [30]. In this study, participants' responses at baseline and 12 weeks posttest are categorized into a new variable named *help-seeking from adults*. Participants who have sought help from any adult source are categorized as yes, with those who have not categorized as no. This variable is then used as a secondary outcome for analyses, alongside the number of participants who report not seeking help at all despite having a need for mental health support.

Generalized Anxiety Disorder Questionnaire

Generalized anxiety is measured using the GAD-7 [22], which is a brief measure of anxiety consisting of 7 items that are rated on a scale from not at all (0) to nearly every day (4). Total scores can be classified as nil to mild (5-9), moderate (10-14), or moderately severe to severe (>15). A score of 10 and above is recommended as a reasonable cutoff point for identifying cases of clinical anxiety [22]. The GAD-7 has been found to have good test-retest reliability and strong criterion validity [22].

Center for Epidemiologic Studies Depression Scale—Child Version

The Center for Epidemiologic Studies Depression Scale (CES-D) was originally developed for measuring depression in an adult population [45] and was later modified into the child version, the Center for Epidemiologic Studies Depression Scale – Child Version (CES-DC) [46]. It is comprised 20 items, rated from not at all (0) to a lot (3), with 4 items reverse-scored. The total possible scores range from 0 to 60, with higher scores indicative of increased levels of depression. The scale has been found to have good psychometric properties among adolescents [47]. The most commonly used cutoff score for the CES-D is 16 and above, which indicates the individual should be further screened for risk of depression [48].

Distress Questionnaire-5

The Distress Questionnaire-5 (DQ5) is a brief, valid screener used for identifying those at risk of psychological distress in the general population [49]. The DQ5 consists of 5 items rated from never (1) to always (5). Total scores range from 5 to 25 and higher scores indicate greater psychological distress. It has been suggested that the DQ5 may be superior to the K6 and K10 in identifying increased risk for common mental disorders, with a screening cutoff point of 11 indicating the likely presence of a mental health condition [49].

Functioning

Taken at baseline and posttest, this is a 1-item question in which students are asked to rate how difficult their symptoms of anxiety and depression have made their daily life and relationships. Participants answer using a 4-point Likert scale ranging from not at all difficult (0) to very difficult (4).

Barriers to Adolescents Seeking Help

The Barriers to Adolescents Seeking Help-Brief [50] is a shortened (11 items) version of the original questionnaire, measuring barriers to help-seeking behavior among adolescents. Participants rate their level of agreement with a list of statements about help-seeking from strongly disagree (1) to strongly agree (6). Total scores range from 11 to 66, with higher mean scores reflecting greater resistance to seeking professional help. The scale has been found to have good test-retest reliability and validity in a population of high school students [50].

Mental Health Literacy and Stigma Scale

This scale measures a person's confidence about help-seeking for mental health and stigmatizing attitudes toward mental illness. It has demonstrated good internal and test-retest reliability and validity [51]. A shortened version of the questionnaire, consisting of 4 items assessing confidence seeking help and 9 items assessing stigmatizing attitudes of mental illness, has been selected for use in this study. Item responses are rated on a 5-point scale from strongly disagree (1) to strongly agree (5) with 9 items reverse-scored. Total scores for this scale range from 13 to 65, with higher scores indicating greater confidence in seeking help for mental health problems and lower levels of stigma.

Tertiary Outcome Measure

Service Adherence and Satisfaction

Service adherence is measured by the number of modules accessed and the number of check-ins completed by the intervention participants. A total of 3 questionnaires assess service satisfaction. The first questionnaire assesses 3 main domains: enjoyment and ease of use, usefulness, and degree of comfort with service requirements. This questionnaire requires participants to rate whether they agree or disagree with 11 statements about the service, for example, *I enjoyed using Smooth Sailing*. The second questionnaire asks participants to report on 18 service use barriers across 3 key domains: technical, personal, or intervention-specific (answered yes or no), for example, *My internet connection didn't work*. The third questionnaire examines participants' use of the websites and CBT programs suggested by the service (answered yes, no, or I can't remember). Participants are also asked to report if they have been contacted by the school counsellor during the study period and if they felt comfortable when this occurred. Participants are also asked to provide an overall rating of helpfulness of the service (answered on a Likert scale of 1 to 5 from extremely unhelpful to extremely helpful) and their suggestions for improvements or any other comments (2 free response options).

Assessment of Safety

Data Monitoring

A data safety monitoring board is not utilized because the study targets a nonclinical population, blinding is not used, and the service provides a real-time monitoring and notification system of participants' thoughts of death and self-harm. An Outcomes Advisory Group has been established to provide specific monitoring, governance, and reporting of adverse events and trial safety. This consists of the CI, chief scientist of the Black Dog Institute, a medical expert in child and adolescent psychiatry, and a clinical psychologist. This team meets monthly, and on an as-needed basis, to monitor the safety of the trial. During baseline, 6-week, and 12-week assessments, the research team notifies the Outcomes Advisory Group by email every fortnight of the total number of students participating and the notifications received for each school. This group is also responsible for responding to any complaints.

Harms

Throughout the study period, the duty of care for all students remains with their participating school. A trained mental health researcher from the Black Dog Institute is present at all school visits, alongside a school staff member. The school counsellor and a private breakout space is made available for any students who become distressed during the assessments. In the control condition, all students are provided with an information sheet of help-seeking resources and services at baseline, 6-week, and 12-week assessment. In the intervention condition, the Smooth Sailing service includes a question assessing thoughts of dying or harming oneself. This is part of the PHQ-9 screening measure for depression and aims to provide an objective measure of symptom severity if completed truthfully. However, further assessment from a trained mental health professional is required for the extent of an individual's suicide risk to be fully determined. If a student reports the presence of these thoughts during the assessments, information about suicide support services is immediately provided by the service through a pop-up. The school counsellors are also notified using the Web-based portal and email system. The school counsellors are also notified when any participant in the intervention condition reports severe depression or anxiety symptoms, or there is risk of significant harm as identified by the school. The counsellors are notified to conduct the follow-up assessments to determine participants' risk, using normal school protocols. A list of local services and resources is provided to the school counsellors to assist them in following up with students. The child and adolescent psychiatrist in the Outcomes Advisory Group is also available to provide advice and assistance to the school counsellors throughout the study period. School counsellors are asked to report on follow-ups to the research team via phone call or email with a record-keeping option on the school counsellor Web portal. The research team also inquire about adverse events at each of the school visits and report on these in their fortnightly reports to the Outcomes Advisory Group. At the final study visit, all students are provided with a list of mental health resources and services information.

Data Handling, Storage, and Access

The study is hosted on the servers of the Black Dog Institute Research Platform, Faculty of Medicine, UNSW. These servers are encrypted with data backups occurring daily. This platform and its associated data are only accessible to authorized and approved personnel. When registering, participants create password-protected accounts and the platform allocates a unique identification number. For analyses, data are deidentified by removing names, email address, and mobile phone numbers. Students who actively withdraw will have their data removed and a withdrawal confirmation email will be sent from the research team. The platform retains data from the students who are lost to follow-up. All data will be stored securely for a minimum of 15 years.

Auditing

Researchers follow step-by-step guidelines for each study visit to ensure consistency across different trial sites and research team members. A decision-making guide for trial conduct is provided to all research staff. School packages are mailed to

XSL•FO

RenderX

schools ahead of time, which included trial documents, such as an explanatory paragraph for the school newsletter, an agenda for the baseline visit, and a list of local support services, to facilitate familiarity with the consent process and study procedure. Instructional documents for intervention school counsellors are provided to ensure adherence with the study protocol and to guide their use of the Web-based portal. All research staff debrief with the CI or research manager after each school visit. Fortnightly research team meetings are held to discuss and audit the conduct of the trial. The research team also have frequent contact with the school counsellors to assess any adverse events.

Analysis

Data are collected using the Black Dog Institute electronic health platform. Data will be downloaded into Microsoft Excel and exported to SPSS Version 22.0 (SPSS Inc) for analysis. Primary analyses will be undertaken on an intention to treat basis, including all participants randomized, regardless of treatment received. Effectiveness of Smooth Sailing will be established by a change in the GHSQ scores between baseline and 12 weeks using a mixed-effects model repeated-measures analysis. The school will be included in analyses as a random effect to evaluate and accommodate clustering effects. In analyses of scaled secondary variables, methods comparable with those of the primary analysis will be used. An analysis of covariance will be conducted for the secondary outcomes of anxiety, depression, distress, functioning, mental health literacy, and barriers to help-seeking, controlling for baseline scores. To determine the effect of the intervention on help-seeking, 2 binary logistic regression analyses will be used. In model 1, the dependent variable will be help-seeking from adults at 12 weeks. The baseline result will be entered at step 1 in the model, with the study arm entered at step 2. In model 2, the dependent variable will be not seeking help at all at 12 weeks despite a need. The baseline result will be entered at step 1, with the study arm entered at step 2.

Dissemination Policy

A summary of the results will be emailed to all participants who request this on their consent form. The results summary will also be published on the Black Dog Institute website. School reports, consisting of the mean aggregate scores for the measures, will be prepared and shared with the school at completion of the data collection period. The primary and secondary outcomes analysis will be prepared for academic publication and presentations. A final report will also be submitted to the funding body. In all documents, participants will not be individually identifiable with data presented at the aggregate level.

Results

The results for this trial are currently under analysis. Ethics approval was provided by UNSW HREC on November 24, 2017, and by the NSW Department of Education SERAP on January 10, 2018. An Expression of Interest for the trial was advertised through the NSW School-Link Initiative (School-Link) newsletter and email. We were subsequently contacted by a school representative (eg, School Counsellor,

Head of Well-Being) from 88 different schools. Throughout January and February 2018, 22 secondary schools across NSW agreed to participate in the trial. These schools were randomized into 2 conditions, with 10 schools assigned to receive the intervention and 12 assigned to the control condition. The first baseline assessment occurred on February 22, 2018, with the 12-week endpoint assessments completed on Friday, June 29, 2018. Control schools are currently receiving the service, due for completion by June 30, 2019.

Discussion

This study protocol outlines the Smooth Sailing trial, which aims to investigate the effectiveness of a Web-based mental health service for secondary school students. Using a randomized controlled design, this study evaluates the efficacy of the Smooth Sailing service for improving a range of mental health outcomes among students compared with school-as-usual. The service was designed to increase adolescents' help-seeking intentions for mental health problems, as well as to improve help-seeking behaviors and reduce symptoms of anxiety, depression, and distress. This is critically important to this age group as mental illness first emerges during adolescence and help-seeking from professionals is low [2,3]. Most notably, this study represents the first attempt at delivering a school-based mental health service that incorporates the principles of stepped care with Web-based screening and intervention alongside direct face-to-face follow-up.

During the development of the service, school counsellors, parents, and general practitioners (GPs) outlined various concerns related to the accuracy, effectiveness, and suitability of the service for delivery in the school setting [18-20]. Some parents questioned the effectiveness of a Web-based service for determining the level of care their child required [20] and whether the service would increase stigma among students. Both parents and school counsellors were concerned that students may not openly disclose their symptoms to a Web-based service [18]. Parents, counsellors, and GPs [19] were also concerned about teens' adherence and engagement with self-directed Web-based CBT programs. This study will be able to determine these outcomes. This trial will also evaluate the benefits of using a Web-based service to initiate school counsellor referrals for symptomatic students compared with traditional referral processes, and the capacity of school counsellors to manage these. This study will also be able to examine the effectiveness of school counselling services in Australia, an area in which effectiveness data are lacking.

The results of this trial will confirm whether an integrated service model that combines screening, triage, intervention, and monitoring is acceptable and effective in the school environment on a large scale. A multicenter trial of a suicide prevention program in Europe revealed that screening alone was not effective for improving outcomes [52]. The authors argued that concurrent activities are needed, including mental health literacy and broader awareness among the student year group. Furthermore, adolescents were more likely to engage in programs and services that acknowledge their autonomy, rather than being adult or teacher driven. These findings have

important implications for this trial, suggesting that the Smooth Sailing service is likely to influence students' mental health outcomes. Overall, the trial will determine whether universal screening with prompted access to Web-based psychoeducation, self-directed CBT and school counselling services will increase help-seeking among secondary school students.

Although the co-design process and pilot study indicated that Smooth Sailing was acceptable and feasible to deliver, practical considerations for the larger trial were outlined. Students revealed that they would disapprove of this type of service being made compulsory at school and instead felt engagement and interest in the service could be increased if an opt-out parental consent process was implemented. This was supported by school staff as they felt it respected students' increasing autonomy and addressed privacy concerns, especially among students who were more reluctant to participate in mental health initiatives or were experiencing mental health problems. To increase student engagement in the service content, students and staff requested the use of multimodal reminders. For example, school staff could be encouraged to remind students to use Smooth Sailing for a certain amount of time per week, alongside SMS text messaging and email reminders for students at fortnightly intervals. It was also suggested that the Web-based psychoeducation content be modified and reduced. School

counsellors also felt that a Web-based portal would increase their efficiency by improving their ability to keep track of follow-ups and better understand the mental health care needs of those students who trigger a notification. All of these changes have been made and instigated in the current protocol.

Certain limitations may affect the findings in this study. Given that the intervention condition involves being followed up by the school counsellor, it may be that students from the intervention schools are less likely to consent than students from the control schools. Furthermore, the increased contact with students in the intervention arm through the service check-ins may impact retention. If found to be effective, the researchers will need to carefully consider the range of implementation barriers that challenge school-based mental health initiatives, including maintaining service fidelity and access beyond trial completion [12]. Major strengths of the study include the RCT design, the novelty of the intervention, and the use of stepped care in the school context. This study will be able to determine the usefulness and effectiveness of a stepped care model, confirming whether a 5-step model is justified. The trial findings will provide vital information about whether this new and innovative model of care for youth mental health is effective for when delivered in the school setting.

Acknowledgments

Authors' Contributions

BOD and HC conceived the study. BOD prepared the protocol and initiated the trial. BOD, CK, MSK, MRA, MA, and BP contributed to the coordination of the trial. AM conceived and conducted randomization. All authors contributed to refinement of the study protocol. All authors read and approved the final manuscript.

Conflicts of Interest

BOD is a section editor for JMIR Mental Health.

Multimedia Appendix 1

Outcome measures schedule.

[PDF File (Adobe PDF File), 350KB - resprot_v8i5e12892_app1.pdf]

References

1. Kessler R, Amminger G, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007 Jul;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
2. Lawrence D, Johnson S, Hafekost J, Boterhoven De Haan K, Sawyer M, Ainley J, et al. The Mental Health of Children and Adolescents. Report on the Second Australian Child and Adolescent Survey of Mental Health and Wellbeing. Canberra, Australia: Department of Health; 2015.
3. Burgess P, Pirkis J, Slade T, Johnston A, Meadows G, Gunn J. Service use for mental health problems: findings from the 2007 National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2009 Jul;43(7):615-623. [doi: 10.1080/00048670902970858] [Medline: 19530018]
4. Leach L, Butterworth P. The effect of early onset common mental disorders on educational attainment in Australia. Psychiatry Res 2012 Aug 30;199(1):51-57. [doi: 10.1016/j.psychres.2012.03.040] [Medline: 22507527]
5. Kessler R, Foster C, Saunders W, Stang P. Social consequences of psychiatric disorders, I: educational attainment. Am J Psychiatry 1995 Jul;152(7):1026-1032. [doi: 10.1176/ajp.152.7.1026] [Medline: 7793438]

6.   Scott J, Fowler D, McGorry P, Birchwood M, Killackey E, Christensen H, et al. Adolescents and young adults who are not in employment, education, or training. Br Med J 2013 Sep 18;347:f5270. [doi: 10.1136/bmj.f5270] [Medline: 24048295]

7.   O'Neil K, Conner B, Kendall P. Internalizing disorders and substance use disorders in youth: comorbidity, risk, temporal order, and implications for intervention. Clin Psychol Rev 2011 Feb;31(1):104-112. [doi: 10.1016/j.cpr.2010.08.002] [Medline: 20817371]

8.   Perry Y, Petrie K, Buckley H, Cavanagh L, Clarke D, Winslade M, et al. Effects of a classroom-based educational resource on adolescent mental health literacy: a cluster randomized controlled trial. J Adolesc 2014 Oct;37(7):1143-1151. [doi: 10.1016/j.adolescence.2014.08.001] [Medline: 25151646]

9.   Calear A, Christensen H, Mackinnon A, Griffiths K, O'Kearney R. The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009 Dec;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]

10. Perry Y, Werner-Seidler A, Calear A, Mackinnon A, King C, Scott J, et al. Preventing depression in final year secondary students: School-based randomized controlled trial. J Med Internet Res 2017 Dec 2;19(11):e369. [doi: 10.2196/jmir.8241] [Medline: 29097357]

11. Wong N, Kady L, Mewton L, Sunderland M, Andrews G. Preventing anxiety and depression in adolescents: a randomised controlled trial of two school based Internet-delivered cognitive behavioural therapy programmes. Internet Interv 2014 Apr;1(2):90-94. [doi: 10.1016/j.invent.2014.05.004]

12. Langley A, Nadeem E, Kataoka S, Stein B, Jaycox L. Evidence-based mental health programs in schools: barriers and facilitators of successful implementation. School Ment Health 2010 Sep;2(3):105-113 [FREE Full text] [doi: 10.1007/s12310-010-9038-1] [Medline: 20694034]

13. Briesch A, Chafouleas S, Chaffee R. Analysis of state-level guidance regarding school-based, universal screening for social, emotional, and behavioral risk. School Ment Health 2017 Nov 4;10(2):147-162. [doi: 10.1007/s12310-017-9232-5]

14. van Straten A, Hill J, Richards D, Cuijpers P. Stepped care treatment delivery for depression: a systematic review and meta-analysis. Psychol Med 2015 Jan;45(2):231-246. [doi: 10.1017/S0033291714000701] [Medline: 25065653]

15. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17. [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]

16. McGorry P, Goldstone S, Parker A, Rickwood D, Hickie I. Cultures for mental health care of young people: an Australian blueprint for reform. Lancet Psychiatry 2014 Dec;1(7):559-568. [doi: 10.1016/S2215-0366(14)00082-0] [Medline: 26361315]

17. Hollis C, Morriss R, Martin J, Amani S, Cotton R, Denis M, et al. Technological innovations in mental healthcare: harnessing the digital revolution. Br J Psychiatry 2015 Apr;206(4):263-265. [doi: 10.1192/bjp.bp.113.142612] [Medline: 25833865]

18. O'Dea B, King C, Subotic-Kerry M, O'Moore K, Christensen H. School counselors' perspectives of a web-based stepped care mental health service for schools: cross-sectional online survey. JMIR Ment Health 2017 Nov 20;4(4):e55 [FREE Full text] [doi: 10.2196/mental.8369] [Medline: 29158207]

19. Subotic-Kerry M, King C, O'Moore K, Achilles M, O'Dea B. General practitioners' attitudes toward a web-based mental health service for adolescents: implications for service design and delivery. JMIR Hum Factors 2018 Mar 23;5(1):e12 [FREE Full text] [doi: 10.2196/humanfactors.8913] [Medline: 29572203]

20. O'Dea B, Leach C, Achilles M, King C, Subotic-Kerry M, O'Moore K. Parental attitudes towards an online, school-based, mental health service: implications for service design and delivery. Advance Ment Health 2018 Aug 30:1-15. [doi: 10.1080/18387357.2018.1514269]

21. Kroenke K, Spitzer R, Williams J. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]

22. Spitzer R, Kroenke K, Williams J, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

23. McDermott B, Baigent M, Chanen A, Fraser L, Graetz B, Hayman N. Clinical Practice Guidelines: Depression in Adolescents and Young Adults. Melbourne, Australia: Beyond Blue; 2010.

24. Husky M, Kaplan A, McGuire L, Flynn L, Chrostowski C, Olfson M. Identifying adolescents at risk through voluntary school-based mental health screening. J Adolesc 2011 Jun;34(3):505-511. [doi: 10.1016/j.adolescence.2010.05.018] [Medline: 20561672]

25. Green J, McLaughlin K, Alegría M, Costello E, Gruber M, Hoagwood K, et al. School mental health resources and adolescent mental health service use. J Am Acad Child Adolesc Psychiatry 2013 May;52(5):501-510 [FREE Full text] [doi: 10.1016/j.jaac.2013.03.002] [Medline: 23622851]

26. Carlton P, Deane F. Impact of attitudes and suicidal ideation on adolescents' intentions to seek professional psychological help. J Adolesc 2000 Feb;23(1):35-45. [doi: 10.1006/jado.1999.0299] [Medline: 10700370]

27. Wilson C, Deane F, Ciarrochi J, Rickwood D. Measuring help seeking intentions: properties of the General Help seeking Questionnaire. Can J Counsel 2005;39(1):15-28 [FREE Full text]

28. Kelly A, Achter J. Self-concealment and attitudes toward counseling in university students. J Couns Psychol 1995;42(1):40-46. [doi: 10.1037/0022-0167.42.1.40]

29. Deane F, Todd D. Attitudes and intentions to seek professional psychological help for personal problems or suicidal thinking. J College Stud Psychother 1996 May 31;10(4):45-59. [doi: 10.1300/J035v10n04_06]

30. Rickwood D, Deane F, Wilson C, Ciarrochi J. Young people's help-seeking for mental health problems. Advance Ment Health 2014 Dec 17;4(3):218-251. [doi: 10.5172/jamh.4.3.218]

31. Chan A, Tetzlaff J, Gøtzsche PC, Altman D, Mann H, Berlin J, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. Br Med J 2013 Jan 8;346:e7586 [FREE Full text] [doi: 10.1136/bmj.e7586] [Medline: 23303884]

32. Lewis J. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. Stat Med 1999 Aug 15;18(15):1903-1942. [doi: 10.1002/(SICI)1097-0258(19990815)18:15<1903::AID-SIM188>3.0.CO;2-F] [Medline: 10440877]

33. Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research. London: Wiley; 2000.

34. Taves D. The use of minimization in clinical trials. Contemp Clin Trials 2010 Mar;31(2):180-184. [doi: 10.1016/j.cct.2009.12.005] [Medline: 20060500]

35. Altman D, Bland J. Treatment allocation by minimisation. Br Med J 2005 Apr 9;330(7495):843 [FREE Full text] [doi: 10.1136/bmj.330.7495.843] [Medline: 15817555]

36. Ryan P. EconPapers. 2018. RCT_MINIM: Stata module to assign treatments to subjects in a controlled trial URL: https://econpapers.repec.org/software/bocbocode/s457029.htm [accessed 2019-04-03] [WebCite Cache ID 77NCgRztx]

37. Calear A, Christensen H, Mackinnon A, Griffiths K, O'Kearney R. The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009 Dec;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]

38. Spence S, Holmes J, March S, Lipp O. The feasibility and outcome of clinic plus internet delivery of cognitive-behavior therapy for childhood anxiety. J Consult Clin Psychol 2006 Jun;74(3):614-621. [doi: 10.1037/0022-006X.74.3.614] [Medline: 16822117]

39. Spence S, Donovan C, March S, Gamble A, Anderson R, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011 Oct;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]

40. MoodGym. URL: https://moodgym.com.au/ [accessed 2019-04-03] [WebCite Cache ID 77N3bOWfM]

41. BRAVE-Online. URL: http://www.brave-online.com/ [accessed 2019-04-03] [WebCite Cache ID 77Omj3Rvv]

42. Löwe B, Wahl I, Rose M, Spitzer C, Glaesmer H, Wingenfeld K, et al. A 4-item measure of depression and anxiety: validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. J Affect Disord 2010 Apr;122(1-2):86-95. [doi: 10.1016/j.jad.2009.06.019] [Medline: 19616305]

43. Christensen H, Griffiths K, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]

44. Rickwood D, Braithwaite V. Social-psychological factors affecting help-seeking for emotional problems. Soc Sci Med 1994 Aug;39(4):563-572. [Medline: 7973856]

45. Radloff L. The CES-D Scale: a self-report depression scale for research in the general population. Appl Psychol Meas 2016 Jul 26;1(3):385-401. [doi: 10.1177/014662167700100306]

46. Radloff LS. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. J Youth Adolesc 1991 Apr;20(2):149-166. [doi: 10.1007/BF01537606] [Medline: 24265004]

47. Fendrich M, Weissman M, Warner V. Screening for depressive disorder in children and adolescents: validating the Center for Epidemiologic Studies Depression Scale for Children. Am J Epidemiol 1990 Mar;131(3):538-551. [doi: 10.1093/oxfordjournals.aje.a115529] [Medline: 2301363]

48. Vilagut G, Forero C, Barbaglia G, Alonso J. Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): a systematic review with meta-analysis. PLoS One 2016;11(5):e0155431 [FREE Full text] [doi: 10.1371/journal.pone.0155431] [Medline: 27182821]

49. Batterham P, Sunderland M, Carragher N, Calear A, Mackinnon A, Slade T. The Distress Questionnaire-5: Population screener for psychological distress was more accurate than the K6/K10. J Clin Epidemiol 2016 Mar;71:35-42. [doi: 10.1016/j.jclinepi.2015.10.005] [Medline: 26464194]

50. Kuhl J, Jarkon-Horlick L, Morrissey R. Measuring barriers to help-seeking behavior in adolescents. J Youth Adolesc 1997 Dec;26(6):637-650. [doi: 10.1023/A:1022367807715]

51. O'Connor M, Casey L. The Mental Health Literacy Scale (MHLS): a new scale-based measure of mental health literacy. Psychiatry Res 2015 Sep 30;229(1-2):511-516. [doi: 10.1016/j.psychres.2015.05.064] [Medline: 26228163]

52. Wasserman D, Hoven C, Wasserman C, Wall M, Eisenberg R, Hadlaczky G, et al. School-based suicide prevention programmes: the SEYLE cluster-randomised, controlled trial. Lancet 2015 Apr 18;385(9977):1536-1544. [doi: 10.1016/S0140-6736(14)61213-7] [Medline: 25579833]

Abbreviations

CBT: Cognitive Behavioral Therapy
CI: Chief Investigator
DQ5: Distress Questionnaire-5

GAD-7: Generalized Anxiety Disorder Scale-7
GHSQ: General Help-Seeking Questionnaire
GPs: General Practitioners
HREC: Human Research Ethics Committee
NSW: New South Wales
PHQ-9: Patient Health Questionnaire-9
PICF: Participant Information and Consent Form
RCT: Randomized Controlled Trial
SCS: Sydney Catholic Schools
SERAP: State Education Research Applications Process
SMS: Short Message Service
UNSW: University of New South Wales

XSL•FO
RenderX

Protocol

Personalized Text Messages and Automated Calls for Improving Vaccine Coverage Among Children in Pakistan: Protocol for a Community-Based Cluster Randomized Clinical Trial

Abdul Momin Kazi[1,2], MBBS, MPH; Nazia Ahsan[1], BSc, MSc; Ayub Khan[1], BComm; Saima Jamal[1], BSc, MA; Hussain Kalimuddin[1], BSc, MCS; Naveera Ghulamhussain[1], BSc, MSc; Zabin Wajidali[1], BSc, MSc; Abdul Muqeet[3], BS; Fabiha Zaidi[3], BE; Meraj Subzlani[3], BComm; William McKellin[4], PhD; Asad Ali[1], MPH, MBBS; Jean-Paul Collet[2,5], MD, PhD

[1]Department of Pediatrics and Child Health, Aga Khan University, Karachi, Pakistan

[2]Department of Pediatrics, University of British Columbia and BC Children's Hospital Research Institute, Vancouver, BC, Canada

[3]Digital Health Resource Center, Aga Khan Development Network, Karachi, Pakistan

[4]Department of Sociology, University of British Columbia, Vancouver, BC, Canada

[5]Beijing Tiantan Hospital and Advanced Innovation Center for Human Brain Protection, Capital Medical University, Beijing, China

Corresponding Author:
Abdul Momin Kazi, MBBS, MPH
Department of Pediatrics and Child Health
Aga Khan University
Stadium Road
Karachi,
Pakistan
Phone: 92 2134864232
Email: momin.kazi@aku.edu

Abstract

Background: A major reason for poor childhood vaccine coverage in developing countries is the lack of awareness among parents and caregivers regarding the need for immunization and the importance of completing the entire series of vaccines. Short message service (SMS)–based interventions have been quite effective in different programs such as smoking cessation, treatment adherence, health care scheduled appointment attendance, antenatal care attendance, and compliance to immunization. However, there are limited data from low- and middle-income countries on the role of SMS and automated call–based messages and interventions to improve routine immunization (RI) coverage.

Objective: The primary objective of this study is to evaluate whether automated mobile phone–based personalized messages (SMS or automated call) can improve RI uptake at 6, 10, and 14 weeks of age per the expanded program immunization schedule, compared with a usual care control group. Secondary objectives include assessing the effects of different types of automated SMS text or calls on RI coverage at 20 weeks of age.

Methods: This is a mixed methods study using a clustered randomized controlled trial with 4 intervention arms and 1 control arm, augmented by qualitative interviews for personalizing the message. The study is being conducted in Pakistan (an urban site in Karachi and a rural site Matiari). In Karachi, 250 administrative structures are taken as 1 cluster, whereas in Matiari, a catchment area of 4 Lady Health Workers is considered as 1 cluster. The intervention targets families to receive weekly 1-way or 2-way (interactive) personalized automated SMS or automated phone call messages regarding vaccination. Possible barriers to vaccination are assessed in each family at the time of inclusion to determine the type of personalized messages that should be sent to the family to increase the chance of a positive response. Finally, in-depth interviews using purposive sampling are conducted before and after the trial to determine the family's vaccination experience and related factors.

Results: All study participants for the cluster randomized trial were enrolled by January 14, 2019. Study exit interviews at 20-weeks follow-up visits will be completed by June 2019.

Conclusions: The results of this study will be useful to understand the respective effects of SMS text messages versus automated phone–based communication to improve RI coverage and timelines. Moreover, information regarding families' perceptions of vaccination and the daily life challenges for timely visits to the vaccine clinic will be used for developing more complex

XSL•FO

RenderX

interventions that use mobile phone messages and possibly other approaches to overcome barriers in the uptake of correct and timely immunization practices.

Introduction

Background

Pakistan is one of the countries with the highest rates of child death in the world [1]. It ranks 4th in child mortality, with 60% deaths due to vaccine-preventable diseases (VPDs) [2]. Table 1 [18] shows the current schedule of routine immunization (RI) in Pakistan, which is provided by the government free of cost. The immunization coverage in Pakistan is estimated to be 59%, which is still well below the desired level, leading to continued polio transmission, large measles outbreaks, and thousands of deaths from vaccine-preventable illnesses [2]. In addition, Pakistan is a major polio epidemic country and among 3 countries in the world requiring proof of polio vaccination for international travel [3]. Pakistan demographic and health survey in 2017-2018 suggests 88% percent of children had received BCG vaccine due at birth, 86% and 95% had received the first dose of pentavalent and polio vaccine respectively due at the 6th week. Furthermore, 75% and 86% of children had received the third dose of the pentavalent and polio vaccines, respectively, due at 14th week and measles vaccination was 73%, which is due at 9 months. However, these rates are at 1 year of age and much higher than vaccination coverage rate at scheduled time and among conflict hits and displaced populations [4]. Improved RI coverage is recommended as the priority public health strategy to reduce VPDs and eradicate polio in Pakistan and worldwide.

According to immunization coverage surveys, 1 in 5 children is unimmunized [5]. A major reason for poor childhood vaccine coverage is low immunization uptake, when parents are unable to complete the entire series of vaccines in accordance with the scheduled timelines. Some of the reasons include: (1) the family is not in favor of getting their child immunized, (2) low trust in vaccines provided through Expanded Programme on Immunization (EPI) and government health care providers, and (3) caregivers have forgotten their child's next vaccination due date or child's EPI card is misplaced [6]. These barriers may be modified with additional support through education and behavior change strategies. In addition, with more pressing issues of food and shelter, preventive health often takes the back seat, and parents and caregivers forget or ignore the subsequent doses of vaccines for their children. There is an immense need to encourage parents' care seeking and collaboration with the health care providers to improve initial vaccine uptake and the completion of all doses according to the schedule. New innovative and cost-effective techniques are necessary for practical solutions to improve vaccination uptake and coverage.

Mobile phones offer a new medium to provide education and advocate families or caregivers to enable behavior change so as to improve immunization uptake. Mobile phone use has also increased in countries with low RI coverage and a high risk of VPDs. Good examples are Nigeria and Pakistan, where there were around 170 and 140 million mobile phone subscribers, respectively, in 2014 [7,8]. There is also a surge in use of short message service (SMS), with 237.58 billion person-to-person SMS generated in 2011 estimating to around 175 SMS per mobile phone on a monthly basis in Pakistan [9].

Table 1. Childhood immunization schedule for Pakistan.

Disease	Causative agent	Vaccine	Doses	Age of administration
Childhood tuberculosis	Bacteria	Bacillus Calmette-Guerin	1	Soon after birth
Poliomyelitis	Virus	OPV (oral poliovirus vaccine); Inactivated polio vaccine (IPV)	4; 1	OPV0: soon after birth, OPV1: 6 weeks, OPV2: 10 weeks, OPV3: 14 weeks; IPV-1: 14 weeks
Diphtheria, Tetanus, Pertussis, Hepatitis B, Hib pneumonia, and meningitis	Bacteria	Pentavalent vaccine (Diphtheria, tetanus toxoids and pertussis +Hepatitis B + Hib)	3	Penta1: 6 weeks, Penta2: 10 weeks, Penta3: 14 weeks
Pneumonia	Bacteria	Pneumococcal vaccine (PCV)	3	PCVIO 1: 6 weeks, PCVIO 2: 10 weeks, PCVIO 3: 14 weeks
Measles	Virus	Measles	2	Measles1: 9 months, Measles2: 15 months
Diarrhea due to rotavirus	Virus	Rotavirus	2	Rota 1: 6 weeks, Rota 2: 10 weeks

XSL·FO

RenderX

Mobile reminders in the form of phone calls have also proven to be a feasible method of improving RI uptake in resource-limited settings with wide cellphone coverage [10]. A Cochrane review reported that automated telephone communication systems including automated calls are effective in a variety of health care settings such as improving clinic attendance rates, screening, adherence to medications, and laboratory tests [11]. Automated calls were also found to be cost-effective in increasing immunization rates in an urban practice in the United States [12]. A study suggests that mobile phones have wide spread abilities to improve health outcomes in low and middle-income countries (LMICs) by targeting larger populations in a cost-effective manner [13]. Furthermore, SMS-based reminders along with small financial incentives could possibly help in improving RI timelines [14]. Text reminders have also proved to reduce vaccine dropout rate and improve parents' compliance to immunization-scheduled visits [15,16].

There are limited data from LMICs set up on the role of SMS-based interventions for improvement of RI coverage, and conventional 1-way reminder SMS text messages were used by most of the studies as the intervention [13,21,22]. Overall, very few studies compared reminders, educational, and interactive SMS messages related to childhood vaccination uptake [13,21-26]. Although some of the studies have shown some behavior change for improvement in vaccination coverage, more rigorous application of health behavior change model needs to be applied to understand the impact of reminder, educational, and interactive messages on behavior change related to improvement in RI coverage [17]. However, data from developing countries regarding the role of automated calls in improving vaccine coverage are limited. Majority of the studies have focused on SMS-based intervention.

In this study, we will identify the factors that affect adherence to RI coverage among Pakistani families and caregivers and will examine an important public health question—do low cost, automated SMS text messages and calls improve RI coverage in resource-constrained settings such as Pakistan? In this study, we will compare the effectiveness of different types of messages: reminder, educational, and interactive SMS text messages and automated calls for improving RI uptake. Different types of messages will be developed to meet the possible RI barriers identified; these messages will be sent specifically to the participants according to the type of RI barriers they faced for immunization. This information will be used to develop strategies to improve vaccine adherence in Pakistan.

Study Objectives

Primary

Our first objective is to evaluate whether personalized automated SMS text messages, 1-way versus 2-way, can improve on-time visits at 6, 10, and 14 weeks of age for RI as compared with standard care. Our second objective is to evaluate whether personalized automated calls, 1-way versus 2-way, can improve on-time visits at 6, 10, and 14 weeks of age for RI as compared with standard care. Finally, our third objective is to compare the efficacy of automated SMS text messages versus automated calls on increasing vaccination uptake.

Secondary

Our secondary objectives are to learn the perception and attitudes of caregivers regarding childhood vaccination and to find out about factors that might influence mobile phone and SMS text–based interventions for vaccination improvement.

Methods

Target Population

Our target population are caregivers of newborn (NB) infants younger than 14 days who are due for their RI at age 6, 10, and 14 weeks according to the EPI schedule of Pakistan.

Study Goal

Our goal is to identify the barriers faced by families in the uptake of correct immunization practices and to develop adapted messages (SMS or automated calls) to improve vaccine adherence in Pakistan. We will collect information from families regarding perceived sociocultural, technical, and economical barriers that may explain the decrease in vaccine coverage, and possible solutions to overcome these constraints. This information will be used to develop personalized educational messages, reinforced by interactive exchanges (2-way automated SMS and automated calls) with caregivers in the initial 20 weeks of their child's life.

Study Hypothesis

Personalized weekly message in the form of automated SMS text message or automated calls according to barriers for RI and language preferences can improve RI uptake according to the schedule for vaccine due at age 6, 10, and 14 weeks as compared with standard counseling by health care providers at EPI center and outreach visits for RI. Interactive voice recording (IVR) or 2-way auto calls might be preferred way of communication to caregivers as compared with other intervention and control for possible improvement in RI coverage and timelines.

Outcomes

Primary

Primary outcomes include: (1) to see a 10% increase in RI through personalized automated mobile phone–based communication (SMS or automated call) at 6, 10, and 14 weeks of age according to the EPI schedule versus standard care; (2) to see a 10% increase in RI within 1 week of the original timeline at 6, 10, and 14 weeks versus standard care; (3) to compare coverage rates of personalized 1-way versus 2-way SMS text messages on improvement in RI at 20 weeks of age; (4) to compare coverage rates of personalized 1-way versus 2-way automated calls on improvement in RI at 20 weeks of age; and (5) to compare coverage rates of personalized SMS text messages and automated calls on improvement in RI at 20 weeks of age.

Secondary

Secondary outcomes include: (1) to understand the perceptions and barriers of caregivers regarding immunization and (2) to understand the perception of caregivers related to personalized mobile phone–based SMS text messages and automated health messages for vaccination improvement.

Methodology

Study Design

The study design is a mixed method clustered randomized controlled trial (RCT) augmented by qualitative interviews (Figure 1). The cluster trial will be used to assess the respective impacts of mobile phone and SMS-based communications on RI coverage rates among children at 20 weeks of age. Initially, structured interviews will be conducted at baseline to identify barriers to immunization and role of SMS and automated call–based messages in improving vaccination coverage. This information will lead to the development of personalized barrier-specific messages (SMS and voice calls) that will be used for each participant or caregiver according to the specific barriers they struggle with, as identified at baseline.

Study Sites

The Department of Paediatrics and Child Health at the Aga Khan University (AKU) conducts active health demographic surveillance systems (HDSS) at several urban sites in Karachi. These sites are (1) Ibrahim Hydri Goth, (2) Ali Akber Shah Goth, (3) Rehri Goth, and (4) Bhains colony, which is part of Bin Qasim Town (Figure 2). Ibrahim Haidry Goth, Ali Akber Shah Goth, and Rehri Goth are located along the sea coast of Karachi, and the main occupation of people living in these communities is fishing. Bhains Colony is located at the outskirts of Karachi, and the main source of income of its population is dairy products.

Figure 1. Study design: a mixed methods randomized controlled trial (RCT) augmented by qualitative interviews. SMS: short message service.
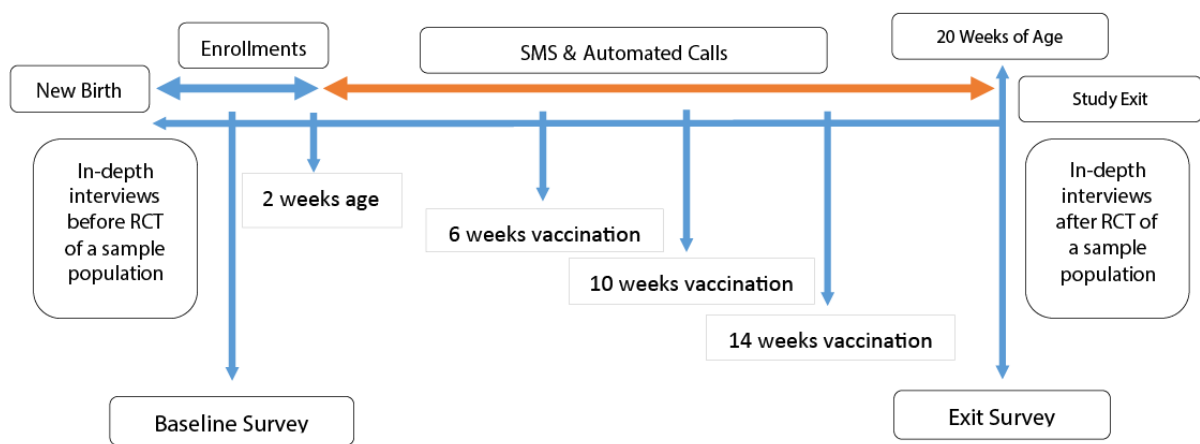
Figure 2. Catchment area of Karachi pre-urban demographic surveillance site.
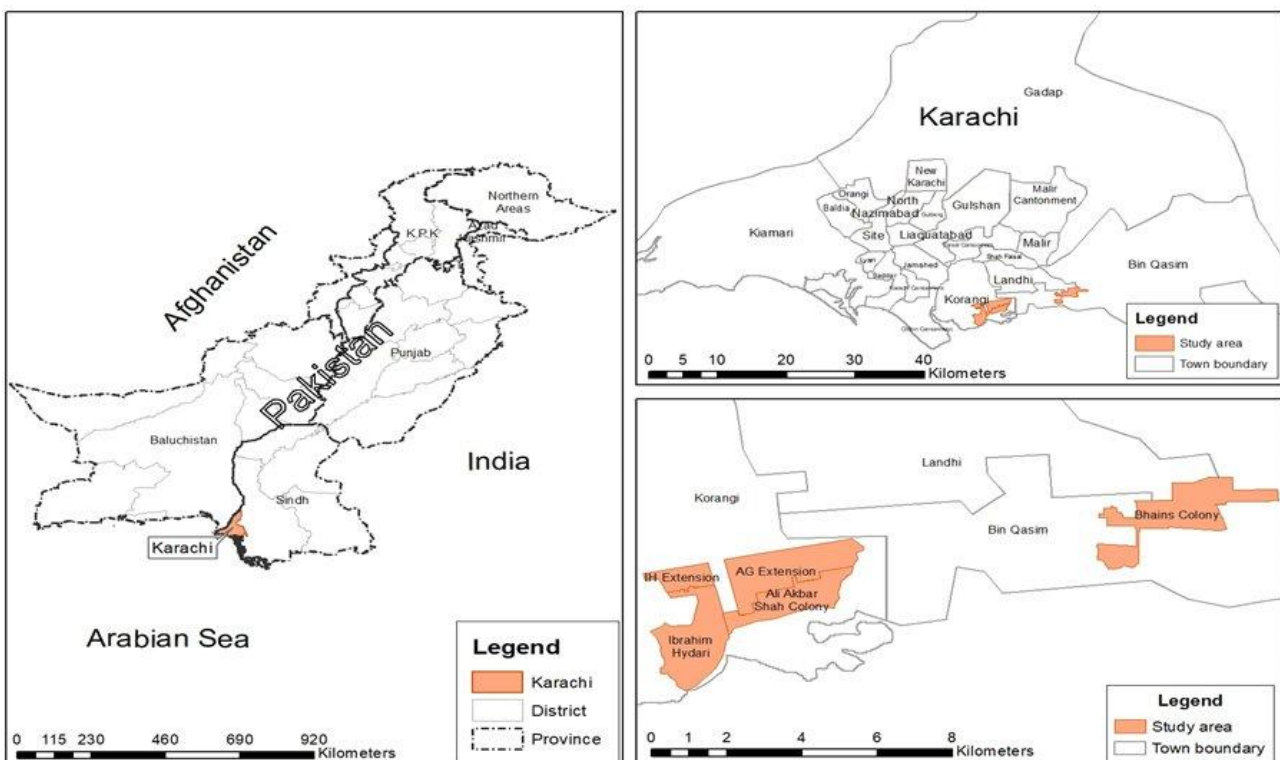
The HDSS sites are divided into 195 blocks, each containing 250 structures and mapped using the geographical information system. There are a total of 42,093 structures where 43,098 households are living in areas of Korangi and Bin Qasim towns [19]. A *structure* is defined as a building with a single entrance and a boundary. These structures can be houses, hospitals, dispensaries, schools, shops, parks, etc. Each structure has a unique number assigned to it. The total population of the active surveillance catchment area is approximately 0.3 million, with around 7115 pregnant women and 6831 NBs being followed annually by the surveillance team. Within 1 block, the 250 administrative structures will be our trial clusters; 4 sites will be participating in the study. NBs will be enrolled from the HDSS, and clusters within the 4 sites will be randomly assigned to the study arms. Hence, participants within each administrative structure (ie, cluster) will receive same intervention.

Data will also be collected from the rural site of Matiari, which is located in Sindh province 185 km north of Karachi (Figure 3). The department of pediatrics will partner with the local Lady Health Worker (LHW) program Sindh. Information about new births within LHW catchment area will be provided to the study team. The catchment areas in Matiari are the areas covered by each LHW; these areas are small and that is why the catchment area of 4 LHWs will be considered as 1 cluster. The total population of the Matiari District area is approximately 0.4 million. A total of 3 main sites from district Matiari will be the part of study, which includes Matiari, Hala, and Saeedabad. The main source of income in this area is agriculture. In Matiari and Karachi, clusters will have a mean of 15 births per cluster.

Figure 3. The Matiari site: catchment area of surveillance site.
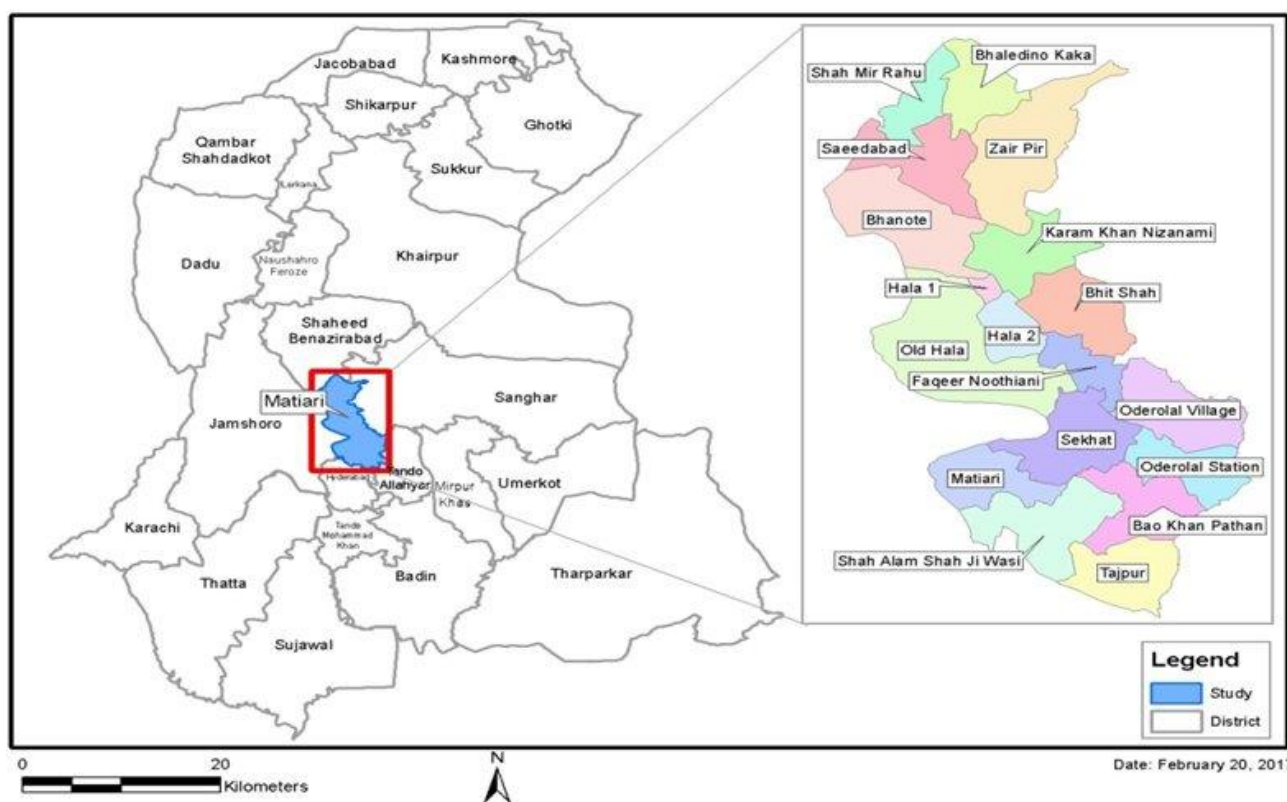
Baseline Survey

A total of 50 interviews will be conducted from Karachi and Matiari sites (ie, 25 interviews from each site). From the HDSS, list of caregivers who have completed the vaccination schedule and those who are dropouts or not vaccinating their child was taken out, and randomly, caregivers will be selected for in-depth interviews. Trained staff will conduct the qualitative interviews. Depending upon the availability, caregivers will be approached at the household, and appointments will be taken from the caregiver so that they can spare enough time for the discussion. Through in-depth interviews, we will explore the perceptions and barriers related to EPI and mobile health from the caregivers. Information collected before the study through qualitative in-depth interviews will help in developing the survey to assess the different categories of barriers. The baseline survey will, overall, collect information on demographics, mobile phone

access and usage, possible barriers for RIs, and factors associated with mobile phone–based health messages.

Qualitative Phase

Assessing Perceptions and Experiences of Caregivers About the Routine Immunization and the Role of Mobile Phone Short Message Service and Automated Calls in Improving the Routine Immunization Coverage and to Develop Messages

Initially, in-depth interviews with a subpopulation in the study catchment area will be conducted to explore RI immunization coverage among Pakistani caregivers and families to assess (1) perceptions regarding risks of infectious diseases preventable by vaccines and vaccine safety and efficacy, (2) barriers to vaccinating children including difficulties in visiting RI centers, and (3) perceptions and barriers that may affect mobile

phone–based interventions to improve immunization coverage. The interviews will be conducted across the site regions. We will use purposive sampling to assure that participants represent different ethnic groups. Information gathered through the interviews will help in (1) understanding the types of barriers perceived by caregivers, (2) designing the RCT and, (3) developing content for SMS text messages in several categories of barriers. We expect 20 to 25 parents or caregivers to be interviewed over a period of 3 months, but the number will be guided by saturation of information. Each interview will take 45 to 60 min.

Randomized Clinical Trial

The study design is a cluster randomized clinical trial including 4 intervention and 1 control arm (Table 2).

Sample Size Calculations

Assumptions used for calculating sample size are as follows: increase in coverage rate from 30% to 40%, having a power of 80% with an alpha error at .05. The sample size per arm would be 615 per arm or 3075 for all 5 arms. Adding a 10% dropout will make it 677 per arm or 3385 recruitment for all 5 study arms. The ICC is taken as .05, with an estimate of 15 NB per cluster (see Table 3).

Table 2. Cluster randomized clinical trial.

| Study arm | Automated SMS[a] text and phone calls | | | |
	1-way SMS text	2-way SMS text	1-way phone calls	2-way phone calls
1	X[b]	—[c]	—	—
2	—	—	X	—
3	—	X	—	—
4	—	—	—	X
5	—	—	—	—

[a]SMS: short message service.

[b]Applicable.

[c]Not applicable.

Table 3. Sample size calculation. Having a difference of 10% and the sample size given (ie, 1230 is 2 groups, 1:1) in our study, we have 5 groups (615x5)+10% dropout=3385.

| Serial # | Interclass correlation | Range of cluster size | | | | | |
| | | Cluster size of 15 | | Cluster size of 20 | | Cluster size of 30 | |
		Clusters, n	Total birth required, n	Clusters, n	Total birth required, n	Clusters, n	Total birth required, n
1	0.01	54	810	44	880	32	960
2	0.05	82	1230	70	1400	58	1740
3	0.09	108	1620	98	1960	86	2580
4	0.1	114	1710	104	2080	94	2820
5	0.13	134	2010	124	2480	114	3420
6	0.15	148	2220	138	2760	128	3840
7	0.2	180	2700	172	3440	162	4860
8	0.21	188	2820	178	3560	168	5040
9	0.25	214	3210	206	4120	196	5880
10	0.3	248	3720	238	4760	230	6900
11	0.35	280	4200	272	5440	264	7920
12	0.37	294	4410	286	5720	278	8340
13	0.4	314	4710	306	6120	300	9000
14	0.45	346	5190	340	6800	334	10,020
15	0.5	380	5700	374	7480	368	11,040

Enrollment Criteria

Inclusion Criteria

The inclusion criteria include being a child from the HDSS site, being younger than 14 days with a parent or guardian or at least 1 person in the household, having a working mobile phone connection, and parent or guardian providing consent to participate in the study.

Exclusion Criteria

The exclusion criteria include a child from outside the HDSS area or if the family plans to stay in the catchment area for less than 20 weeks.

Sampling Strategy and Randomization

The sampling strategy adapted for both sites are as follows (Figure 4).

Figure 4. Cluster Randomization.
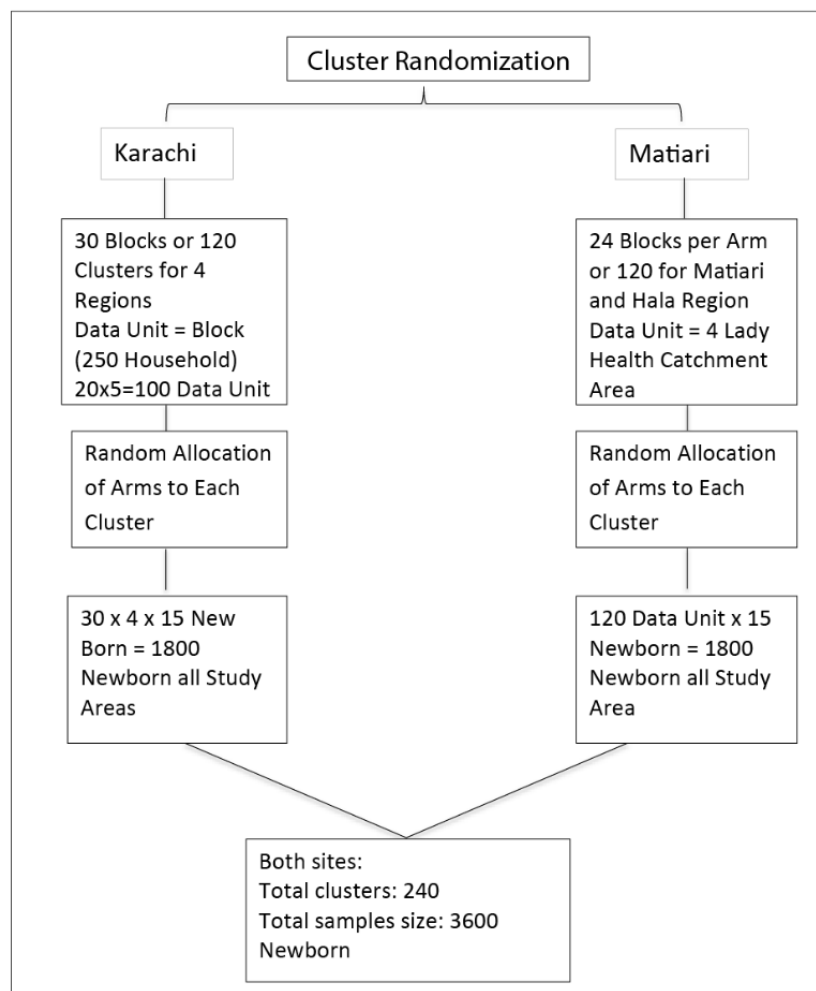
Karachi Site

The data for live births and population of each site in Karachi will be obtained from HDSS team. Data will be scrutinized and divided according to 4 sites in the study, with each site consisting of 5 arms (5 clusters). Administrative structures will be assigned to each intervention through stratified (ie, site) randomization.

Matiari Site

As Matiari site has not been stratified into clusters according to HDSS, we randomly allocated 30 clusters per arm. The catchment area of 4 LHWs will be considered as 1 cluster. Each intervention arm will then be randomly allocated to 24 administrative structures (clusters; see Figure 5).

Figure 5. Sampling strategy for Karachi and Matiari. For 10% effect size, sampling strategy cluster size=15.
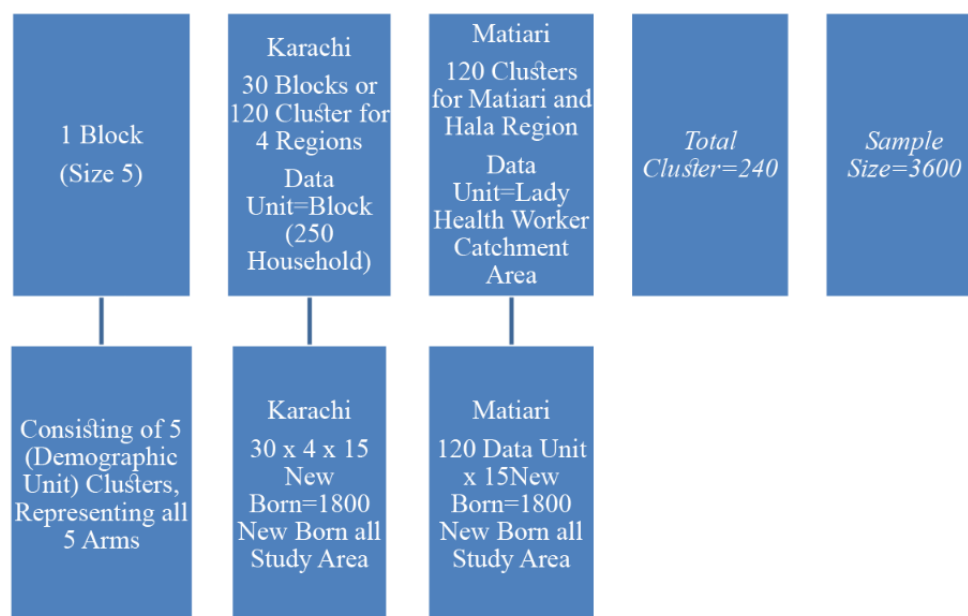
Recruitment

For Karachi, birth data of the clusters will be shared by central HDSS team on weekly basis and at Matiari the LHWs will inform births to the study team daily according to the study clusters. The study staff will approach families who have an infant younger than 2 weeks in each cluster.

Staff will explain the study objectives to the parents or caregivers. If the parent or caretaker is interested in the study, the infant will be evaluated for enrollment. One child per household will be selected. In a household, where there is more

than 1 child (younger than 14 days), a random selection will be made by a program designed in the mobile phone device. After meeting the eligibility criteria, informed consent will be obtained, and the infant will be enrolled in the study and a baseline survey will be conducted.

Follow-Up Process

A second survey will be conducted at 20 weeks of child's age (end of follow-up) to identify vaccination coverage according to the schedule, to be confirmed by physical examination of EPI card (see Figure 6).

Figure 6. Participant flow diagram. HDSS: health demographic surveillance systems; SMS: short message service.
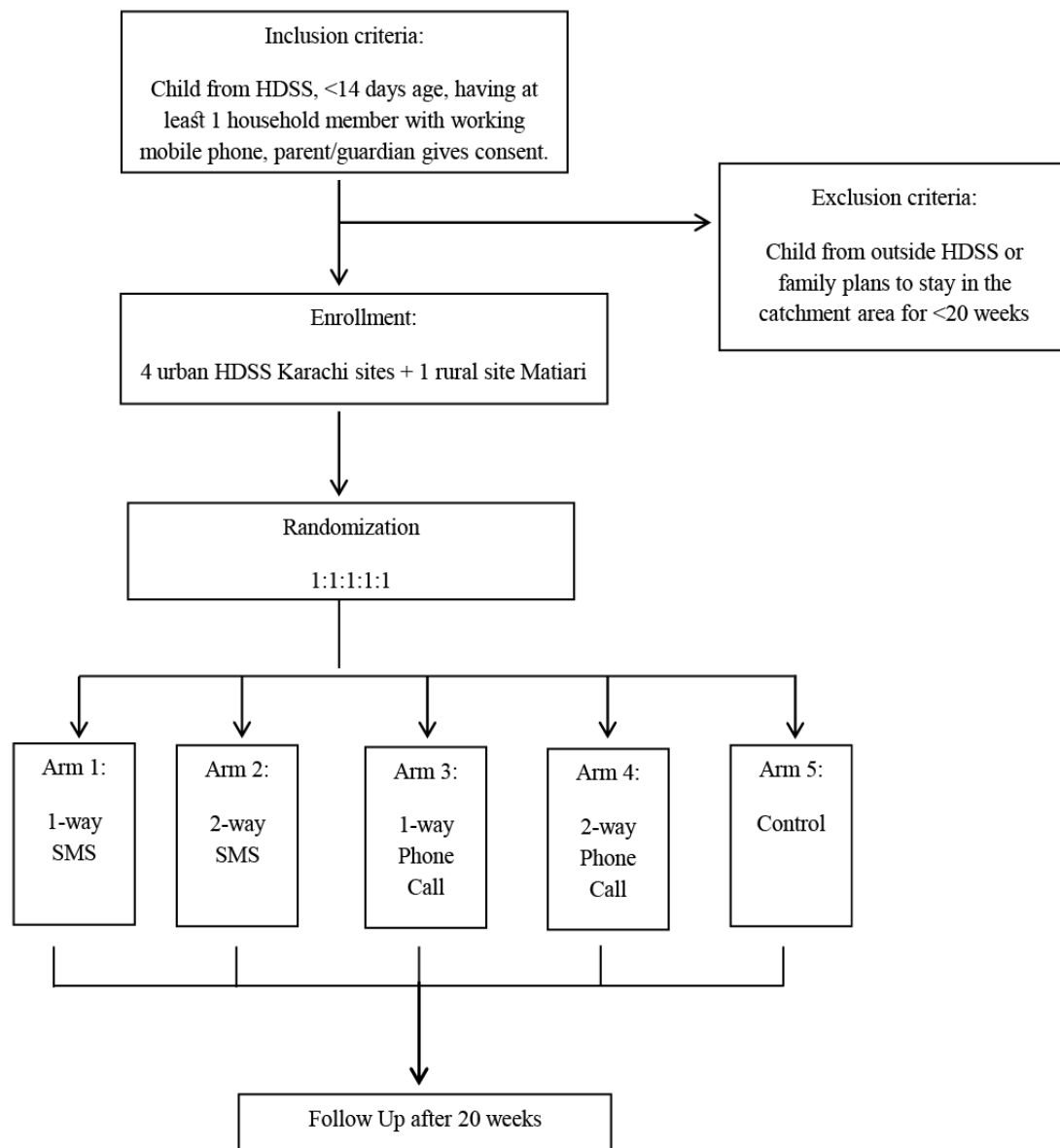
Study Intervention

Intervention Group

The intervention arms in addition to the standard counseling will include receiving personalized 1-way or 2- way (interactive) automated SMS or 1-way or 2-way (IVR) automated phone call messages regarding vaccination. In personalized 1-way messages and automated calls, parents will receive educational or reminder or proactive messages related to RI once a week till the child turns 20 weeks. SMS and automated phone calls will contain the same information content and interactive features, only the delivery method will differ. In the interactive arms, in addition

to personalized weekly educational or reminder or proactive message, parents will have the option to reply to messages or calls and receive more information related to immunization through SMS text messages or calls (see Figure 7 and Table 4 and Multimedia Appendices 1 and 2). The study timeline has been provided in Table 5.

Control Group

The control group will receive 1 time standard verbal counseling at the time of initial visit for on-time EPI vaccines at 6, 10, and 14 weeks of age as recommended by EPI, government of Pakistan.

Figure 7. Details of study arm explained.
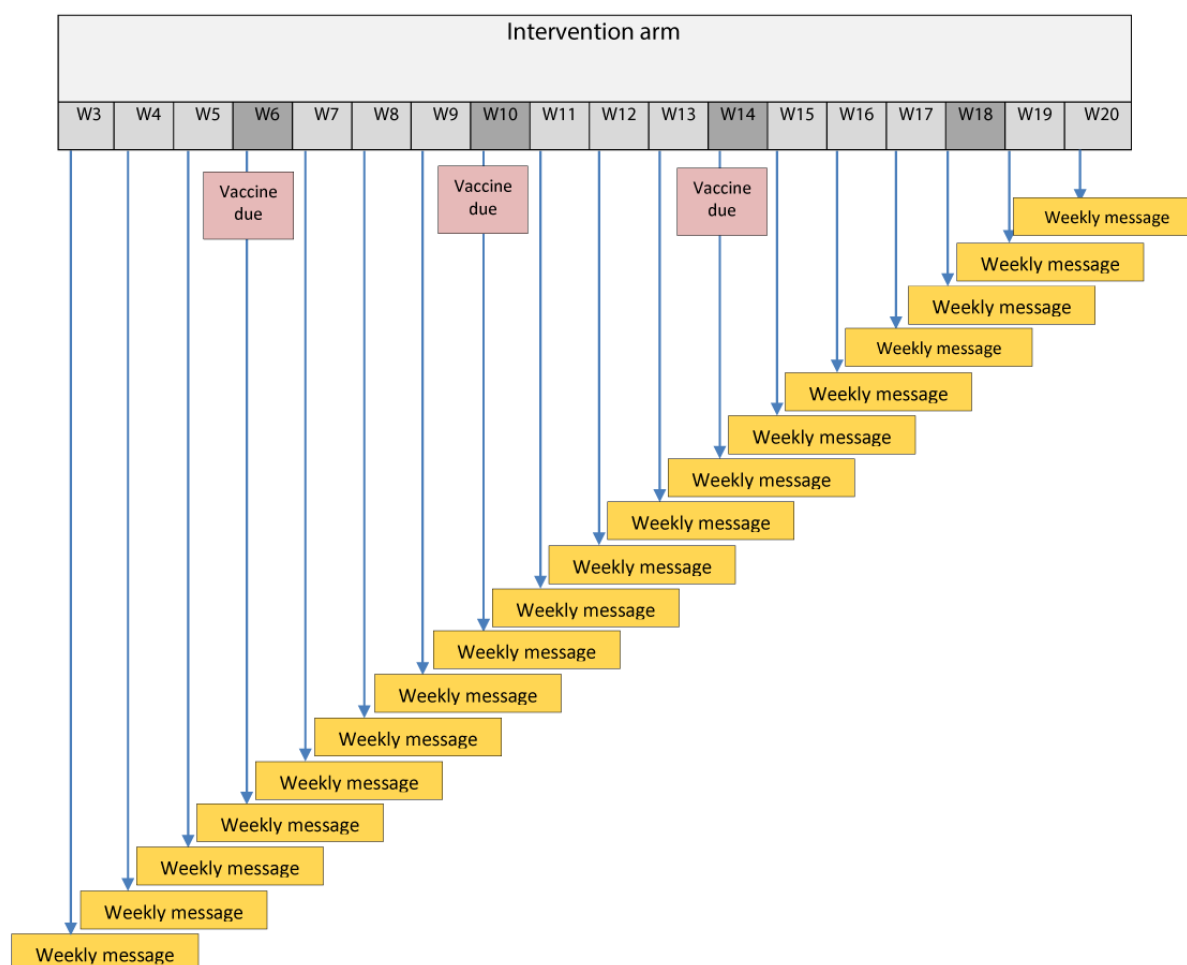
Table 4. Detail on the personalized weekly short message service and automated calls from enrollment to 20 weeks of life.

Intervention arm	Weekly automated SMS[a] text message and automated calls from enrollment till 20 weeks of life
Arm 1 (intervention)	Parents or caregivers will receive 1-way educational or reminder or proactive SMS messages related to routine immunization once a week till 20 weeks of age
Arm 2 (intervention)	Parents or caregivers will receive 2-way (interactive) educational or reminder or proactive SMS messages related to routine immunization once a week till 20 weeks of age—parents will have the option to reply and receive more information related to immunization through SMS text messages
Arm 3 (intervention)	Parents or caregivers will receive 1-way educational or reminder or proactive automated phone call related to routine immunization once a week till 20 weeks of age
Arm 4 (intervention)	Parents or caregivers will receive 2-way (interactive) educational or reminder or proactive automated phone call related to routine immunization once a week till 20 weeks of age—parents will have the option to reply and receive more information related to immunization through phone call
Arm 5 (control)	One time counseling at the baseline survey

[a]SMS: short message service.

XSL•FO
RenderX

Table 5. Study timeline.

Steps	2018			2019						
	Jan/Feb	Mar/Apr	May/Jun	Jul/Aug	Sep/Oct	Nov/Dec	Jan/Feb	Mar/Apr	May/Jun	Jul
Protocol development	X[a]	—[b]	—	—	—	—	—	—	—	—
Ethical review submission	X	—	—	—	—	—	—	—	—	—
Standard operating procedures	X	—	—	—	—	—	—	—	—	—
Training and implementation of field team	X	—	—	—	—	—	—	—	—	—
Qualitative interviews	X	X	X	—	—	—	—	—	—	X
Trial recruitment	—	—	—	X	X	X	X	—	—	—
Trial follow-up	—	—	—	—	—	X	X	X	X	—
Data entry cleaning and analysis	—	—	—	X	X	X	X	X	X	—
Report and scientific paper writing	—	—	—	—	—	—	—	—	X	X
Thesis completion	—	—	—	—	—	—	—	—	X	X
Additional—baseline, analyses	—	—	—	—	—	—	—	—	X	X

[a]Applicable.

[b]Not applicable.

Study End Interview

Interviews will be conducted after the trial, with a purposive sample of a subpopulation included in the RCT representing all study arms. From these structured in-depth interviews, we expect to further explore participants' experience of the different study arms having different levels of coverage. We will also assess the factors related to SMS text messages and automated calls associated with vaccination uptake. Sample size will be guided by saturation of the information.

Statistical Analysis

Analyses will be conducted according to the intention-to-treat principle. When the information cannot be gathered, the vaccine status of the child will be *failure* (conservative approach). Missing the vaccine EPI card to verify vaccine status at home will lead to a deeper investigation in the family and also the vaccine clinic. From previous experience, we expect this process to be successful in identifying the outcome of over 95% (95/100) of the children who remain in the study at 20 weeks. Families that have left the study for any reason (moving, withdrawal, child death) will be interviewed (if possible) and the information gathered used for the outcome classification. If the information is not clear or incomplete, the child's status regarding vaccination will be *failure*.

The main analyses will be comparing our 4 study intervention arms against the control, and with each other. For the purposes of our study, we will compare each of the 4 interventions with each other and with the control arm. For example, 1-way versus 2 SMS, reminder versus educational SMS, 1-way reminders SMS versus control, 1-way educational SMS versus control, 2-way reminders SMS versus control, 2-way educational SMS versus control, 2-way education versus 1-way reminder, and 2-way reminder versus 1-way educational.

Analysis of Quantitative Data

Analyses will be conducted according to the intention-to-treat principle, and the Bonferroni correction will be used to control for multiple testing. The primary outcome is to assess difference in vaccination status. Chi-square test will first be used to compare the groups' percentage. Multivariable logistic regression analyses will be conducted to adjust for confounders. As our randomization unit is cluster-based rather than individual-based, generalized linear mixed effects models or generalized estimating equations models will be employed to account for within-cluster correlation, and hierarchal models will be used. The statistical model for the primary and secondary analyses will be developed as explanatory variables. As secondary analyses of the primary outcome, time-to-immunization curves will be constructed using the Kaplan-Meier method and Cox regression for multivariate analysis.

Analyses of Qualitative Component

In the first section of the semistructured interview guide, we will ask the caregivers regarding common barriers they encounter at the time of RI, whereas the second section comprises perceptions and barriers related to a mobile phone. The data will be transcribed into written form from audio recordings and will be analyzed via qualitative data analysis software NVivo 11 (QSR International). Written transcripts will then be uploaded into NVivo 11 software to offer easy and organized retrieval of data for analysis. The data analysis will be conducted according to discourse analysis. The interview guides will also be pretested in a similar community

Data Management, Confidentiality and Privacy Protection, and Quality Assurance

The audio recording and transcripts will have a unique identifier; original and backup files will be archived in password-protected computer or server at AKU. Only transcriptions and themes

will be shared at University of British Columbia (UBC) without nominative information. All study-related data including the recording will be stored in an encrypted server with password protection and having access only to the study specific personnels.

Baseline and end line data will be collected on a smartphone device (Multimedia Appendices 3 and 5). The entry program will be designed to capture data as well as the location of the interviewer along with some monitoring parameters. Each child participant will be given a structured unique identifier. Business rules, skips, and consistency checks will be incorporated, and important fields will be marked as must enter in the questionnaire to maintain data collection quality. The database will reside on a central computer at AKU managed by the study staff. A Web-based dashboard will be designed to report daily study progress. Mobile numbers will not be shared except to track patterns of use. Only relevant study staff will have access to study data allowed by the local ethics committee. All study staff will undergo basic research ethics training. Participants' information will be given a study code, and no personal identifiers will be shared. Data confidentiality will be maintained at all times. No personal identifiers will be used in any reports or publication of the study. Mobile phone companies will be only provided mobile phone numbers to send SMS text messages and automated calls messages through gateway (Multimedia Appendix 4). No individual identifier such as names of participants and area of location will be shared. In addition, a confidentiality agreement has been signed with the university and the mobile companies stating that the numbers provided will only be used for the purpose of the trial. All study data and recordings will be destroyed within 5 years of the study according to the recommendation of the UBC and AKU ethics committees.

Ethical Considerations

The study protocol and associated study instruments, including consent forms in English and local language, were approved by UBC's and AKU's Ethics Review Committee before commencement of any study activities. The study will be conducted in accordance with the Helsinki declaration and established guidelines. All participants will be administered informed consent before participation. Participants have the right to refuse to participate in the study or leave the study at any time; this will not affect any services provided at the health center. Data confidentiality will be maintained at all times. No personal identifiers will be used in any reports or publication of the study.

Training and Refresher

Initial training of the study staff related to study protocol, SMS text messages and phone calls will be conducted. This will be followed by refresher training in the middle of the study.

Limitations or Mitigations

We recognize that not all barriers identified are modifiable and amenable to phone calls and SMS text messages. However, most of them reflect a priority ranking that does not favor child routine vaccination. One major reservation about SMS-based interventions is the recipient's level of literacy. However, there

has been mixed input related to the preference of phones calls as compared with SMS text messages in populations of low literacy and resource-constrained settings [20,27]. Mobile phone SMS text messages in local languages, pictorial messages, and in combination with phone calls can further reduce this gap. In our study, we will send messages in local languages (as per the preference of the participants). Although SMS text messages have a limitation of 160 characters and even fewer if translated in other characters, these limitations might help in making the messages simple and brief, especially for populations with low literacy levels [22,28].

The type of intervention will not be blinded, but all families in the same cluster will receive the same intervention. The primary outcome (vaccine completion) is not likely to be influenced by any knowledge of the intervention arm.

Result

The baseline survey was conducted from July 2018 to January 2019 where it was found that 97.9% (3797/3877) of care givers either owned or have a sharing access to a working mobile phone. A total of 50 IDIs were conducted before the start of the study, that is, 25 interviews per site, based on which the content of the SMS text messages and automated calls were developed. FGDs with caregivers and stake holders were then conducted to validate the developed content. The RCT enrollment was completed on January 31, 2019, and the participants continued receiving messages till the child was 20-week old. Currently, the study is in its final phase and expecting completion of exit surveys and interviews by June 2019.

Discussion

This study is the first one of this type to assess the efficacy of different types of SMS text messages and automated calls messages, with or without interactive messages in LMICs. Different types of SMS text messages and automated calls messages will also be personalized based on the identification of families' possible concerns or challenges regarding RI. Each family will receive one weekly message, and we expect families to talk to each other about the messages within the demographic unit, therefore facilitating circulation of information (this aspect will be assessed through the end-of-study interview). This information will be extremely useful to understand the effect of different types of messages in different contexts.

The qualitative aspect of our study will generate useful information regarding family's perception of vaccination and the daily life challenges for timely visits to the vaccine clinic. This information will be used for developing more complex interventions that use mobile phone messages and possibly other approaches to overcome some of the barriers. Finally, conducting this study in Pakistan will generate results that will be useful in most LMICs worldwide.

Representatives of public health, Ministry of Health, and community leaders will be part of the study steering committee to ensure transparency, direct communication, and shared decision making as part of on-going knowledge transfer strategy. We will also inform major policy makers, donors, and mobile

XSL·FO

RenderX

phone service providers. Study findings will be presented at   international scientific and programmatic forums.
and published in a scientific journal and national and

Conflicts of Interest

None declared.

Multimedia Appendix 1

Strategy for health messages content.

[PDF File (Adobe PDF File), 99KB - resprot_v8i5e12851_app1.pdf]

Multimedia Appendix 2

Example of 1-way and 2-way personalized messages.

[PDF File (Adobe PDF File), 1MB - resprot_v8i5e12851_app2.pdf]

Multimedia Appendix 3

Study database.

[PDF File (Adobe PDF File), 275KB - resprot_v8i5e12851_app3.pdf]

Multimedia Appendix 4

Study gateway.

[PDF File (Adobe PDF File), 316KB - resprot_v8i5e12851_app4.pdf]

Multimedia Appendix 5

Administrative information.

[PDF File (Adobe PDF File), 53KB - resprot_v8i5e12851_app5.pdf]

References

1. USAID. Research and Development Solutions. 2012. Childhood Immunization in Pakistan: Policy Briefs Series No. 3 URL: http://resdev.org/files/policy_brief/3/3.pdf [accessed 2019-05-08] [WebCite Cache ID 78DFvzdgZ]
2. Shen AK, Fields R, McQuestion M. The future of routine immunization in the developing world: challenges and opportunities. Glob Health Sci Pract 2014 Dec 10;2(4):381-394 [FREE Full text] [doi: 10.9745/GHSP-D-14-00137] [Medline: 25611473]
3. Kazi AM, Khalid M, Kazi AN. Failure of polio eradication from Pakistan: threat to world health. J Pioneer Med Sci 2014;4(1):8-9.
4. National Institute of Population Studies. 2017. Demographic and Health Survey, 2017-18 URL: https://www.dhsprogram.com/pubs/pdf/FR354/FR354.pdf [accessed 2019-05-08] [WebCite Cache ID 78DG3mndU]
5. Reid M, Fleck F. The immunization programme that saved millions of lives. Bull World Health Organ 2014 May 01;92(5):314-315. [doi: 10.2471/BLT.14.020514]
6. Anwer J, Saleem M, Babar M. Knowledge of vaccination in the community: a contributing factor for vaccination in children. Pak J Med Health Sci 2013;7:1007-1010 [FREE Full text]
7. Attaa A. Propakistani. 2014. Mobile Phone Users in Pakistan Cross 139.2 Million Mark URL: https://propakistani.pk/2014/07/10/mobile-phone-users-in-pakistan-cross-139-2-million-mark/ [accessed 2019-05-03] [WebCite Cache ID 785Wwde7e]
8. Nigerian Communications Commission. Subscriber statistics URL: http://www.ncc.gov.ng/index.php?option=com_content&view=article&id=125:art-statistics-
9. Pakistan Telecommunication Authority. Proliferation of SMS & MMS in Pakistan with emphasis on Premium Rate SMS services SMS/MMS URL: https://www.pta.gov.pk/media/sms_mms_report_130313_1.pdf [accessed 2019-05-03] [WebCite Cache ID 785Y4vVHQ]

XSL•FO
RenderX

10.   Brown V, Oluwatosin O. Feasibility of implementing a cellphone-based reminder/recall strategy to improve childhood routine immunization in a low-resource setting: a descriptive report. BMC Health Serv Res 2017 Dec 04;17(Suppl 2):703 [FREE Full text] [doi: 10.1186/s12913-017-2639-8] [Medline: 29219093]

11.   Posadzki P, Mastellos N, Ryan R, Gunn L, Felix L, Pappas Y, et al. Automated telephone communication systems for preventive healthcare and management of long-term conditions. Cochrane Database Syst Rev 2016 Dec 14;12:CD009921 [FREE Full text] [doi: 10.1002/14651858.CD009921.pub2] [Medline: 27960229]

12.   Franzini L, Rosenthal J, Spears W, Martin H, Balderas L, Brown M, et al. Cost-effectiveness of childhood immunization reminder/recall systems in urban private practices. Pediatrics 2000 Jul;106(1 Pt 2):177-183. [Medline: 10888689]

13.   Domek G, Contreras-Roldan IL, O'Leary S, Bull S, Furniss A, Kempe A, et al. SMS text message reminders to improve infant vaccination coverage in Guatemala: a pilot randomized controlled trial. Vaccine 2016 Dec 05;34(21):2437-2443 [FREE Full text] [doi: 10.1016/j.vaccine.2016.03.065] [Medline: 27026145]

14.   Gibson D, Ochieng B, Kagucia E, Were J, Hayford K, Moulton L, et al. Mobile phone-delivered reminders and incentives to improve childhood immunisation coverage and timeliness in Kenya (M-SIMU): a cluster randomised controlled trial. Lancet Glob Health 2017 Dec;5(4):e428-e438 [FREE Full text] [doi: 10.1016/S2214-109X(17)30072-4] [Medline: 28288747]

15.   Haji A, Lowther S, Nganga Z, Gura Z, Tabu C, Sandhu H, et al. Reducing routine vaccination dropout ratesvaluating two interventions in three Kenyan districts, 2014. BMC Public Health 2016;16:152. [doi: 10.1186/s12889-016-2823-5]

16.   Garcia-Dia MJ, Fitzpatrick JJ, Madigan EA, Peabody JW. Using text reminder to improve childhood immunization adherence in the Philippines. Comput Inform Nurs 2016 Nov 08;35(4):212-218. [doi: 10.1097/CIN.0000000000000307] [Medline: 27828815]

17.   Ahlers-Schmidt C, Chesser A, Paschal A, Hart T, Williams K, Yaghmai B, et al. Parent opinions about use of text messaging for immunization reminders. J Med Internet Res 2012 Jun 06;14(3):e83 [FREE Full text] [doi: 10.2196/jmir.1976] [Medline: 22683920]

18.   World Health Organization. Expanded Programme on Immunization URL: http://www.emro.who.int/pak/programmes/expanded-programme-on-immunization.html [accessed 2019-05-03] [WebCite Cache ID 785bC2CFF]

19.   Illyas M, Naseem K, Fatima U, Nisar M, Kazi AM, Jehan F, et al. Profile: health and demographic surveillance system in peri-urban areas of Karachi, Pakistan. Gates Open Research 2018. [doi: 10.12688/gatesopenres.12788.1]

20.   Jo Y, Labrique A, Lefevre A, Mehl G, Pfaff T, Walker N, et al. Using the lives saved tool (LiST) to model mHealth impact on neonatal survival in resource-limited settings. PLoS One 2014;9(7):e102224 [FREE Full text] [doi: 10.1371/journal.pone.0102224] [Medline: 25014008]

21.   Eze G, Adeleye O. Enhancing routine immunization performance using innovative technology in an urban area of Nigeria. West Afr J Med 2015;34(1):3-10. [Medline: 26902809]

22.   Bangure D, Chirundu D, Gombe N, Marufu T, Mandozana G, Tshimanga M, et al. Effectiveness of short message services reminder on childhood immunization programme in Kadoma, Zimbabwe - a randomized controlled trial, 2013. BMC Public Health 2015 Feb 12;15:137 [FREE Full text] [doi: 10.1186/s12889-015-1470-6] [Medline: 25885862]

23.   Stockwell M, Kharbanda E, Martinez R, Vargas C, Vawdrey D, Camargo S. Effect of a text messaging intervention on influenza vaccination in an urban, low-income pediatric and adolescent population: a randomized controlled trial. J Am Med Assoc 2012 Apr 25;307(16):1702-1708. [doi: 10.1001/jama.2012.502] [Medline: 22535855]

24.   Stockwell M, Kharbanda E, Martinez R, Lara M, Vawdrey D, Natarajan K, et al. Text4Health: impact of text message reminder-recalls for pediatric and adolescent immunizations. Am J Public Health 2012 Feb;102(2):e15-e21. [doi: 10.2105/AJPH.2011.300331] [Medline: 22390457]

25.   Stockwell M, Hofstetter A, DuRivage N, Barrett A, Fernandez N, Vargas C, et al. Text message reminders for second dose of influenza vaccine: a randomized controlled trial. Pediatrics 2015 Jan;135(1):e83-e91 [FREE Full text] [doi: 10.1542/peds.2014-2475] [Medline: 25548329]

26.   Poorman E, Gazmararian J, Parker R, Yang B, Elon L. Use of text messaging for maternal and infant health: a systematic review of the literature. Matern Child Health J 2015 May;19(5):969-989. [doi: 10.1007/s10995-014-1595-8] [Medline: 25081242]

27.   Wakadha H, Chandir S, Were E, Rubin A, Obor D, Levine O, et al. The feasibility of using mobile-phone based SMS reminders and conditional cash transfers to improve timely immunization in rural Kenya. Vaccine 2013 Jan 30;31(6):987-993 [FREE Full text] [doi: 10.1016/j.vaccine.2012.11.093] [Medline: 23246258]

28.   Lund S, Rasch V, Hemed M, Boas I, Said A, Said K, et al. Mobile phone intervention reduces perinatal mortality in zanzibar: secondary outcomes of a cluster randomized controlled trial. JMIR Mhealth Uhealth 2014 Mar 26;2(1):e15 [FREE Full text] [doi: 10.2196/mhealth.2941] [Medline: 25098184]

Abbreviations

AKU: Aga Khan University
EPI: Expanded Programme on Immunization
HDSS: health demographic surveillance systems
IVR: interactive voice recording

LHW: Lady Health Worker
LMIC: low and middle-income country
NB: newborn
RCT: randomized controlled trial
RI: routine immunization
SMS: short message service
UBC: University of British Columbia
VPD: vaccine-preventable disease

XSL•FO
RenderX

Protocol

Effective Information Provision About the Side Effects of Treatment for Malignant Lymphoma: Protocol of a Randomized Controlled Trial Using Video Vignettes

Nanon Labrie[1,2], PhD; Sandra van Dulmen[3,4,5], PhD; Marie José Kersten[6], MD, PhD; Hanneke JCM de Haes[1], PhD; Arwen H Pieterse[7], PhD; Julia CM van Weert[8], PhD; Dick Johan van Spronsen[9], MD, PhD; Ellen MA Smets[1], PhD; INSTRUCT Group[1]

[1]Medical Psychology, Amsterdam Public Health Research Institute, Cancer Center Amsterdam, Amsterdam University Medical Centers, Amsterdam, Netherlands

[2]Athena Institute, Vrije Universiteit van Amsterdam, Amsterdam, Netherlands

[3]Netherlands Institute for Health Services Research (NIVEL), Utrecht, Netherlands

[4]Radboud Institute for Health Sciences, Department of Primary and Community Care, Radboud University Medical Center, Nijmegen, Netherlands

[5]Faculty of Health and Social Sciences, University of South-Eastern Norway, Drammen, Norway

[6]Department of Hematology, Cancer Center Amsterdam, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands

[7]Medical Decision Making, Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, Netherlands

[8]Amsterdam School of Communication Research, University of Amsterdam, Amsterdam, Netherlands

[9]Department of Hematology, Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Ellen MA Smets, PhD
Medical Psychology
Amsterdam Public Health Research Institute, Cancer Center Amsterdam
Amsterdam University Medical Centers
Amsterdam University Medical Centers
Amsterdam,
Netherlands
Phone: 31 205664768
Fax: 31 205669104
Email: e.m.smets@amc.uva.nl

Abstract

Background: Informing patients with cancer about the possible implications of prospective treatment is a crucial yet challenging task. Unfortunately, patients' recall of medical information is generally poor and their information needs are not met. Effective information giving entails that oncologists help patients understand and recall the implications of their treatment, meanwhile fostering a trusting physician-patient relationship. Communication strategies that are often suggested to be effective are *structuring* and *tailoring* (cognition-oriented) but also are oncologists' expressions of *caring* or empathy (affect-oriented).

Objective: The aim of this study is to provide evidence concerning the pathways linking physician communication to (improved) consultation outcomes for patients. More specifically, the aim is to determine the effects of information *structuring* and information *tailoring*, combined with physician *caring*, on information recall, satisfaction with information, and trust in the physician (primary objective) and on symptom distress (secondary objective).

Methods: A randomized controlled trial, systematically testing the effects of information *structuring* and information *tailoring*, each combined with *caring*, in 2 video-vignette experiments (2×2 and 2×2×2 design). Using an online survey platform, participants will be randomly allocated (blinded) to 1 of 12 conditions in which they are asked to view a video vignette (intervention) in which an oncologist discusses a treatment plan for malignant lymphoma with a patient. The independent variables of interest are systematically varied across conditions. The outcome measures are assessed in a survey, using validated instruments. Study participants are (former) patients with cancer and their relatives recruited via online panels and patient organizations. This protocol discusses the trial design, including the video-vignette design, intervention pretesting, and a pilot study.

Results: Data collection has now been completed, and preliminary analyses will be available in Spring 2019. A total of 470 participants completed the first part of the survey and were randomized to receive the intervention.

Conclusions: The results of the proposed trial will provide evidence concerning the pathways linking physician information, giving skills to (improved) consultation outcomes for patients.

Trial Registration: Netherlands Trial Register NTR6153; https://www.trialregister.nl/trial/6022 (Archived by Webcite at http://www.webcitation.org/76xVV9xC8).

International Registered Report Identifier (IRRID):   DERR1-10.2196/12453

Introduction

Background

Informing patients with cancer about the possible implications of prospective treatment is a crucial yet challenging task. Cancer treatment plans are typically complex, and the effects on patients' physical and psychological well-being can be severe. Although it is important that patients remember treatment information, research consistently shows that patients' recall of medical information is poor [1-7]. Patients forget approximately 40% to 80% of the information that is provided by their oncologist [5,8-10].

Lack of information is not only potentially harmful but has also been cited as among the greatest causes of dissatisfaction in patients with cancer [1,6,11-13]. Patients mostly want information about treatment [14], particularly about symptoms and side effects, both in the short and long term [15-18]. Having information about symptoms and treatment may provide patients with a sense of control, reduce their anxiety and distress, and provide support coping with the physical and psychological demands of cancer treatment [18,19]. Additionally, by discussing current and future symptom experiences, physicians can influence patients' expectations of symptoms and their ability to control symptoms [20]. These expectations may subsequently affect patients' actual symptom experiences, either positively (placebo effect) or negatively (nocebo effect) [20,21].

Finally, providing comprehensive and understandable information that is congruent with patients' needs is known to increase patients' trust in the physician [22], which is associated with a higher tolerance for symptoms [21,23]. Indeed, patients with cancer who feel more able to cope with the disease and its treatment are better adjusted and experience greater quality of life than patients who feel less in control [24]. Therefore, effective information giving entails that oncologists help patients understand and recall the implications of their treatment, meanwhile fostering a trusting physician-patient relationship.

Strategies of Effective Information Giving

Communication strategies to enhance information provision can be described as either cognition- or affect-oriented [25,26]. Cognition-oriented strategies are typically aimed at enhancing patient-related outcomes that are cognitive in nature, such as patients' recall of information [27-31]. Two prominent cognition-oriented strategies are information *structuring* and *tailoring*. Affect-oriented strategies target patients' emotions and include, for example, oncologists' expressions of *caring* or empathy [9,31-35]. Owing to the inherent interplay between the cognition-oriented and affect-oriented aspects of information giving [36], these strategies should ideally be studied in conjunction.

Structuring

Structuring treatment information, that is, a clear organization of information provision during a consultation, is assumed to improve patients' recall. Structure allows patients to systematically organize and store information in their working memory, such that it is easier to remember at a later moment [28,29]. Similar to the way in which newspaper articles or books are structured by means of, for example, (sub)titles and paragraph/chapter headings, physicians can use verbal structure *signals* to guide their patients through the information (also called the *book metaphor*) [28]. A total of 4 types of explicit verbal structure signals can be distinguished [37,38]: (1) Statements that set the agenda and announce key topics that will be dealt with in detail later (eg, "The most important issues to be discussed are..."); (2) Statements used to conclude or summarize the most important issues discussed (eg, "All in all, there are four main issues to be considered..."); (3) Ordinal or numeral signals that indicate elements of a series (eg, "first, second,..." and "in addition"); and (4) Statements expressing an opinion or a point of view (*"unfortunately"*; *"in my opinion"*) [37,38]. As early as in the 1970s, Ley et al found that the use of explicit structure signals in medical information can improve students' memory [12]. Four decades later, Langewitz et al [28] demonstrated that providing verbal structure signals during discharge consults in emergency medicine significantly improved students' recall. These results are yet to be replicated in clinical contexts.

Tailoring

Tailoring, that is, adjusting the (amount of) information to meet an individual patient's information need, is proposed to be more effective than the provision of generic information [39,40]. People tend to pay more attention to information that they perceive as personally relevant, which leads to improved information processing and consequently better recall [40-42]. For patients with cancer, providing them with *less* information than they wish is known to cause dissatisfaction [16] and

providing them with *more* information than they desire can also be harmful [43]. Consequently, a tailored approach, that is, congruence between the patients' information need and physician's information provision, is generally advocated [40,41,43-46].

Caring

Caring refers to a communication style in which the physician displays behaviors of empathy for and affective engagement with the patient [9,47,48,49], thereby potentially reducing patients' emotional distress [9,32,48] and enhancing patients' memory of information [49,50]. A sense of a caring relationship with the physician has been shown to increase patients' satisfaction with the provided information [34]. Other studies demonstrate that patients' trust in their physician reduces the need to subsequently seek detailed information [44,51]. Moreover, research suggests that physicians can help alleviate symptom distress by affective communication, rather than by information giving alone [21,52].

The effectiveness of the aforementioned cognitive (ie, structuring and tailoring) and affective (ie, caring) strategies may differ, depending on patients' individual characteristics known to affect information processing, such as their age [5], their degree of anxiety [35] or coping style [53], or their medical history.

Research Objectives

The randomized controlled trial described in this study protocol aims to provide evidence concerning the pathways linking physician communication to (improved) consultation outcomes for patients. More specifically, it seeks to determine the effects of information *structuring* (experiment 1) and information *tailoring* (experiment 2), combined with physician *caring*, on information recall, satisfaction with information, and trust in the physician (primary objective). Additionally, it aims to determine the effects of these independent variables on expected symptom distress (secondary objective). The planned trial consists of a single study in which 2—analytically distinct—subexperiments can be identified—each with specific objectives related to the independent variables.

Hypotheses and Research Questions

The hypotheses and research questions are as follows (see Figure 1):

1. Experiment 1: Structuring
   - H1: Information *structuring* positively affects patients' recall of treatment information.

2. Experiment 2: Tailoring
   - H2: Information *tailoring* positively affects patients' recall of treatment information.

3. Experiments 1 and 2: Caring
   - H3: Oncologists' expressions of *caring* positively affect patients' recall of, and satisfaction with, treatment information and their trust in the oncologist.
   - RQ1: Is there an interaction effect of information *structuring* or *tailoring* and oncologists' expressions of *caring*?
   - RQ2: Do patients' recall of, and satisfaction with, treatment information and their trust in the oncologist affect patients' expected symptom distress?
   - RQ3: Do patient characteristics, including sociodemographics (eg, gender, age, and health literacy), medical history (eg, type and year of diagnosis and treatment), and personality traits (eg, coping style and trait anxiety), moderate the hypothesized relationships?

Figure 1. Hypotheses and research questions.
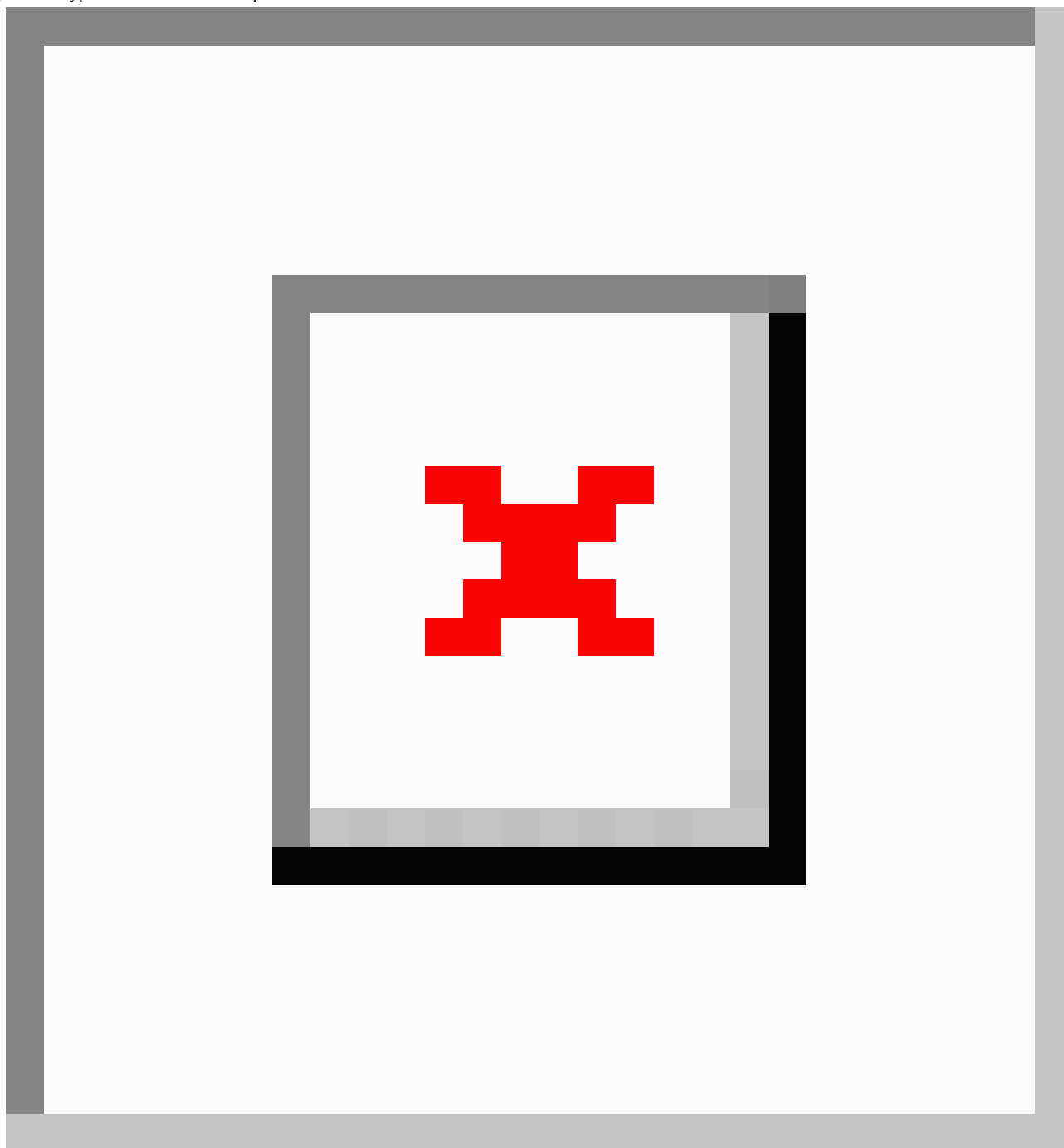
Methods

Trial Design

The trial employs a between-subjects single-message factorial design in which the independent variables of interest are systematically manipulated. This is done using the video vignette methodology. In experiment 1, hematologists' information structuring and expressions of caring are varied in a 2 (standard versus enhanced structuring) × 2 (standard versus enhanced caring) design. In experiment 2, hematologists' information tailoring and expressions of caring are varied in a 2 (high versus low need for information) × 2 (high versus low amount of information provided) × 2 (standard versus enhanced caring) design. This results in 12 experimental conditions, or interventions, across the trial (see Tables 1 and 2).

Table 1. Experiment 1: Manipulation of provider information structuring and caring (2×2 design).

Manipulations	Standard of care	Caring
Standard of care	Standard of care	Enhanced caring
Structuring	Enhanced structuring	Enhanced caring and structuring

Table 2. Experiment 2: Manipulation of provider caring and information tailoring in a match/mismatch design (2×2×2 design).

Manipulations	Provider			
	Enhanced caring		No caring	
	Additional information provision	No additional information provision	Additional information provision	No additional information provision
Patient				
High information need	Tailoring match; (++[a])	Tailoring mismatch; (+ −[b])	Tailoring match; (++)	Tailoring mismatch; (+ −)
Low information need	Tailoring mismatch; (− +[c])	Tailoring match; (− −[d])	Tailoring mismatch; (− +)	Tailoring match; (− −)

[a]Patient with high information need receiving additional information provision.

[b]Patient with high information need yet not receiving additional information.

[c]Patient with low information need nevertheless receiving additional information.

[d]Patient with low information need not receiving additional information.

Study Setting

The trial is set within the specific context of hematology and the treatment of malignant lymphoma (diffuse large B-cell lymphoma [DLBCL]). Treatment information, including information about likely symptoms, is particularly important in this context because treatments are highly unpleasant and require patients' commitment and adherence in the face of these demands [36]. The trial is conducted in the Netherlands. This study has been evaluated by the medical ethical committee of the Academic Medical Center. The institutional ethics committee has determined that the study is exempt from the need for review according to the Dutch regulations for research involving human subjects (W16_054 # 16.069, date: February 2016).

Eligibility

Participants are (former) patients with cancer and their relatives. Eligible participants (1) have experience with oncology consultations, as a (former) patient or relative; (2) are fluent in Dutch; (3) are at least 18 years of age, and (4) have access to the internet at their home computer to complete the survey and view a video vignette. Inclusion of (former) patients with cancer and their relatives is believed to maximize identification with the patient displayed in the video vignette intervention.

Video Vignette Intervention

Video Vignettes

Eligible participants are randomized in equal proportions to view 1 of 12 video vignette interventions. Video vignettes are scripted, hypothetical scenarios of real-life (medical) consultations, which allow for the systematic variation of verbal and/or nonverbal behaviors across experimental conditions [48,54-63]. Video vignettes are preferred in health communication research when ethical or practical considerations prevent the manipulation of physician behaviors in clinical practice. Participants are asked to view and evaluate vignettes while imagining themselves to be the patient in the video, that is, participants act as *analogue patients* [61,62]. Several studies demonstrated the validity of this type of methodology [60-63] and show that video vignettes allow for good levels of experimental control through script standardization and manipulation checks. Formal guidelines for video vignette

development proposed by Hillen et al [61] and van Vliet et al [62] are used in this study to enhance internal and external validity.

Scenario Development

The first author (NL) used existing recordings (n=12) [64] and 2 days of real-life observations of consultations between hemato-oncologists and lymphoma patients to develop a script of a prototypical treatment-related consultation in hematology. Additionally, instructional materials for patients and evidence-based publications about malignant lymphoma and, in particular, DLBCL, its treatment (R-CHOP), and possible side effects were consulted. The main side effects and complaints associated with DLBCL and its treatment, as derived from the literature, included fatigue, nausea, infections, and anxiety.

When possible, exact excerpts from the transcribed consultations and information materials were embedded in the basic script to enhance ecological validity. To further ensure realism of the script, the basic script was discussed and revised in several discussion rounds with the project's lead hematologist (MJK). Furthermore, the script was role played by 2 medical communication researchers to test the natural flow of the dialogue. The basic script was then sent out for commentary to an expert panel consisting of 8 hematologists, radiotherapists, and oncologists from academic and regional hospitals, ranging in experience from resident in training to senior attending. Additionally, 11 patients with a history of lymphoma or blood cancer were recruited to provide written feedback. This was done via PanelCom [65], an online panel for patient-provider research. Expert panel members were subsequently excluded from further trial participation. Finally, a professional scriptwriter also commented on the script. The physicians, patients, and scriptwriter were asked to specifically provide feedback on the script's (medical) realism and the interaction between the physician and patient. The script was revised accordingly and discussed for final revision by the overall project group, including 4 medical psychologists, 1 hematologist, 3 medical communication researchers, and 2 medical education experts.

Experimental Manipulations

The basic script was designed such that the independent variables of interest—structuring, tailoring, and caring—could be integrated into the dialogue in the form of blocks of text fragments. These fragments sometimes consisted of a turn-taking sequence between the hematologist and patient or of short utterances added to the hematologist's text. Experimental manipulations were thus operationalized primarily as verbal expressions, which were in turn supported by nonverbal behaviors, if possible (eg, using hand signals to support statements such as "first," "second").

Structuring

On the basis of theoretical conceptualizations of text structuring [37,38,66,67], providing structure when giving information was operationalized (1) by having the physician provide verbal signals that introduce content and set the agenda (eg, "In today's consultation I will tell you more about the treatment with chemotherapy and what to expect"; "I would like to discuss four possible side-effects and complaints with you"), (2) by having the physician summarize information (eg, "In short, you could thus suffer from nausea, fatigue, and infections"), and (3) by having the physician use ordinal or numeral text signals to indicate separate elements in a series (eg, "first, second"; "additionally, moreover, finally"). These structure markers were absent in the standard script (see Table 2).

Tailoring

Patients' need for information was operationalized in the video script by having the patient respond to a prompt by the physician (eg, "Would you like to know more about this?"). The patient either confirmed a preference for more information (eg, "I would like to know as much as possible") or stated that the information received was considered sufficient for the time being (eg, "It's clear for now. I would like to let it all sink in"). This was done twice, once at the beginning and once toward the end of the consultation, keeping the patient's need for information (high vs low) consistent across the script. Tailoring was defined as a match between a high information need of the patient and the provision of further information (tailoring$^+$) or a match between a low-information need of the patient and the absence of further information provision (tailoring$^-$). In contrast, lack of tailoring was defined as a mismatch between need for information and information provision (no tailoring$^+$ and no tailoring$^-$; see Table 2).

Caring

Physician caring was operationalized based on Hillen et al [33] who developed, tested, and used doctors' verbal expressions of caring in a scripted video vignette study to test the effect of caring on trust. These verbal utterances were modified to fit the hemato-oncology setting, based on feedback from the expert panel, the patients, screenwriter, and project group during script development. The overall effectiveness of these manipulations of provider caring was established previously [33]. In the standard script, these expressions of caring were absent (Table 2).

Pretest Script Manipulations

To test the efficacy of the manipulations pertaining to information structuring and information preference tailoring, a pretest was conducted among a convenience sample of 63 participants (76%, [48/63] female; age range: 20 to 72 years; mean 41.5), including 19% (12/63) physicians (5 of which were hematologists), 11% (7/63) patients with lymphoma and blood cancer, 33% (21/63) researchers, and 37% (23/63) participants without previous experience in hemato-oncology. Participants were randomly assigned to one of the experimental conditions in which they were asked to read short, relevant excerpts of the script. Depending on the condition, participants were asked to rate the extent to which the physician *structured* the provided information or *adjusted* the amount of information to the patient's personal needs on a scale from 1 (not at all) to 10 (a lot). In an open-ended question, they were asked to explain their judgment. The findings from the pretest suggested that these manipulations were largely recognized. Nonetheless, participants did not always correctly distinguish between structured and unstructured information provision. To resolve this, information structuring signals were made more explicit in the script, for example, by reformulating text fragments more strongly and by emphasizing verbal statements with nonverbal behaviors.

Filming and Editing

The roles of physician and patient were played by professional actors with ample experience as standardized patients in the medical context and with video vignette research in particular. The role of the hematologist was played by a 51-year-old white male; the role of the patient was played by a 57-year-old white female. The video vignettes were recorded by a professional film crew, over the course of 2 days, at our hospital. The first day of filming was used as a training day and resulted in a preliminary video clip, shot with a single-camera setup. This clip was shown to a group of 10 medical communication and education experts who provided feedback on, for example, aspects such as quality of the image, as well as the acting skills and realism of the set. Changes were made where necessary. On the second day of filming, the entire script was filmed using a multicamera setup: the scenario was shot from 3 different angles. Subsequent editing resulted in 12 experimental video vignettes, ranging in length between 9 and 11 min.

Outcomes, Survey Development, and Testing

Survey (Outcome) Measures

As preparatory work for the trial, the experimental survey, including the study outcome measures, was developed and tested in a pilot study.

Background Measures

Survey questions concerning participants' sociodemographic background included gender, birth year, ethnicity, living situation, educational level, and occupational status. In addition, questions concerning participants' health literacy [68], medical knowledge (general and lymphoma-specific), overall health (1 item), and cancer (treatment and family) history were included. Personality trait measures included the avoidance scale of the Impact of Event Scale (8 items, 4-point scale) [69] to assess the tendency to avoid cancer-related issues, the Trait Anxiety

Inventory (20 items, 4-point scale) [70] for the assessment of generalized anxiety, the Threatening Medical Situations Inventory (TMSI-2, Monitoring scale) [71] to assess a monitoring coping style, and a single item assessing information preferences in medical consultations (5-point scale).

Manipulation Checks

To assess manipulation success of the independent variables, 3 items similar to those used in the pretest were included (scale 1-10; perceived structuring, tailoring, and caring). Open text boxes were added for participants to explain their judgments. The Video Engagement Scale (15 items; 7-point scale) [60] was included to measure participants' involvement with the video vignette.

Information Recall

Information recall was measured following the protocol of the Netherlands Patient Information Recall Questionnaire [8]. On the basis of the video vignette script, an item pool was developed, pairing open-ended questions (active recall) with analogous multiple-choice questions (recognition). A code sheet was developed by the authors (NL and ES) to assess correctly recalled items and calculate active recall and recognition scores. The results from the pilot test were used to refine the scale items. Two coders independently scored the pilot answers. In the case of disagreement, results were discussed to reach consensus (for further details, see pilot testing results below).

Participants' *satisfaction with the information* provided in the video was measured with 7 single items (5-point Likert scale). In total, 4 items were taken from the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-INFO25 survey (items 52-55) [72]. Items assessed participants' satisfaction with the *content* and *amount* of information provided by the hematologists; their *desire for more* or *less* information; the perceived *usefulness* and *clarity* of the information; and their satisfaction with the hematologist's *information giving style*. Participants were asked to explain their answer in an open-ended question box.

To assess participants' *trust* in the video hematologist, the 5-item short trust in oncologist scale (5-point Likert scale) was added [73].

Expected Symptom Distress

Expected symptom distress, that is, the perceived probability (1=very improbable to 5=very probable), severity (1=not at all severe to 5=very severe), and controllability (1=very little to 5 a lot) of physical as well as emotional distress was measured using separate items for each of the possible complaints discussed by the hematologist (fatigue, nausea, infections, and anxiety). Hemato-oncologists whom we consulted for this study differed in their opinion as to whether patients would appreciate information about the possible evolution of anxiety. For this reason, we added a single item (1=very unimportant to 5=very important) to measure the extent to which participants find it important that hematologists explicitly discuss possible feelings of anxiety and insecurity following diagnosis and treatment.

Pilot Testing

Survey validity and usability as well as the ecological validity of the video vignettes were pilot tested among 53 healthy participants aged 45 years and above. This age range was based on the incidence of non-Hodgkin lymphoma, which occurs mostly in older adults [74]. Participants were recruited in collaboration with Qualtrics panel services. In total, 500 invitation emails were sent out through the panel. A total of 145 (29.0%, 145/500) participants entered the survey, 87 (60.0%, 87/145) of which were filtered out as they did not match our participant requirements (Dutch; aged 45 years and older; and equal distribution of gender and region). Of the remaining 58 participants, a total of 53 (91%, 53/58) completed the survey including 1 random version of the 12 video vignettes.

The majority of participants (98%, 52/53) were Dutch. Participants (51%, 27/53 male) were on average aged 55.2 years (range 41 to 71); 64% (34/53) of participants indicated to have a partner. Participants had a diverse educational background; 42% (24/53) had a high school diploma; 30% (1/53) had completed vocational training; and 26% (14/53) had obtained a higher educational degree. In total, 45% (24/53) of the participants were not employed at the time of the survey. This was likely because of the average age of the target population.

Participants indicated to have little to average (medical) knowledge about cancer and lymphoma in particular (*cancer:* mean 2.68, range 1 to 5; *lymphoma*: mean 1.94, range 1 to 5). They judged their own health as average (mean 2.68; range 1 to 4) as compared with others their age. The majority of participants (79%) knew someone in their direct circle of friends and family who has (had) cancer. In total, 6 participants (11%) had received a cancer diagnosis (between 1993 and 2015), including skin cancer (4), breast cancer (1), and vocal cord cancer (1); two had received chemotherapy treatment.

Participants found the video realistic (mean 5.81; range 1 to 7), believable (mean 6.04; range 1 to 7), and the events displayed lifelike (mean 6.15; range 2 to 7). They found it easy to pay attention to the video (mean 5.85; range 1 to 7). More so, they perceived the physician as friendly (mean 5.96; range 1 to 7), likeable (mean 5.87; range 1 to 7), and credible in both his behavior (mean 5.94, range 3 to 7) and looks (mean 6.06, range 4 to 7). This provided support for the ecological validity of the video vignettes.

On average, participants found the physician's information structured (mean 8.13, range 3 to 10). They also deemed the amount of information provided by the physician quite adapted to the patient's needs (mean 7.72, range 2 to 10). Finally, participants perceived the physician as relatively empathetic (mean 7.57, range 3 to 10). Owing to the small group sizes (n=3 per condition), between-group differences were not tested. However, these overall scores suggested the potential for ceiling effects in the item's responses. Items were revised slightly to minimize these effects, but attention should be paid to this during the trial.

The recall instrument required revision (NL and ES), as some items appeared overly easy or complex. Adaptations resulted in a total pool of 28 items (14 active recall and 14 recognition).

For active recall, possible scores now range from 0 to 33 for recognition from 0 to 14. The pilot test indicated sufficient variation for information satisfaction, physician trust, and expected symptom distress.

Taken together, the results of the pilot test and subsequent revisions support the start of the trial. The procedures of the planned trial are further detailed below.

Participant Timeline

Participant recruitment and data collection are expected to last up to 2 months to reach the required sample size. This is considered feasible, based on previous experience using participant panels for study recruitment.

Sample Size

The required sample size is estimated at N=420 participants (*structuring* N=180 and *tailoring* N=240), based on *a priori* power analyses in G*Power [75] with the alpha set at .05, a probability level of .80, and estimated medium effect sizes of .10 to .25 for the dependent variables, that is, information recall, information satisfaction, and trust in the physician.

Recruitment Procedures

First, members of the PanelCom panel will be invited to participate in the experiment via mass emailing, receiving up to 2 reminders [65]. Second, (former) patients with cancer and their family members will be recruited in collaboration with several cancer patient support organizations, including the Dutch Cancer Society and Hematon, the Dutch association for patients with blood cancer and lymphoma. Through these organizations, potential participants will be informed about the study and invited to sign up for participation.

Participants are informed that the study is part of a research project about information giving in the context of cancer treatment. Furthermore, they are informed that study participation includes an online survey and a scripted video of a hematology consultation that will take approximately 30 min to complete and can be entered from a home computer. Participants are asked to complete the survey individually and in one sitting. Finally, participants will be notified that all data will be treated confidentially and remain anonymous. Participants provide informed consent upon entering the online survey.

Allocation and Blinding

Participants are automatically and randomly assigned to 1 of 12 conditions, or interventions, (1:1 ratio) in either the structuring or the tailoring experiment. Allocation is achieved by computer-generated randomization. Participants are unaware to which condition, that is, intervention, they are assigned (blinded). Researchers are unaware, for the duration of the trial, to which condition participants are assigned.

Analysis

Data will be cleaned in a stepwise procedure [76], including the identification of missing values because of dropout, standardization and normalization of data, and outlier analysis. Data will subsequently be analyzed using IBM SPSS statistical package. In the first step, it will be determined whether the experimental groups, within each of the 2 subexperiments of the trial, are indeed comparable in terms of participant characteristics, such as sociodemographics, personality traits, and disease history. If differences are found, these will be controlled for in subsequent analyses. Then, for each of the 2 experiments, one-way analysis of variance or covariance (when aspects need to be controlled for) will be conducted to test the effects of information structuring, information tailoring, and caring on information recall, satisfaction, and trust (H1, H2, and H3, respectively). Additionally, the interaction effect between caring and information structuring as well as tailoring will be added to these models (RQ1). It will be investigated to what extent patient characteristics moderate the hypothesized relationships (RQ3). If necessary, because of violations of the assumption of homogeneity of variance, Welch F statistic will be employed. Posthoc comparisons, using Bonferroni or Games-Howell, will be used as applicable, to create a better understanding of between-group differences. Finally, linear regression analyses will be used to assess possible relationships of information recall, satisfaction, and trust with expected symptom distress (RQ2). Significance levels are determined at $P<.05$.

Results

Data collection has now been completed. A total of N=607 participants went to our homepage, provided informed consent online, and started the survey. A total of N=470 completed the first part of the survey and were randomized to receive 1 of the 12 video vignette interventions within 1 of the 2 experiments (77.4%, 470/607). Participants did not differ from those who dropped out, except for their age (in experiment 1, N=148): completers were younger (mean 3.8 years; $P=.002$) and consequently less likely to be retired (41.2%, 194/470) vs 56.4% (83/148; $P=.006$). The first analyses will be available in Spring 2019.

Discussion

Strengths and Limitations

This study protocol describes the procedures for a randomized controlled trial in which 2 video vignette experiments are used to test the effects of physician information giving about side effects of cancer treatment on patient outcomes. Specifically, the effects of cognitive-oriented communication strategies (ie, information structuring and tailoring) as well as affect-oriented strategies (ie, caring) on patient recall, satisfaction, and trust are tested in conjunction. The outlined approach has both advantages and limitations.

Video vignette experiments allow researchers to experimentally test the causal relationships between communicative behaviors and consultation outcomes. This is particularly relevant when systematic manipulation of physicians' behaviors is undesirable for ethical or practical reasons. This trial thus has the potential to yield critical evidence to support interventions to change communicative behaviors in clinical practice. In the preparatory phase of the trial, the video vignettes were carefully developed in a stepwise procedure, involving a panel of hematologists, patients with cancer, health communication researchers, and

medical psychologists as well as a pilot test. Through this procedure, the script and its manipulations were thoroughly evaluated to ensure vignette realism as well as the effectiveness of the separate manipulations.

However, it should be noted that design artificiality can hamper ecological validity and that the use of analogue patients can hinder participants' ability to identify with the portrayed clinical situation. To ensure vignette realism, the script was based on a transcript of a full-length hematology consultation. As such, the script sought to represent a true-to-life hematology consultation rather than an ideal situation. Duration differences between the different versions of vignettes might account for differences in outcomes rather than the manipulation. However, duration differences are characteristic for realistic consultations, and compensating for these differences by adding *fillers* to the script may produce its own, undesirable, effects [63].

The pilot study demonstrated that participants were indeed able to identify and engage with the video patient. Inclusion of study participants who have previous experience with oncology consultations is expected to further improve identification with the vignettes, although our research group previously found no difference in identification between patients with cancer and cancer naïve participants [63].

As participants will be recruited via a panel of (former) patients with cancer and their relatives, as well as via patient organizations, it should be taken into account that panel participants may not be fully representative of the patient population. However, this is not deemed problematic as we primarily aim to identify pathways underlying effective information giving rather than to generalize patient outcomes to the population. The use of participants with previous experience with cancer does raise ethical concerns, as participants may experience feelings of anxiousness or sadness as a result of viewing a video in which a cancer treatment plan is discussed. This was reviewed by the institutional ethics committee. Participants are extensively debriefed following the experiment to minimize any negative impact of study participation.

Finally, it should be noted that, although video vignettes provide an effective method to study communication effects among (oncology) patients and their relatives, sometimes the use of so-called analogue patients poses a challenge. When striving to test the effects of information tailoring, the communication in the scripted vignettes is tailored to the video-patient rather than to the study participant. This may have implications for the findings. Direct effects of tailoring on participants' recall of information cannot be assumed. To overcome this issue, we added an item to the survey, assessing participants' personal information preferences (amount). Consequently, we can control for this variable in our statistical models.

Implications

The results of the proposed trial will provide evidence concerning the pathways linking physician communication to (improved) consultation outcomes for patients. In particular, the relationships between physicians' information structuring and tailoring (cognition-oriented skills) and caring (affect-oriented skill) and patients' recall (cognitive outcome), satisfaction, trust in the physician, and—ultimately—symptom distress (affective outcomes) will be clarified. The trial will allow researchers to further define what effective information provision about treatment precisely entails. Thereby, this study is highly relevant for patient-provider research in oncology settings. However, there are also practical implications. The results can be used to improve medical education about information provision. Within the scope of this study, it is indeed aimed to develop an innovative, evidence-based training module for hematologists about treatment information provision in cancer care [77]. The ultimate aim of this study is to contribute to our understanding of how oncologists can best inform patients about future symptoms to eventually improve patients' well-being and minimize potential suffering.

Conflicts of Interest

None declared.

References

1. Evans BD. The experiences and needs of patients attending a cancer support group. Int J Palliat Nurs 1995 Oct 02;1(4):189-194. [doi: 10.12968/ijpn.1995.1.4.189] [Medline: 29324115]
2. Fallowfield L, Ford S, Lewis S. No news is not good news: information preferences of patients with cancer. Psychooncology 1995 Oct;4(3):197-202. [Medline: 11655006]
3. Ishikawa H, Son D, Eto M, Kitamura K, Kiuchi T. The information-giving skills of resident physicians: relationships with confidence and simulated patient satisfaction. BMC Med Educ 2017 Feb 08;17(1):34 [FREE Full text] [doi: 10.1186/s12909-017-0875-6] [Medline: 28178986]

4. Jenkins V, Fallowfield L, Saul J. Information needs of patients with cancer: results from a large study in UK cancer centres. Br J Cancer 2001 Jan 05;84(1):48-51 [FREE Full text] [doi: 10.1054/bjoc.2000.1573] [Medline: 11139312]

5. Kessels RP. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222 [FREE Full text] [Medline: 12724430]

6. Mills M, Sullivan K. The importance of information giving for patients newly diagnosed with cancer: a review of the literature. J Clin Nurs 1999 Nov;8(6):631-642. [Medline: 10827609]

7. Strull W, Lo B, Charles G. Do patients want to participate in medical decision making? J Am Med Assoc 1984 Dec 07;252(21):2990-2994. [Medline: 6502860]

8. Jansen J, van Weert J, van der Meulen N, van Dulmen S, Heeren T, Bensing J. Recall in older cancer patients: measuring memory for medical information. Gerontologist 2008 Apr;48(2):149-157. [Medline: 18483427]

9. Sep MS, van Osch M, van Vliet LM, Smets EM, Bensing JM. The power of clinicians' affective communication: how reassurance about non-abandonment can reduce patients' physiological arousal and increase information recall in bad news consultations. An experimental study using analogue patients. Patient Educ Couns 2014 Apr;95(1):45-52 [FREE Full text] [doi: 10.1016/j.pec.2013.12.022] [Medline: 24485947]

10. van der Meulen N, Jansen J, van Dulmen S, Bensing J, van Weert J. Interventions to improve recall of medical information in cancer patients: a systematic review of the literature. Psychooncology 2008 Sep;17(9):857-868. [doi: 10.1002/pon.1290] [Medline: 18050149]

11. Ong LM, Visser MR, Lammes FB, de Haes JC. Doctor-patient communication and cancer patients' quality of life and satisfaction. Patient Educ Couns 2000 Sep;41(2):145-156. [Medline: 12024540]

12. Ley P, Bradshaw P, Eaves D, Walker C. A method for increasing patients' recall of information presented by doctors. Psychol Med 1973 May;3(2):217-220. [Medline: 4715854]

13. Williams S, Weinman J, Dale J. Doctor-patient communication and patient satisfaction: a review. Fam Pract 1998 Oct;15(5):480-492. [Medline: 9848436]

14. Rutten LJ, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). Patient Educ Couns 2005 Jun;57(3):250-261. [doi: 10.1016/j.pec.2004.06.006] [Medline: 15893206]

15. Gamble K. Communication and information: the experience of radiotherapy patients. Eur J Cancer Care (Engl) 1998 Sep;7(3):153-161. [Medline: 9793006]

16. Lamers RE, Cuypers M, Husson O, de Vries M, Kil PJ, Ruud Bosch JL, et al. Patients are dissatisfied with information provision: perceived information provision and quality of life in prostate cancer patients. Psychooncology 2016 Dec;25(6):633-640. [doi: 10.1002/pon.3981] [Medline: 26403417]

17. van Weert JC, Jansen J, de Bruijn GJ, Noordman J, van Dulmen S, Bensing JM. QUOTEchemo: a patient-centred instrument to measure quality of communication preceding chemotherapy treatment through the patient's eyes. Eur J Cancer 2009 Nov;45(17):2967-2976. [doi: 10.1016/j.ejca.2009.06.001] [Medline: 19615889]

18. Estacio CF, Butow PN, Lovell MR, Dong ST, Clayton JM. What is symptom meaning? A framework analysis of communication in palliative care consultations. Patient Educ Couns 2017 Nov;100(11):2088-2094. [doi: 10.1016/j.pec.2017.05.006] [Medline: 28619270]

19. de Haes H, Bensing J. Endpoints in medical communication research, proposing a framework of functions and outcomes. Patient Educ Couns 2009 Mar;74(3):287-294. [doi: 10.1016/j.pec.2008.12.006] [Medline: 19150197]

20. Peerdeman KJ, van Laarhoven AI, Peters ML, Evers AW. An integrative review of the influence of expectancies on pain. Front Psychol 2016;7:1270 [FREE Full text] [doi: 10.3389/fpsyg.2016.01270] [Medline: 27602013]

21. Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J. Influence of context effects on health outcomes: a systematic review. Lancet 2001 Mar 10;357(9258):757-762. [Medline: 11253970]

22. Hillen MA, de Haes HC, Smets EM. Cancer patients' trust in their physician-a review. Psychooncology 2011 Mar;20(3):227-241. [doi: 10.1002/pon.1745] [Medline: 20878840]

23. Caterinicchio RP. Testing plausible path models of interpersonal trust in patient-physician treatment relationships. Soc Sci Med Med Psychol Med Sociol 1979 Jan;13(1):81-99. [doi: 10.1016/0271-7123(79)90011-7] [Medline: 551529]

24. Merluzzi TV, Nairn R, Hegde K, Martinez Sanchez MA, Dunn L. Self-efficacy for coping with cancer: revision of the Cancer Behavior Inventory (version 2.0). Psychooncology 2001;10(3):206-217 [FREE Full text] [Medline: 11351373]

25. Neumann M, Wirtz M, Bollschweiler E, Mercer SW, Warm M, Wolf J, et al. Determinants and patient-reported long-term outcomes of physician empathy in oncology: a structural equation modelling approach. Patient Educ Couns 2007 Dec;69(1-3):63-75. [doi: 10.1016/j.pec.2007.07.003] [Medline: 17851016]

26. Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [Medline: 7792630]

27. Bartlett EE, Grayson M, Barker R, Levine DM, Golden A, Libber S. The effects of physician communications skills on patient satisfaction; recall, and adherence. J Chronic Dis 1984;37(9-10):755-764. [Medline: 6501547]

28. Langewitz W, Ackermann S, Heierle A, Hertwig R, Ghanim L, Bingisser R. Improving patient recall of information: Harnessing the power of structure. Patient Educ Couns 2015 Jun;98(6):716-721. [doi: 10.1016/j.pec.2015.02.003] [Medline: 25770048]

29. Linn AJ, van Dijk L, Smit EG, Jansen J, van Weert JC. May you never forget what is worth remembering: the relation between recall of medical information and medication adherence in patients with inflammatory bowel disease. J Crohn's Colitis 2013;7(11):e543-e550. [doi: 10.1016/j.crohns.2013.04.001] [Medline: 23660489]

30. McGuire LC. Remembering what the doctor said: organization and adults' memory for medical information. Exp Aging Res 1996;22(4):403-428. [doi: 10.1080/03610739608254020] [Medline: 8968711]

31. Fogarty LA, Curbow BA, Wingard JR, McDonnell K, Somerfield MR. Can 40 seconds of compassion reduce patient anxiety? J Clin Oncol 1999 Jan;17(1):371-379. [doi: 10.1200/JCO.1999.17.1.371] [Medline: 10458256]

32. Hillen MA, de Haes HC, Stalpers LJ, Klinkenbijl JH, Eddes EH, Butow PN, et al. How can communication by oncologists enhance patients' trust? An experimental study. Ann Oncol 2014 Apr;25(4):896-901. [doi: 10.1093/annonc/mdu027] [Medline: 24615411]

33. Hillen MA, de Haes HC, van Tienhoven G, van Laarhoven HW, van Weert JC, Vermeulen DM, et al. Oncologists' non-verbal behavior and analog patients' recall of information. Acta Oncol 2016 Jun;55(6):671-679. [doi: 10.3109/0284186X.2015.1120884] [Medline: 27031166]

34. van Osch M, Sep M, van Vliet LM, van Dulmen S, Bensing JM. Reducing patients' anxiety and uncertainty, and improving recall in bad news consultations. Health Psychol 2014 Nov;33(11):1382-1390. [doi: 10.1037/hea0000097] [Medline: 25089344]

35. Visser LN, Tollenaar MS, van Doornen LJ, de Haes HC, Smets EM. Does silence speak louder than words? The impact of oncologists' emotion-oriented communication on analogue patients' information recall and emotional stress. Patient Educ Couns 2019 Dec;102(1):43-52. [doi: 10.1016/j.pec.2018.08.032] [Medline: 30224275]

36. Atherton K, Young B, Salmon P. Understanding the information needs of people with haematological cancers. A meta-ethnography of quantitative and qualitative research. Eur J Cancer Care (Engl) 2017 Nov;26(6). [doi: 10.1111/ecc.12647] [Medline: 28185337]

37. Meyer BJ. The Organization of Prose and Its Effects on Memory. New York, NY: North-Holland Publishing Co; 1975:249.

38. León JA, Carretero M. Signal effects on the recall and understanding of expository texts in expert and novice readers. In: Oliveira A, editor. Hypermedia Courseware: Structures of Communication and Intelligent Help. Berlin Heidelberg: Springer - Verlag; 1992:97-111.

39. Rimer B, Kreuter M. Advancing tailored health communication: a persuasion and message effects perspective. J Communic 2006;56(Suppl 1):S184-S201. [doi: 10.1111/j.1460-2466.2006.00289.x]

40. Smit ES, Linn AJ, van Weert JC. Taking online computer-tailoring forward: The potential of tailoring the message frame and delivery mode of online health behaviour change interventions. Eur Health Psychol 2015;17(1):25-31 [FREE Full text]

41. Kreuter M, Wray R. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27(Suppl 3):S227-S232 [FREE Full text] [Medline: 14672383]

42. Petty RE, Cacioppo JT. Communication and Persuasion: Central and Peripheral Routes to Attitude Change. New York: Springer-Verlag; 1986.

43. Butow PN, Maclean M, Dunn SM, Tattersall MH, Boyer MJ. The dynamics of change: cancer patients' preferences for information, involvement and support. Ann Oncol 1997 Sep;8(9):857-863. [Medline: 9358935]

44. Leydon GM, Boulton M, Moynihan C, Jones A, Mossman J, Boudioni M, et al. Cancer patients' information needs and information seeking behaviour: in depth interview study. Br Med J 2000 Apr 01;320(7239):909-913 [FREE Full text] [Medline: 10742000]

45. Davey HM, Butow PN, Armstrong BK. Cancer patients' preferences for written prognostic information provided outside the clinical context. Br J Cancer 2003 Oct 20;89(8):1450-1456 [FREE Full text] [doi: 10.1038/sj.bjc.6601287] [Medline: 14562016]

46. Skalla KA, Bakitas M, Furstenberg CT, Ahles T, Henderson JV. Patients' need for information about cancer therapy. Oncol Nurs Forum 2004;31(2):313-319. [doi: 10.1188/04.ONF.313-319] [Medline: 15017447]

47. Adler HM. The sociophysiology of caring in the doctor-patient relationship. J Gen Intern Med 2002 Nov;17(11):874-881 [FREE Full text] [Medline: 12406360]

48. Medendorp NM, Visser LN, Hillen MA, de Haes HC, Smets EM. How oncologists' communication improves (analogue) patients' recall of information. A randomized video-vignettes study. Patient Educ Couns 2017;100(7):1338-1344. [Medline: 28233585]

49. Reblin M, Ellington L, Latimer S, Uchino BN, Roter D, Maxwell A. Communication style affects physiological response in simulated cancer risk reduction interactions. J Appl Biobehav Res 2012;17(2):129-156. [doi: 10.1111/j.1751-9861.2012.00080.x]

50. Schwabe L, Wolf OT. Learning under stress impairs memory formation. Neurobiol Learn Mem 2010 Feb;93(2):183-188. [doi: 10.1016/j.nlm.2009.09.009] [Medline: 19796703]

51. Salander P, Henriksson R. Severely diseased lung cancer patients narrate the importance of being included in a helping relationship. Lung Cancer 2005 Nov;50(2):155-162. [doi: 10.1016/j.lungcan.2005.05.023] [Medline: 16043257]

52. Potter J. Fatigue experience in advanced cancer: a phenomenological approach. Int J Palliat Nurs 2004 Jan;10(1):15-23. [doi: 10.12968/ijpn.2004.10.1.12015] [Medline: 14966441]

XSL•FO
RenderX

53. Roussi P, Miller SM. Monitoring style of coping with cancer related threats: a review of the literature. J Behav Med 2014;37(5):931-954 [FREE Full text] [doi: 10.1007/s10865-014-9553-x] [Medline: 24488543]

54. van Vliet L, van der Wall E, Plum NM, Bensing JM. Explicit prognostic information and reassurance about nonabandonment when entering palliative breast cancer care: findings from a scripted video-vignette study. J Clin Oncol 2013 Sep 10;31(26):3242-3249. [doi: 10.1200/JCO.2012.45.5865] [Medline: 23940230]

55. Verheul W, Sanders A, Bensing J. The effects of physicians' affect-oriented communication style and raising expectations on analogue patients' anxiety, affect and expectancies. Patient Educ Couns 2010 Sep;80(3):300-306. [doi: 10.1016/j.pec.2010.06.017] [Medline: 20638815]

56. Labrie NH, Schulz PJ. The effects of general practitioners' use of argumentation to support their treatment advice: results of an experimental study using video-vignettes. Health Commun 2015;30(10):951-961. [doi: 10.1080/10410236.2014.909276] [Medline: 25256304]

57. Saha S, Beach MC. The impact of patient-centered communication on patients' decision making and evaluations of physicians: a randomized study using video vignettes. Patient Educ Couns 2011 Sep;84(3):386-392 [FREE Full text] [doi: 10.1016/j.pec.2011.04.023] [Medline: 21600723]

58. Blanch-Hartigan D, Hall JA, Krupat E, Irish JT. Can naive viewers put themselves in the patients' shoes?: reliability and validity of the analogue patient methodology. Med Care 2013 Mar;51(3):e16-e21. [doi: 10.1097/MLR.0b013e31822945cc] [Medline: 22498688]

59. McKinstry B. Do patients wish to be involved in decision making in the consultation? A cross sectional survey with video vignettes. Br Med J 2000 Oct 07;321(7265):867-871 [FREE Full text] [Medline: 11021866]

60. Visser LN, Hillen MA, Verdam MG, Bol N, de Haes HC, Smets EM. Assessing engagement while viewing video vignettes; validation of the Video Engagement Scale (VES). Patient Educ Couns 2016 Feb;99(2):227-235. [doi: 10.1016/j.pec.2015.08.029] [Medline: 26395312]

61. Hillen MA, van Vliet LM, de Haes HC, Smets EM. Developing and administering scripted video vignettes for experimental research of patient-provider communication. Patient Educ Couns 2013 Jun;91(3):295-309. [doi: 10.1016/j.pec.2013.01.020] [Medline: 23433778]

62. van Vliet LM, Hillen MA, van der Wall E, Plum N, Bensing JM. How to create and administer scripted video-vignettes in an experimental study on disclosure of a palliative breast cancer diagnosis. Patient Educ Couns 2013 Apr;91(1):56-64. [doi: 10.1016/j.pec.2012.10.017] [Medline: 23219482]

63. Visser LN, Tollenaar MS, Bosch JA, van Doornen LJ, de Haes HC, Smets EM. Analogue patients' self-reported engagement and psychophysiological arousal in a video-vignettes design: Patients versus disease-naïve individuals. Patient Educ Couns 2016 Dec;99(10):1724-1732. [doi: 10.1016/j.pec.2016.04.012] [Medline: 27157320]

64. van Bruinessen IR, van Weel-Baumgarten EM, Snippe HW, Gouw H, Zijlstra JM, van Dulmen S. Active patient participation in the development of an online intervention. JMIR Res Protoc 2014 Nov 06;3(4):e59 [FREE Full text] [doi: 10.2196/resprot.3695] [Medline: 25379679]

65. Amsterdam Center for Health Communication. Wat is PanelCom? URL: http://www.healthcommunication.nl/panelcom/wat-is-panelcom/ [accessed 2019-03-17] [WebCite Cache ID 76wKa4Vwf]

66. Meyer BJ, Freedle RO. Effects of discourse type on recall. Am Educ Res J 1984 Jan 01;21(1):121-143. [doi: 10.3102/00028312021001121]

67. Meyer BJ, Poon LW. Effects of structure strategy training and signaling on recall of text. J Educ Psychol 2001;93(1):141-159. [doi: 10.1037/0022-0663.93.1.141]

68. van der Vaart R, Drossaert CH, Taal E, ten Klooster PM, Hilderink-Koertshuis RT, Klaase JM, et al. Validation of the Dutch functional, communicative and critical health literacy scales. Patient Educ Couns 2012 Oct;89(1):82-88. [doi: 10.1016/j.pec.2012.07.014] [Medline: 22878030]

69. van der Ploeg E, Mooren TT, Kleber RJ, van der Velden PG, Brom D. Construct validation of the Dutch version of the impact of event scale. Psychol Assess 2004 Mar;16(1):16-26. [doi: 10.1037/1040-3590.16.1.16] [Medline: 15023089]

70. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. Manual for the State-Trait Anxiety Inventory, STAI (form Y). Palo Alto, CA: Mind Garden; 1983.

71. Ong LM, Visser MR, van Zuuren FJ, Rietbroek RC, Lammes FB, de Haes JC. Cancer patients' coping styles and doctor-patient communication. Psychooncology 1999;8(2):155-166. [doi: 10.1002/(SICI)1099-1611(199903/04)8:2<155::AID-PON350>3.0.CO;2-A] [Medline: 10335559]

72. Arraras JI, Greimel E, Sezer O, Chie WC, Bergenmar M, Costantini A, et al. An international validation study of the EORTC QLQ-INFO25 questionnaire: an instrument to assess the information given to cancer patients. Eur J Cancer 2010 Oct;46(15):2726-2738. [doi: 10.1016/j.ejca.2010.06.118] [Medline: 20674333]

73. Hillen MA, Postma RM, Verdam MG, Smets EM. Development and validation of an abbreviated version of the Trust in Oncologist Scale-the Trust in Oncologist Scale-short form (TiOS-SF). Support Care Cancer 2017 Dec;25(3):855-861 [FREE Full text] [doi: 10.1007/s00520-016-3473-y] [Medline: 27830396]

74. cijfers over kanker [Figures About Cancer]. Dutch Cancer Registration URL: https://www.cijfersoverkanker.nl/ [accessed 2019-03-17] [WebCite Cache ID 76wLU9Vk0]

75.   Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]

76.   Morrow J. SlideShare. Brief Introduction to the 12 Steps of Evaluation Data Cleaning URL: http://www.slideshare.net/jamorrow/brief-introduction-to-the-12-steps-of-evaluagio [accessed 2019-03-17] [WebCite Cache ID 76wLju3hj]

77.   Stuij SM, Labrie NH, van Dulmen S, Kersten MJ, Christoph N, Hulsman RL, INSTRUCT project group. Developing a digital communication training tool on information-provision in oncology: uncovering learning needs and training preferences. BMC Med Educ 2018 Sep 24;18(1):220 [FREE Full text] [doi: 10.1186/s12909-018-1308-x] [Medline: 30249221]

Abbreviations

DLBCL: diffuse large B-cell lymphoma

XSL•FO

RenderX

Protocol

Communicating Uncertainty From Limitations in Quality of Evidence to the Public in Written Health Information: Protocol for a Web-Based Randomized Controlled Trial

Roland Brian Büchter[1], MPH; Cornelia Betsch[2], PhD; Martina Ehrlich[1], Dr oec troph; Dennis Fechtelpeter[1], MSc; Ulrich Grouven[1], Dr rer biol hum; Sabine Keller[1], Dipl Soz; Regina Meuer[1], MPH; Constanze Rossmann[2], PhD; Andreas Waltering[1], MD

[1]Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany

[2]Media and Communication Science, University of Erfurt, Erfurt, Germany

Corresponding Author:
Roland Brian Büchter, MPH
Institute for Quality and Efficiency in Health Care (IQWiG)
Im Mediapark 8
Cologne, 50670
Germany
Phone: 49 35685 ext 413
Email: roland.buechter@iqwig.de

Abstract

Background: Uncertainty is integral to evidence-informed decision making and is of particular importance for preference-sensitive decisions. Communicating uncertainty to patients and the public has long been identified as a goal in the informed and shared decision-making movement. Despite this, there is little quantitative research on how uncertainty in health information is perceived by readers.

Objective: The objective of this study is to design an experiment to examine how different degrees of uncertainty (Q1) and different types of uncertainty (Q2) impact patients' perception of treatment effectiveness, the body of evidence, text quality, and hypothetical treatment intention. The experiment also examines whether there is an additive effect when multiple sources of uncertainty are communicated (Q3).

Methods: We developed 8 variations of a research summary set in a hypothetical scenario for a treatment decision in the context of tinnitus. These were modified only in the degree of uncertainty relating to the evidence of the presented treatment. We recruited members of the German public from a Web-based research panel and randomized them to one of 8 variations of the research summary to examine the 3 research questions. The trial was only open to the members of the research panel. The outcomes are perception of the effectiveness of the treatment (primary), certainty in the judgement of treatment effectiveness, perception of the body of evidence relating to the treatment, text quality, and decisional intention (secondary). Outcomes were self-assessed. We aimed to recruit 1500 participants to the trial. The recruitment and data collection was fully automated. Ethical approval was waivered by an ethics committee because of the negligible risk to participants.

Results: This protocol is retrospectively published in its original format. In the meantime, the trial was set up and the data collection was completed. Data collection was conducted in May 2018. A total of 1727 eligible panel members were enrolled.

Conclusions: We aim to publish the results in a peer-reviewed journal by the end of 2019. In addition, results will be presented at conferences and disseminated among developers of guidance for the development of evidence-based health information and decision aids.

Trial Registration: German Clinical Trials Register DRKS00015911; https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00015911 (Archived by WebCite at http://www.webcitation.org/77zyZTGzk)

International Registered Report Identifier (IRRID): DERR1-10.2196/13425

XSL•FO
RenderX

Introduction

Background

Uncertainty pervades health care and is integral to evidence-informed decision making. The many layers of uncertainty, however, have hampered a common understanding of the subject. Han et al have developed a helpful taxonomy by identifying various types of uncertainty and classifying them into 3 dimensions [1]:

- Sources of uncertainty: These include, for example, ambiguity arising from conflicting evidence or statistical uncertainty.
- Issues arising from uncertainty: These include difficulties in decision making resulting from scientific uncertainty, for example, regarding treatment effects.
- Their loci, that is, uncertainty may exist in the mind of the patient, the health care provider, or both.

In terms of these dimensions, our experiment examines how communication of scientific uncertainty affects the perception of treatment effectiveness by patients and the public. Helping patients and consumers to understand and deal with uncertainty has been identified as one of the goals of the shared decision making and informed choices movement [2]. Understanding uncertainty is of particular importance for preference-sensitive decisions, that is, when there is a close trade-off between benefits and harms, and patient values and preferences are highly variable. Communicating uncertainty, however, poses many difficulties. Often information providers have to decide which of the many sources of uncertainty are most relevant to patients. Selection is required to prevent information overload, which can prompt people to base their decisions on heuristics rather than evidence [3]. Furthermore, communication of uncertainty may also have detrimental effects, for example, by hampering understanding or decreasing the credibility of the information provider [4,5].

Research on how to communicate uncertainty regarding the benefits and harms of treatments to patients and the public is limited. A systematic review by the Agency for Healthcare Research and Quality identified 8 controlled studies with 9 comparisons, including 6 randomized trials [6]. Of these studies, 4 examined statistical uncertainty, 4 studied different ways of communicating net benefit, and 1 addressed uncertainty arising from the use of a surrogate outcome. These studies were very heterogeneous in terms of context (including cancer screening and treatment decision making), interventions (including written information, drug fact boxes, and multifaceted interventions), and outcomes (including risk perception and decision making).

Objectives

We are not aware of any studies investigating whether perceptions of uncertainty depend on the degree, type, or amount of uncertainty presented in written health information. Thus, we decided to address the following 3 questions in our study:

1. Degree of uncertainty: Do members of the public perceive treatment effects differently depending on the choice of words used to express the certainty of those treatment effects?
2. Type of uncertainty: Do members of the public interpret uncertain treatment effects differently depending on the type of uncertainty?
3. Number of sources of uncertainty: Is there an additive effect of multiple sources of uncertainty?

We investigated these questions using 8 variations of a written piece of hypothetical consumer health information (research summary) set in a treatment decision scenario in the context of tinnitus. The research summaries were presented to a broad group of members of the German public using a Web-based research panel. Although the study was conducted with Web-based health information, we believe the results will be applicable to all types of written health information, including printed material.

We designed the experiment as a Web-based randomized superiority study, with 8 parallel groups allocated in an equal ratio (between-group design).

Methods

Procedures

We recruited members of the public from a Web-based research panel. Panel members had to be at least aged 18 years and able to read and write German. No other inclusion restrictions were applied.

The participants were first provided with a short introduction to the study and an informed consent sheet. We then collected information on age, sex, and educational degree. Participants were then asked to imagine having tinnitus and having unsuccessfully tried several treatments (see Multimedia Appendix 1). They were then randomly presented with one version of different variations of the research summary on the internet. These contained information on the medical condition and a short summary of evidence for a fictitious new tinnitus medication called *Oroxil* (see Multimedia Appendix 1). After presenting participants with the research summaries, we collected data on different outcomes using a questionnaire developed for the purpose of this study. We asked participants to return to the research summary as needed while answering the questions. At the end of the experiment, participants were asked about their profession (medical or nonmedical) and their history of tinnitus (present or not present). Participants were neither aware of the specific research questions nor of the alternative presentations. The original research summaries were written in German and translated into English for this publication.

Interventions

We chose a treatment scenario in the context of tinnitus and developed 8 variations of the research summary based on our

XSL•FO

RenderX

experience and use of language in providing evidence-based health information to consumers through Germany's statutory health website [7]. In accordance with the objectives of our study, we altered the research summary regarding the degree of uncertainty, the sources of uncertainty, and the number of sources of uncertainty. This resulted in 8 variations, two of which were used in 2 (statistically independent) comparisons (Table 1). An exemplary version of the research summary is provided in the Multimedia Appendix 1.

For the first objective of the study (Q1), we formulated 3 versions of the research summary with different degrees of uncertainty of the treatment effect. One version (A) describes a certain treatment effect and the other (B), a possible, but not certain treatment effect (indication of effect). The third version (B1) is identical to variation B but includes an additional statement on the need for further research. The semantic variations in the degrees of uncertainty of the treatment effect

were based on the methods for the assessment of treatment benefits developed by the German Institute for Quality and Efficiency in Health Care (IQWiG) [8].

For the second objective (Q2), we drew on the GRADE (Grading of Recommendations Assessment, Development and Evaluation) framework to categorize different types of uncertainty. According to GRADE, uncertainty can arise from risk of bias, (unexplained) inconsistency, indirectness, imprecision, and other threats to validity, such as publication bias or vested interests [9]. We therefore formulated 3 additional variations of the research summary describing publication bias and vested interests (B2), indirectness (B3), and imprecision (B4). We will also include variation B1 in this comparison.

For the third objective of the study (Q3), we developed 2 further variations that contained a combination of 2 or 3 sources of uncertainty (B42 and B432). A variation including only 1 source of uncertainty (B4) is included in this comparison.

Table 1. Variations of the research summary used to examine the 3 overarching research questions (translated from German).

Variations for research objective and identifier	Version	Variation in text
Q1. Degree of uncertainty		
A	Effect shown	Studies show that Oroxil can reduce tinnitus.
B	Indication of effect	Studies indicate that Oroxil may reduce tinnitus.
B1	Indication of effect with general explanation	Studies indicate that Oroxil may reduce tinnitus. (...) The pros and cons of Oroxil cannot be fully judged, however. This requires further research.
Q2: Type of uncertainty		
B1	Indication of effect with general explanation	As above
B2	Publication bias/vested interests	Studies indicate that Oroxil may reduce tinnitus. (...) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that the company that developed the drug has not published all the studies on Oroxil.
B3	Indirectness (population)	Studies indicate that Oroxil may reduce tinnitus. (...) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that the people who took part in the study were exposed to loud noises at work. It is uncertain whether the results also apply to other people.
B4	Imprecision (small sample size)	Studies indicate that Oroxil may reduce tinnitus. (...) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that only a small number of people took part in the studies.
Q3: Multiple sources of uncertainty		
B4	Imprecision (small sample size)	As above
B42	Publication bias/vested interests and imprecision	Studies indicate that Oroxil may reduce tinnitus. (...) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that only a small number of people took part in the studies. Furthermore, the company that developed the drug has not published all the studies on Oroxil.
B432	Publication bias/vested interests and imprecision and indirectness	Studies indicate that Oroxil may ease tinnitus. (...) The pros and cons of Oroxil cannot be fully judged, however. The reason for this is that the studies were small. Furthermore, the people who took part in the studies were exposed to loud noises at work. It is uncertain whether the results also apply to other people. Finally, the company that developed the drug has not published all studies on Oroxil

Outcomes

Our primary outcome is the perception of treatment effectiveness. The secondary outcomes are subjective certainty in the judgement of treatment effectiveness, perception of the body of evidence, hypothetical treatment intention, and perception of text quality.

The perception of treatment effectiveness was measured with 1 item on an ordinal scale (*How do you judge the effectiveness of Oroxil?*) with 5 possible answers: (a) It is proven that Oroxil can help; (b) Oroxil may possibly help; (c) It is unclear, whether Oroxil helps; (d) Oroxil may not help; and (e) Oroxil definitely does not help.

Subjective certainty in the judgement of treatment effectiveness was measured on a 5-point Likert scale, ranging from *not certain at all* to *very certain*. As this relates to the first question on perceptions of treatment effectiveness, data on this item were gathered immediately after answering the first question.

The perception of the body of evidence was measured with a 6-item semantic differential (Cronbach alpha=.81), with each item measured on a 5-point Likert scale. Participants were asked to rate the body of evidence as follows:

- Certain to uncertain
- Reliable to unreliable
- Valid to not valid
- Generalizable to not generalizable
- Excellent to poor
- Trustworthy to untrustworthy

The hypothetical treatment intention was measured using 1 item (*How would you decide?*) measured on a 5-point Likert scale with 2 poles: (a) definitely not take Oroxil and (b) definitely take Oroxil.

The perception of text quality was measured with a 9-item semantic differential (Cronbach alpha=.81) based on previous literature [10,11]. The construct included the following items measured on a 5-point Likert scale:

- Interesting to uninteresting
- Balanced to 1-sided
- Comprehensible to incomprehensible
- Credible to incredible
- Clear to unclear
- Well done to not well done
- Professional to unprofessional
- Appealing to not appealing
- Respectable to not respectable

Mediators (Explorative)

We collected data on the following possible mediating variables for the purpose of explorative analyses:

- Decisional conflict (German version of the uncertainty subscale of the decisional conflict scale, Cronbach alpha=.76) [12]
- Perceived knowledge about the treatment measured on a visual analogue scale ranging from 0 for *no knowledge* to 100 for *all possible knowledge* [13]

- Perceived sufficiency of knowledge about the treatment for decision making, measured on an ordinal scale with 3 possible answers (more knowledge, the amount of knowledge provided, and less knowledge)
- Perception of the credibility of the information provider (based on previous scales, Cronbach alpha=.93) [14,15]

Moderators (Explorative)

We collected data on the following possible moderator variables, again, for the purpose of explorative analyses:

- Sex
- Age
- Educational degree based on the German school system (none/basic secondary/higher secondary/general entry qualification for university/university degree)
- Subjective health literacy (using the German version of the Brief Health Literacy Screening Tool (known as BRIEF), Cronbach alpha=.76) [16]
- Numeracy (using the 1-item version of the Berlin Numeracy Test) [17]
- Objective subscale of the perceived uncertainty of scientific evidence scale (Cronbach alpha=.76) [18]
- Medical degree or profession (yes/no)
- Previous experience with tinnitus (history of tinnitus/currently symptomatic/never present)

We piloted a paper-and-pencil version of the questionnaire with 2 versions of the research summary in a convenience sample of 40 students to test the reliability of the constructs, comprehensibility of instructions, the stimuli, and the questions. The reliability of the constructs was good to very good as reported above (Cronbach alpha ranging from .76 to .93). On the basis of the pretest, we omitted 2 items from the pilot questionnaire for the outcome of perceived text quality to increase reliability. We also amended the instructions to improve comprehensibility.

Statistical Analysis

We will present demographic characteristics of the sample using frequencies, in case of categorical data, and measures of location (mean and median) and variation (SD, interquartile ranges [IQRs], and ranges), in case of continuous data.

We will treat the primary outcome variable as an ordinal scale with 5 possible values, where higher values indicate an increase in the perception of effectiveness (5=it has been proven that the treatment can help to 1=treatment definitely cannot help). We will present data as medians, IQRs, and ranges for each group. We will also present means and SDs, as well as the proportions for each possible answer in descriptive tables. This will also help to establish the practical relevance of the results.

For the secondary outcomes *perception of the body of evidence* and *text quality*, we will combine the items of each of the scales into 1 index by averaging their values, where a higher value will indicate better perception of the body of evidence or text quality.

For our confirmatory statistical analyses of the primary and secondary outcomes, we will conduct Kruskall-Wallis tests to test for overall differences between the groups within each of

our 3 primary study questions. We chose to use a nonparametric test to account for the types of scales used (ordinal scaling or unequal differences between items).

In case of statistical significance, we will conduct a multiple testing procedure to perform pairwise comparisons within the groups by means of the Dwass-Steel-Critchlow-Fligner multiple comparison analysis, which is based on pairwise 2-sample Wilcoxon comparisons. All comparisons between groups across the 3 overarching study questions will be considered explorative (eg, A vs B432).

We will conduct sensitivity analyses by means of an analysis of variance (ANOVA) to test for the overall differences between the groups within each of our 3 overarching study questions. In case of statistical significance, we will conduct pairwise comparisons by means of Tukey honestly significant difference procedure. We will inspect data to ensure that they meet distributional assumptions (normality and equal variance) before applying statistical tests.

Statistical analyses will be conducted in SAS version 9.4 (SAS Institute Inc). All statistical tests will be 2-sided and performed using a 5% significance level. Where applicable, differences in means between groups will be presented together with a 95% CI. All analyses will be conducted on an intention-to-treat basis. Explorative analyses based on potential moderators and mediators are not predetermined.

As we will collect outcome data immediately after the presentation of the research summaries and panel members need to finish the questionnaire to receive an incentive, we have no major concerns regarding missing data. Furthermore, we assume that any dropouts will be likely to be missing at random, as we believe it is unlikely that the intervention has an influence on the responses to the questionnaire. Thus, we do not plan to employ any imputation methods. In case of missing data, we will present this information descriptively.

As the experiment is Web-based and participants come from a panel that provides incentives for participation, there is a risk that some participants only participate to collect their incentive and do not provide valid answers. As a measure of quality assurance, we will exclude data from participants who answer all questions in less than 2 min, spend less than 20 seconds on the page displaying the research summary, and spend less than 1.5 min between reading the research summary and completing the questionnaire (so called speeders). These time limits were determined by a priori test readings. We will also exclude participants who provide all answers in the same column for the matrix questions, that is, when more than 1 item is displayed on the screen (so called straightliners).

Sample Size Calculation

We based the sample size calculations for all 3 research questions on the following considerations and assumptions. We used the comparison of 4 groups (which corresponds to the second study question) as a basis and proceeded from a 1-way ANOVA with equal sample sizes in each group. We assumed a significance level of 5% and a statistical power of 90%. We decided to assume an effect size of $F=0.15$ for the primary outcome (confirmatory analysis), where F denotes the ratio of

the SD of the group means and the common SD within each group. This decision was made based on a pretest of 2 of the research summaries, where we found small effect sizes in the range of up to $F=0.3$, depending on the outcome variable. According to Cohen, the chosen value of F lies between a small ($F=0.10$) and a medium ($F=0.25$) effect size [19]. Sample size calculations were conducted with nQuery version 3.0 (Statistical Solutions Ltd). This resulted in a number of 159 participants in each group. As the primary analysis is a nonparametric test, we added 15% according to a general rule of thumb [20]. To allow some leeway, we decided to randomize a total of 1500 participants, equaling an average of 187.5 participants per group.

Data Collection and Allocation Procedure

The data collection was run by the Survey Centre Bonn (uzbonn—Gesellschaft für empirische Sozialforschung und Evaluation), a spinout company of the Center for Evaluation and Methods at the University of Bonn. UNICOM Intelligence (formerly IBM SPSS Data Collection) was used for data collection (UNICOM Systems, Inc). This software uses Microsoft's .NET Framework 4.0 random generator to generate random numbers to allocate the participants to the research summaries. A quota was used to ensure equal representation of different age groups and sexes. Once a quota cell was full, enrollment for this quota was closed. Thus, allocation happened after panel members answered demographic questions and were computer-checked for eligibility. As the experiment was entirely Web-based, the allocation sequence was concealed from investigators and data collectors. The data analyses will be conducted by a statistician from the Medical Biometry Department at IQWiG.

Ethics and Dissemination

The study was presented to the ethics committee at the University of Erfurt (EV-20180921). The committee decided that the research was exempt from the requirement of ethical approval because of its negligible risk to participants and as only nonidentifiable data were collected. The study results will be disseminated via publication and conference presentations.

Results

The trial was set up between February and April 2018. Data collection was completed in May 2018. Recruitment and data collection were Web-based. First, a website, only accessible via a link available to invited panel members, was set up (password protected site). Font and color use matched the appearance of the national German consumer health website [7]. Then, 6 of the authors (RBB, ME, DF, UG, RM, and AW) read and reread the recruitment texts and the research summaries. The same authors also tested the questionnaire used for data collection. After 2 rounds of debugging, the website and questionnaire were finalized and data collection was started.

Participants were recruited by the Survey Centre Bonn from a Web-based panel of members of the German public. The panel was provided by the online access panel provider Bilendi. Participation was only permitted with use of a desktop computer.

Participants were first informed about the general purpose of the study (to study the perception of health information), the

duration required for answering the questionnaire, and the use of data. Only anonymous data were collected. The outcome data were collected over 10 consecutive screens. Participants were able to move forward and backward between the screens. However, once they completed the data collection for the primary and secondary outcomes, moving backward to that section of the questionnaire was not possible anymore. The order of items in the multiitem outcomes was presented in random order. We used soft reminders to encourage participants to answer all questions, that is, the participants were asked to complete unanswered questions before proceeding but were not obliged to answer them. The only mandatory questions were regarding age and sex, as this information was needed to check

eligibility. Repeated participation by the same panel member was prevented by the use of a password encrypted access link, which was provided to participants via email.

In total, 2099 invited panel members were assessed for eligibility and 1727 were randomized to 1 of the 8 groups.

Discussion

We aim to publish the results in a peer-reviewed journal by the end of 2019. In addition, results will be presented at conferences and disseminated among developers of guidance for the development of evidence-based health information and decision aids.

Acknowledgments

Authors' Contributions

RBB, ME, DF, SK, AW, and RM had the initial idea for the study, conceived the study design, developed the research summaries, and drafted a preliminary version of the questionnaire. CB and CR elaborated, extended, and pretested the questionnaire and commented on the study design. CB and UG performed sample size calculations. UG and RBB developed the statistical analysis plan. RBB drafted the first version of this manuscript. All authors critically reviewed and approved the final version.

Conflicts of Interest

RBB, ME, DF, UG, SK, RM and AW are employees of IQWiG.

Multimedia Appendix 1

Introduction text and exemplary research summary.

[PDF File (Adobe PDF File), 31KB - resprot_v8i5e13425_app1.pdf]

References

1. Han PK, Klein WM, Arora NK. Varieties of uncertainty in health care: a conceptual taxonomy. Med Decis Making 2011;31(6):828-838 [FREE Full text] [doi: 10.1177/0272989X10393976] [Medline: 22067431]
2. Institute of Medicine Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington (DC): National Academies Press (US); 2001. URL: https://www.ncbi.nlm.nih.gov/pubmed/25057539 [Medline: 25057539]
3. Gigerenzer G, Gaissmaier W. Heuristic decision making. Annu Rev Psychol 2011;62:451-482. [doi: 10.1146/annurev-psych-120709-145346] [Medline: 21126183]
4. Longman T, Turner RM, King M, McCaffery KJ. The effects of communicating uncertainty in quantitative health risk estimates. Patient Educ Couns 2012 Nov;89(2):252-259. [doi: 10.1016/j.pec.2012.07.010] [Medline: 22858415]
5. Sladakovic J, Jansen J, Hersch J, Turner R, McCaffery K. The differential effects of presenting uncertainty around benefits and harms on treatment decision making. Patient Educ Couns 2016 Dec;99(6):974-980. [doi: 10.1016/j.pec.2016.01.009] [Medline: 26823204]
6. McCormack L, Sheridan S, Lewis M, Boudewyns V, Melvin CL, Kistler C, et al. Communication and dissemination strategies to facilitate the use of health-related evidence. Evid Rep Technol Assess (Full Rep) 2013 Nov(213):1-520 [FREE Full text] [Medline: 24423078]
7. Informed Health. URL: https://www.informedhealth.org/our-approach.2059.en.html [accessed 2019-03-01] [WebCite Cache ID 76XrwKJhQ]
8. Institute for Quality and Efficiency in Health Care (IQWiG). General Methods 4.2. Cologne: Institute for Quality and Efficiency in Health (IQWiG); 2015. URL: https://www.ncbi.nlm.nih.gov/books/NBK332872 [Medline: 27403465]
9. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Br Med J 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]

XSL•FO
RenderX

10. Brosius H, Birk M. [Text-image correspondence and information through television news]. In: [Radio and Television]. Munich: Ludwig Maximilian University; 1994:171-183.

11. Gehrau V. [TV Rating and TV Act]. Munich: R Fischer; 2008.

12. Buchholz A, Hölzek L, Kriston L, Simon D, Härter M. [The Decisional Conflict Scale in German (DCS-D)-Dimensional structure in a sample of family doctor patients]. Klin Diagnostik Evaluation 2011;4:15-30.

13. Hwang Y, Jeong SH. Information insufficiency and information seeking: an experiment. Sci Commun 2016;38(6):679-698. [doi: 10.1177/1075547016673200]

14. Flanagin AJ, Metzger MJ. The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. New Media Soc 2007;9(2):319-342. [doi: 10.1177/1461444807075015]

15. Roobina O. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. J Advert 1990;19(3):39-52.

16. Haun J, Luther S, Dodd V, Donaldson P. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. J Health Commun 2012;17(Suppl 3):141-159. [doi: 10.1080/10810730.2012.712615] [Medline: 23030567]

17. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring risk literacy: the Berlin Numeracy test. Judgm Decis Mak 2012;7(1):25-47 [FREE Full text]

18. Retzbach J, Otto L, Maier M. Measuring the perceived uncertainty of scientific evidence and its relationship to engagement with science. Public Underst Sci 2016;25(6):638-655. [Medline: 25814513]

19. Cohen J. Statistical Power Analysis For The Behavioral Sciences (2nd Edition). United Kingdom: Routledge; 1988.

20. Lehman EL. Nonparametrics: Statistical Methods Based on Ranks. London: Pearson; 1998.

Abbreviations

ANOVA: analysis of variance
GRADE: Grading of Recommendations Assessment, Development and Evaluation
IQRs: interquartile ranges
IQWiG: the German Institute for Quality and Efficiency in Health Care

Protocol

A Mobile and Web-Based Self-Directed Complementary and Integrative Health Program for Veterans and Their Partners (Mission Reconnect): Protocol for a Mixed-Methods Randomized Controlled Trial

Jolie N Haun[1], EdS, PhD; Lisa M Ballistrea[1], MSPT, DPT; Christine Melillo[1], PhD, RN; Maisha Standifer[1], MPH, PhD; Kevin Kip[1,2], PhD; Jacquelyn Paykel[3], MD; Jennifer L Murphy[4], PhD; Carol E Fletcher[5], PhD, RN; Allison Mitchinson[5], MPH; Leila Kozak[6,7,8], PhD; Stephanie L Taylor[9,10], PhD; Shirley M Glynn[11], PhD; Matthew Bair[12,13,14], MS, MD

[1]Rehabilitation Outcomes Research Section, James A Haley Veterans' Hospital and Clinics, Veterans Health Administration, Tampa, FL, United States

[2]College of Public Health, University of South Florida, Tampa, FL, United States

[3]Whole Health Service, James A Haley Veterans' Hospital and Clinics, Veterans Health Administration, Tampa, FL, United States

[4]Mental Health and Behavioral Sciences Service, James A Haley Veterans' Hospital and Clinics, Veterans Health Administration, Tampa, FL, United States

[5]Veterans Affairs Ann Arbor Healthcare System, Veterans Health Administration, Ann Arbor, MI, United States

[6]Department of Family Medicine, University of Washington School of Medicine, University of Washington, Seattle, WA, United States

[7]Veterans Affairs Puget Sound Health Care System, Veterans Health Administration, Seattle, WA, United States

[8]Integrative Health Coordinating Center, Office of Patient Centered Care and Cultural Transformation, Veterans Health Administration, Washington, DC, United States

[9]Health Services Research and Development, Veterans Health Administration, Los Angeles, CA, United States

[10]Department of Health Policy and Research, University of California - Los Angeles, Los Angeles, CA, United States

[11]Research Service, Veterans Affairs Greater Los Angeles Healthcare System at West Los Angeles, Los Angeles, CA, United States

[12]School of Medicine, Indiana University, Indianapolis, IN, United States

[13]Regenstrief Institute, Inc, Indianapolis, IN, United States

[14]Center for Health Information and Communication, Veterans Affairs, Indianapolis, IN, United States

Corresponding Author:
Jolie N Haun, EdS, PhD
Rehabilitation Outcomes Research Section
James A Haley Veterans' Hospital and Clinics
Veterans Health Administration
13000 Bruce B Downs Blvd
Tampa, FL,
United States
Phone: 1 813 558 7622
Email: joliehaun@gmail.com

Abstract

Background: Complementary and integrative health (CIH) is a viable solution to PTSD and chronic pain. Many veterans believe CIH can be performed only by licensed professionals in a health care setting. Health information technology can bring effective CIH to veterans and their partners.

Objective: This paper describes the rationale, design, and methods of the Mission Reconnect protocol to deliver mobile and Web-based complementary and integrative health programs to veterans and their partners (eg, spouse, significant other, caregiver, or family member).

Methods: This three-site, 4-year mixed-methods randomized controlled trial uses a wait-list control to determine the effects of mobile and Web-based CIH programs for veterans and their partners, or dyads. The study will use two arms (ie, treatment intervention arm and wait-list control arm) in a clinical sample of veterans with comorbid pain and posttraumatic stress disorder,

and their partners. The study will evaluate the effectiveness and perceived value of the Mission Reconnect program in relation to physical and psychological symptoms, global health, and social outcomes.

Results: Funding for the study began in November 2018, and we are currently in the process of recruitment screening and data randomization for the study. Primary data collection will begin in May 2019 and continue through May 2021. Projected participants per site will be 76 partners/dyads, for a total of 456 study participants. Anticipated study results will be published in November 2022.

Conclusions: This work highlights innovative delivery of CIH to veterans and their partners for treatment of posttraumatic stress disorder and chronic pain.

Trial Registration: ClinicalTrials.gov NCT03593772; https://clinicaltrials.gov/ct2/show/NCT03593772 (Archived by WebCite at http://www.webcitation.org/77Q2giwtw)

International Registered Report Identifier (IRRID): PRR1-10.2196/13666

Introduction

Background

Chronic pain and posttraumatic stress disorder (PTSD) are highly prevalent comorbid conditions in the veteran population and present a challenge for traditional interventions [1,2]. Pharmacological options have potential consequences of opioid use disorder and overdose [3]. The Veterans Affairs (VA)'s Opioid Safety Initiative and the US Department of Health and Human Services 2016 National Pain Strategy prioritize the need for nonpharmacological pain treatment options that leverage self-management [4,5].The VA is invested in providing virtual services to improve veteran and family member access to care that promotes self- management and improves patient outcomes. The Veterans Health Administration (VHA) currently views touch-based therapies as interventions requiring professional delivery. Massage costs approximately US $60/hour [6], is primarily an out-of-pocket expense, and is often not affordable. Massage is not currently widely available in the VA; however, with the emerging Whole Health initiative, massage and other mind-body modalities will be in increasing demand. Simply put, the VA will not be able to meet the increasing demand for massage services.

Mission Reconnect (MR) provides a potentially low-cost, accessible, and sustainable intervention for veterans in home settings. The need for low-cost interventions to support well-being and optimal functioning among veterans and their families is clear [7]. Furthermore, the VHA's Secretary for Health's Critical Priorities for Strategic Action identified access, pain management, and putting veterans first to achieve health for veterans. To be responsive to these priorities, the 2016 VA State-of-the-Art Conference and VA Comprehensive Addiction and Recovery Act mandate [8] indicate VA's commitment to conduct rigorous research to integrate nonpharmacological and complementary and integrative health (CIH) approaches into care at every level, with emphasis on pain management. To support these priorities, the Office of Patient-Centered Care and Cultural Transformations' Whole Health Program provides an innovative approach to integrative care that leverages CIH and family participation. Finally, the Connected Care Office seeks

to support self-care management and health care using mobile technology.

Severity of outcomes, cost of care, and initiatives prioritizing access to CIH and whole health care warrant the need for a multidimensional approach, such as MR. The proposed research will advance science by (1) testing a safe self-directed sustainable CIH approach for treating a clinically defined high-risk population of veterans diagnosed with chronic pain and PTSD-related symptoms to improve psychological and physical outcomes, (2) introducing a nonpharmacologic chronic pain management program option for veterans, (3) testing the usefulness of MR as a remotely delivered Internet/mobile program delivered in the users' natural environment, (4) leveraging a partnered approach (eg, spouses or family members) to implementing interventions that address the biopsychosocial aspect of wellness, and (5) conducting longitudinal repeated-measures analyses, which are not common in CIH research.

Impact of Chronic Pain

Chronic pain is a highly prevalent condition among veterans (81.5%) [1]. Musculoskeletal ailments are some of the most frequent reasons that veterans seek care at the VA [2]. In addition to discomfort and mood and sleep disturbances associated with pain, veterans with chronic pain have a high coprevalence of medical, mental health, and substance use disorders [4]. Veterans with pain have higher prescribed opioid doses, which is associated with risk of accidental poisoning death and suicide death [6]. In sum, chronic pain is a high priority area in the VA due to its prevalence and severity of impact on quality of life for veterans and their families.

Compounded Impact of Chronic Pain and Comorbid Posttraumatic Stress Disorder on Outcomes

We chose to address chronic pain and PTSD due to their prevalence, compounded impact, and their priority in the VA. As many as 70% of veterans with chronic pain treated in the VA have PTSD, and up to 80% of those with PTSD have pain [2]. Prevalence of PTSD is higher in patients with chronic pain [8,9]. Patients with both PTSD and chronic pain generally present with more complicated clinical profiles [10], and patients

with these comorbidities report lower quality of life than their veteran counterparts [11]. Chronic pain and PTSD are associated with high rates of depression, anxiety, and fatigue [12-16], which detrimentally impact work, social functioning, relationships, independent living, and ability to enjoy life [17-19]. Effectiveness of pharmacotherapy is limited and can result in other negative consequences, such as substance use disorders [20]. Further, veterans with PTSD receive more frequent and higher-dose opioids for pain diagnoses than veterans without PTSD [6]. Innovative CIH approaches are needed to help veterans and their families cope with chronic pain and PTSD without the side effects and adverse events associated with pharmacological management [21-23].

Theoretical Mechanisms Connecting Pain and Posttraumatic Stress Disorder

The biopsychosocial model represents the complex interrelationships between physical, psychological, and social factors [24,25]. Within the context of our research, the application of the biopsychosocial model will center around the bidirectional relationship between pain and PTSD. Psychological trauma induces change in biological substrates, which in turn alter pain transduction pathways and pain processing mechanisms in the brain, intensifying pain experience in individuals with PTSD [2]. Neurologically, PTSD is characterized by hyperactivation of the amygdala and hippocampus, and lower activation and imbalance in the medial prefrontal cortex. The amygdala integrates nociceptive information and plays a dual faciliatory and inhibitory role in the modulation of emotional pain behavior [2]. Therefore, interventions that ameliorate pain may be expected to reduce symptoms of PTSD [26] and vice versa [2]. On the basis of this bidirectional biopsychosocial model, we propose a multidimensional CIH intervention to support positive multifactorial outcomes associated with comorbid chronic pain and PTSD.

Veterans' chronic conditions can affect their families' well-being. In alignment with the biopsychosocial model, we contend the complexity of comorbid chronic pain and PTSD does not end with the veteran. We recognize the critical role of the veteran's family in supporting veterans' well-being. Chronic conditions, such as comorbid pain and PTSD, can have substantial impacts on personal relationships. Studies have linked unhealthy family responses to poorer outcomes in the person with pain [27]. Maladaptive patterns of interaction may reinforce disability, fear of activity, and dependency in the patient, thereby inhibiting their recovery, rehabilitation, and treatment outcomes. Family members or other support partners may respond in a manner that is solicitous, thereby unintentionally reinforcing the sick role and disabled lifestyle of the person with pain [27]. Though partners often serve as a protective factor through social support and advocacy, problematic effects of caregiver burden are common [28,29]. Family and loved ones' involvement in treatment can help support positive outcomes [30-32].

Partner-Family–Assisted Interventions

Including partners-family in treatments has been conceptualized in four ways: (1) partner-assisted interventions, (2)

disorder-specific interventions, (3) general therapy, and (4) education-facilitated engagement [33,34]. MR qualifies as a partner-family assisted intervention, as MR is not disorder-specific nor a focused couples therapy nor primarily educational. Instead, it involves providing guidance and encouragement to the partners or family members so they can support the veterans' engagement in and experience of MR. The majority of family-partner–focused interventions focus on mental health and couples-family therapy [33,34]. MR is unique in its conjoint approach to supporting the veteran through the use of partnered CIH. A veteran-partnered sample participated in a multimodal intervention to address PTSD symptoms using CIH modalities for stress reduction and resource building. Findings indicate significant reduction in PTSD symptoms and benefits to their family members [35]. Though current research in partner-assisted CIH based programs is limited, these programs hold potential for supporting veterans with chronic conditions and their partners/families.

Introduction of an Innovative Partner-Family–Assisted Complementary and Integrative Health Intervention

Mission Reconnect (MR) provides a novel CIH approach that supports veterans' symptom management using evidence-based CIH modalities (ie, massage, meditation, positive psychology) presented in a self-paced partner-family–assisted program that can be accessed remotely. MR is distinctive in its:

1. Approach to treating chronic pain using a nonpharmacological CIH approach to managing chronic pain that has been shown successful in a community-based nonclinically defined cohort [36].
2. Use of CIH therapy skills education, which teaches massage techniques as a home-based, interpersonal skill between veterans and their selected partners.
3. Focus on the relationship dyad as the unit of intervention. The proposed *partner-family–assisted intervention* directly applies the biopsychosocial model, engaging veterans in their natural social contexts rather than relying solely on a clinical setting. Targeting the dyad leverages the natural resources of the relationship—trust, commitment, compassion, and mutual reinforcement of participation.

Complementary and Integrative Health Impact on Pain and Posttraumatic Stress Disorder Outcomes

Massage is the most preferred CIH modality of all complementary health approaches, with broad appeal among veterans [37]. Research indicates 82% of veterans with chronic pain reported use of at least one CIH therapy and nearly all (99%) were willing to try such approaches [37]. These findings are consistent with military treatment facilities that report CIH is most often used use for pain and mental health conditions [38]. The preference for massage is intuitive given evidence suggesting the therapeutic benefits of massage including reduction of pain, stress, depressive symptoms, anxiety, and improvement of sleep across diverse populations [36,39-43]. A 2016 VA evidence-based synthesis report identified 21 high-quality systematic reviews on massage for pain. Findings described potential benefits of massage, but evidence strength is limited due to methods used in reviewed studies [44]. An independent meta-analysis of randomized controlled trials

(RCTs) addressing pain concluded that massage effectively reduced pain compared to sham and active comparators, and improved mood and health-related quality of life compared to active comparators, and concluded that massage is a viable pain management option [45]. A limited number of massage studies have shown results on PTSD, but many studies have demonstrated results on related symptoms such as anxiety, stress, and depression [40,42,46-48].

Though limited research connects the impact of massage on PTSD outcomes, a secondary analysis of four trials with veterans with PTSD suggests that mindfulness-based stress reduction, another MR component, is a viable mechanism of treatment [49]. Mindfulness reduces pain for veterans with chronic pain [50] and improves anxiety, depression, and suicidal ideations [51]. Based on a VA review of the current evidence and best practices, VA/Department of Defense (DOD) PTSD clinical practice guidelines suggest that, although CIH is not indicated as primary treatment, it holds promise to improve wellness and promote recovery [52]. Even with these clinical guidelines, it is noted that study limitations leave current evidence inconclusive [52]. Another systematic review of CIH among veterans and military personnel indicated benefits from mind-body modalities; however, the quality of most RCTs was rated poorly [53]. The evidence base regarding the effectiveness of CIH interventions [44] for reducing pain and PTSD symptoms in veterans is inconclusive; this study will fill the gap in this area of research [54]. This project will contribute to the knowledge base of this field of research by using strong methodology related to sample size, contextual specificity of the target population, RCT design, and measurement of long-term outcomes.

Methods

Conceptual Framework

In our conceptual model (Figure 1), deployment impacts on veterans and their families are summarized in the left box while target biopsychosocial outcomes in the boxes on the right are the expected effects of the proposed MR intervention, while controlling for previous and current treatment history (non-MR). MR has potential to provide a critically needed CIH option to manage deployment-related impacts, such as pain symptoms, sleep issues, and relationship issues.

Figure 1. Conceptual model. CIH: complementary and integrative health; H: hypothesis; MR: Mission Reconnect; PTSD: posttraumatic stress disorder; RQ: research question.
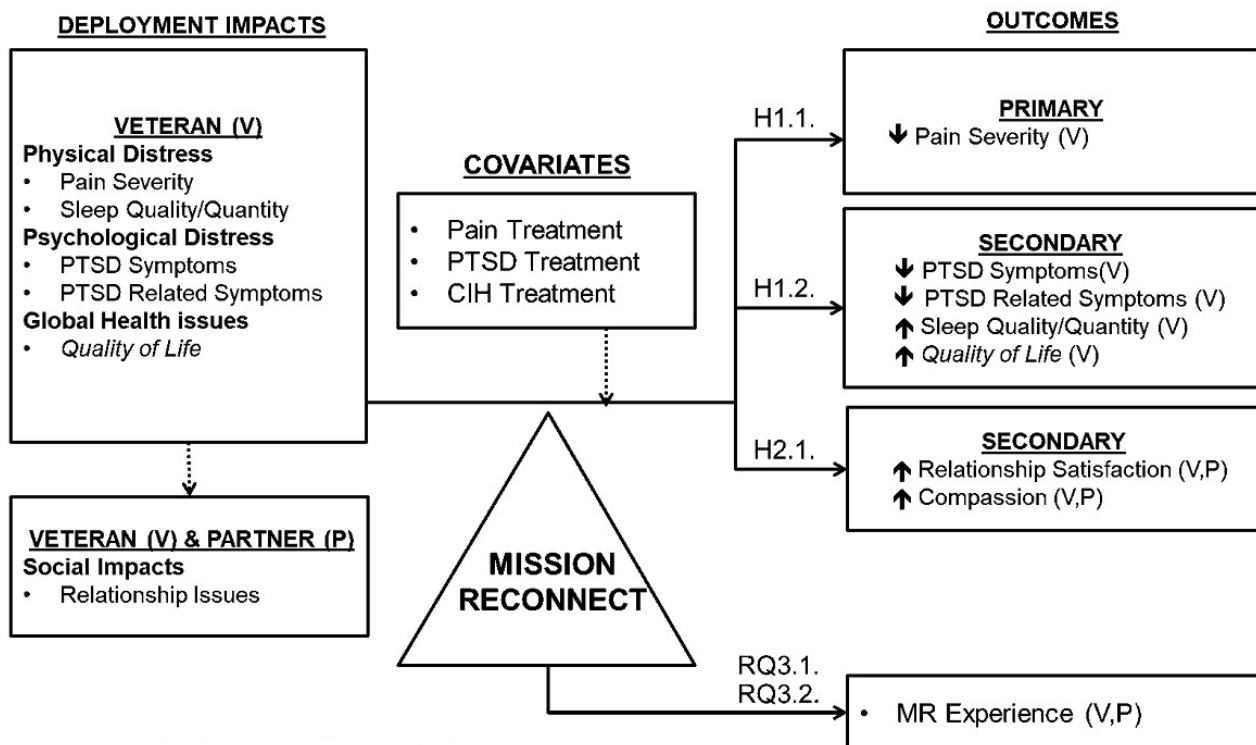
Hypothesized Mechanisms of Mission Reconnect

The biopsychosocial model was proposed to encourage exploration into categories of human distress that did not neatly fit the physiology-only, biomedical concept of disease and the processes by which disease developed [55]. The biopsychosocial model supports interdisciplinary programs that combine intervention components at the biological, psychological, and social levels [24,25]. MR leverages the mechanisms of action of multiple evidence-based CIH approaches, offering users multiple pathways to achieve clinical benefit. Users can benefit from the discrete therapies themselves but also the synergy of diverse therapies that mutually reinforce each other's effects. The primary modalities taught in MR—cognitive and behavioral approaches such as activation of mindfulness, gratitude and compassion, and massage—have very rich evidence bases for reducing pain and anxiety. Finally, the interpersonal support conferred by collaborative participation in MR makes possible the buffering effects of social support while reinforcing practice

activities [56]. Impacts of MR participation on pain have been demonstrated in both the Phase I and Phase II trials [57].

Research Design and Methods

This 4-year RCT with one intervention arm and one wait-list control arm will use mixed methods to evaluate the effectiveness and perceived value of the MR program in relation to physical and psychological symptoms, global health, and social outcomes. An underpinning of the MR program is that it is intended to be adjunctive (complementary) to conventional evidence-based modalities for treatment of pain and PTSD (ie, usual care). This is consistent with the previous MR trial conducted in a non-VA setting and different patient population whereby usual care served as the comparator group. Therefore, a waitlist control condition has been selected as the comparator for the proposed trial. This will allow assessment of the MR program above and beyond the effects of usual care being received. In addition, the relatively brief waitlist control period insures that all subjects will ultimately be offered the MR program. We believe this will be an incentive for both recruitment and retention of study subjects for this new recruitment setting. We will test MR using an RCT with concurrent mixed-methods data collection. We will recruit veteran and partner dyads. Study flow for each site is shown in Figure 2.

Aims 1 and 2 assessment data will be collected from 228 veteran/partner dyads (76 dyads per site, 38 dyads per arm) via a secure project website at baseline, 1, 2, and 4 months to assess primary and secondary outcomes. MR utilization and pain ratings will be assessed weekly for the first 8 weeks of the intervention for the treatment group. This data collection plan will allow assessment of acute changes, rate of change with repeated measures over 2 months and sustained changes.

In Aim 3, a subsample of 42 treatment group dyads (14 per site) will be purposively selected (high- vs low-volume MR use) to participate in telephone interviews to evaluate their experiences using the program. All consenting participants will be randomly assigned in a 1:1 ratio using a permuted variable block design to one of two arms (treatment vs control). Participants will receive study information, instructions on how to complete the online data collection, and contact information in case they experience an adverse event. Participants will provide preferred contact information for data collection reminders.

Figure 2. Study flow chart. and abbreviation CPRS: Computerized Patient Record System; MR: Mission Reconnect; VSSC: Veterans Health Administration Support Service Center Capital Assets.
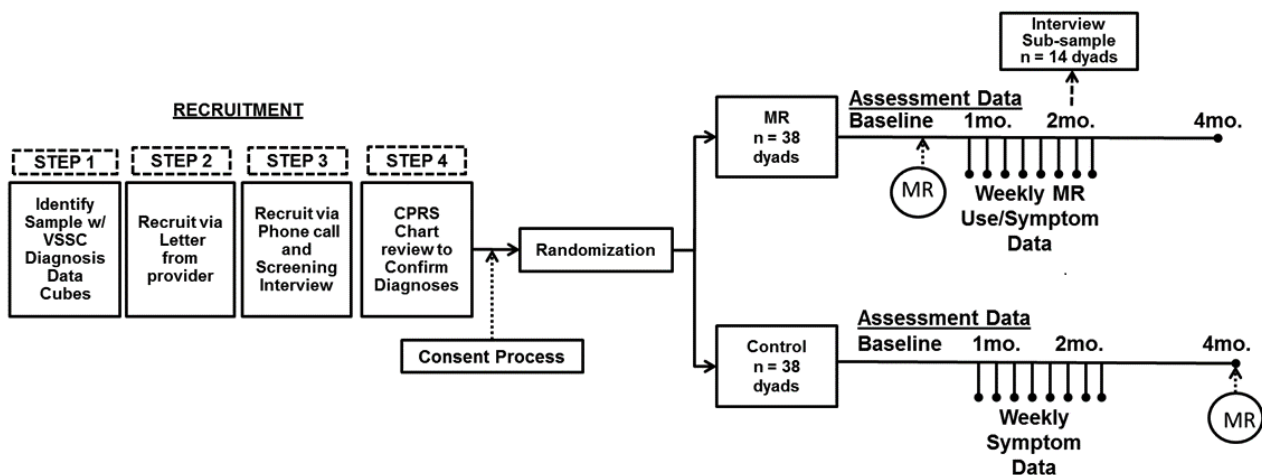
Treatment Arm

MR is a user-driven, dyadic, self-care management program developed with the National Institute of Mental Health funding (R43/44) for use by veterans and their selected partners, individually or together, to reduce pain and distress and support physical, mental, and relationship health. MR was designed for veterans who face obstacles accessing formal mental health services. It can also be used to complement formal services. MR is a patient-centered intervention, allowing users to determine the pace at which to proceed in each program component.

Mission Reconnect Content

The program provides video and audio instruction in a set of 11 evidence-based wellness activities in three thematic categories: Connecting with Yourself, Connecting with Quiet, and Connecting with Your Partner. All instruction is accessible via the Mission Reconnect website and mobile device apps.

Video content totals 91 minutes and was produced by filming 2 days of workshops to teach the practices to veteran/partner dyads. The Program Overview video (54 minutes) introduces the MR instructional sequences accompanied by commentary by workshop participants. Detailed massage instruction is presented in the separate Massage Instruction video (34 minutes) and Massage Video Supplement (3 minutes) addressing use with home furniture. Users are encouraged to give and receive at least one massage per week. Audio content totals 67 minutes and was recorded in studio, with nine instructional audios ranging from 1-22 minutes. Users are encouraged to listen, learn the practices, and then use each technique with or without the guided instruction as they wish.

A *Massage Instruction Booklet* and illustrated *Massage Reminder Handout* are downloadable from the website. A *What if?* feature enables users to access advice on how to apply program techniques in challenging situations such as experiencing problems with sleep, focus, and concentration.

Users can submit questions and suggestions for future content through the *What if?* interface. Optional Audios give users choice of audio gender voice. A Resources section provides links to hotlines, Vet centers, and VA facilities.

Users will be instructed to (1) try all the practices at least once during the first 2 weeks, and thereafter (2) do at least one exchange of massage per week with their partner, and (3) practice other methods of their choice at least 3-4 times per week. Dyads will be informed that greater use may result in greater benefit to reinforce practice.

Usual Care Waitlist Control Arm

For ethical reasons, this study will use a waitlist control arm to ultimately offer all participants exposure to the MR intervention. Partner dyads in the control arm will participate in all assessments like those in the treatment group; however, they will be asked to agree to not access the public website during their participation. Study team will maintain participant engagement with this program through a variety of methods, including email blasts, reminder prompts, and marketing blasts. Wait-list control participants will be instructed to seek advice about treatment from their providers. Other than this initial advice, there will be no attempt by study personnel to influence condition management unless an issue arises (eg, suicidal ideation). The control condition will account for potential temporal effects that occur from passage of time (brief) and expectation effects associated with anticipation of MR participation. While dyad randomization suggests primary physicians will have intervention and control patients in their practices, numerous effectiveness trials have shown there is little spillover of the intervention to usual care patients. The control group will receive access to MR after they complete data collection.

Sampling

We will use a four-step process to purposively recruit study participants (Figure 2). First, using an Institutional Review Board–approved procedure with a of waiver of consent process; a secondary administrative data query of the *International Classification of Diseases, Ninth and Tenth Revisions* (ICD 9-10) conditions for chronic pain and PTSD will identify a sample pool of veterans in the previous fiscal year. With our access levels and expertise, this process should take 2 weeks. Second, we will confirm approval to recruit from providers using signed letters from veterans' providers. Providers are being notified about study details and will provide recruitment letters to patients for them to follow up with a study team member about participating in the study.

Screening interviews will assess the occurrence of chronic pain (ie, pain for 6 months or more during the prior year) and symptoms of PTSD (ie, diagnosed, treated, or experienced symptoms of PTSD in the past 6 months); assess if they have visual, hearing, or other cognitive impairments; assess for previous diagnosis of moderate or severe traumatic brain injury (TBI); recent history of psychosis; and determine availability of a partner and the dyad's interest in participation. Additional details on pain treatment history, severity and treatments of PTSD symptoms, potential TBI exposure, and recent use of

complementary and alternative therapies for pain will be captured from the baseline data collection forms. In tandem, the fourth step (with waiver of consent) will allow chart review if needed for interested potential participants to (1) confirm comorbid conditions (ie, chronic pain and PTSD), (2) determine substance use disorder treatment status, and (3) determine TBI diagnosis and severity to exclude individuals with moderate or severe TBI.

Eligible veterans will be invited to participate after screening, be consented, and receive information to access the data collection site and determine their group assignment. Veterans and their partners will receive separate stipends via mail for their time.

The sample will consist of 152 participants (76 dyads) over approximately 24 months at each site. Women and minorities recruitment numbers will mirror site distribution based on sex and race, as this study is not powered to examine differences by race and sex. We anticipate considerable racial variability due to having three geographically diverse sites with markedly different census profiles. We anticipate a considerable representation of female participants with our invested partnership with a women's physician. This will make a considerable contribution to the knowledge of female veterans which is often lacking in VA studies. Recognizing the potential limitation of this strategy given the inclusion/exclusion criteria we, will also gain Institutional Review Board approval to collaborate with the local Rehabilitation Outcomes Research Section Veteran Engagement Council and co-investigators to recruit veteran participants using other recruitment means such as referrals and advertisements (eg, poster, brochures).

Inclusion criteria for participants will comprise the following: age 18 years or older, English-speaking veterans, and chronic musculoskeletal pain [58] as initially identified through secondary administrative data query of the ICD-9-10 conditions for chronic pain. In the initial query, musculoskeletal pain is present if the veteran meets either of two validated criteria: (1) ≥2 occurrences of any of targeted musculoskeletal *International Classification of Diseases and Related Health Problems, Ninth Revision, Clinical Modification* (ICD-9-CM) codes "likely to represent chronic pain" identified by Tian et al [59] recorded at visits separated by at least 30 days within past 6 months; or (2) high impact chronic pain equal to ≥2 occurrences of targeted musculoskeletal ICD-9-CM codes (adapted from Goulet et al [4]) separated by at least 30 days within the 6 months prior to study recruitment *and* ≥2 pain scores greater ≥4 separated by at least 30 days within past 6 months [59]. For pain scores, we will use the 0-10 numeric pain rating scale that is routinely collected at the VA. We will count two ICD-9-CM codes or pain scores recorded on the same day as one code/score.

Previous history of PTSD will be present if the veteran has a flag in their medical record indicating confirmed condition by the VA Compensation and Benefits program, has at least two outpatient visits in the year with the primary diagnosis being listed as PTSD (ICD-9-CM code 309.81), and/or had PTSD listed on the problem list. Confirmation of chronic pain for more than 6 months in the past year and diagnosis, treatment, or symptoms of PTSD in the past 6 months will be confirmed by

the telephone screening interview and use of the Computerized Patient Record System if needed for further confirmation. To participate in the study, the veteran must also have ability to access and use an electronic platform (eg, mobile device, Internet, DVD) for MR delivery, with a willing partner to participate in the study and MR program.

Exclusion criteria will include the following: moderate to severe TBI; diagnosis or documented treatment for psychosis in previous 6 months; currently in high-intensity substance use disorder treatment; non-English speaking; visual, hearing, cognitive impairment that prevents participation or ability to consent; self-report history of partner/family member physical abuse in the past year; and/or lack of Internet access. As stated above, potential participants who screen for aggression or violence will also be excluded from study.

Informed Consent

A waiver of informed consent process for recruitment purposes (medical record review) and a waiver of consent for a Verbal Consent Document for participant phone screening will occur. The study team will also be using a Written Consent Document or verbal consent via telephone.

Participants will complete the informed consent and Health Insurance Portability and Accountability Act (HIPAA) authorization for the study either in-person or over the phone. Research team members will review the consent and HIPAA content with participants to ensure review and comprehension. Since both partners will be full and equal participants in the study, both will be individually screened and consented by individual interview. We will employ three self-report items to address physical threat to and by partner, and fear reprisal. We will use a standard protocol advised and used in VA Family Services in its couples/family therapy: (1) initial brief – individual consent/self-report, (2) identify urgent need, (3) provide follow-up call and referral for community resources, and (4) exclude from study.

Participants will be provided an option to receive a printed copy of their informed consent and HIPAA authorization for review to ensure their understanding. Communication, such as informed consent explanations, will be supplemented using the phone to provide respondents opportunities for questions and clarification as needed. This is an effective means of communication for studies evaluating electronic health services such as MR. These remote methods of communication and consent have been effectively used in other studies conducted by the PI. These forms of communication (1) reduce participant burden, (2) conserve resources, and (3) leverage electronic communication devices, which support and promote virtual care.

Randomization

Participants (partner dyads) will be randomized to treatment (MR) or control using a per site variable block randomization method (blocks of 6 and 8) to facilitate a balance in treatment and control group sizes per site and over time. Variable blocks of 6 and 8 per site will be randomly used so that no one will know for sure the next random assignment. Dyad is the unit of randomization; this ensures that members of the same couple receive the same intervention. The randomization will be

stratified by whether or not the veteran is currently receiving, or has received in the past 2 weeks, a Level 1 evidence-based treatment for pain and/or PTSD, per current VA-DoD treatment guidelines.

This will facilitate subgroup analyses of the effect of the MR intervention in the presence versus absence of concomitant first-line evidence-based treatment for pain and PTSD. A computer-generated random allocation sequence will be generated by the study biostatistician, separately for each site/era of service combination to ensure balance across sites (ie, control for site effects by design). The sequence will be concealed to participants until immediately after their consent and baseline data collection, when group assignment information will be revealed on the last page of baseline survey. Participant dyads randomized to the treatment group will receive a link and sponsor code to access the MR site and will be instructed to create personal user accounts. Control group members will receive access to MR after completing data collection.

Sample Size Power Analysis

Estimates of statistical power are based on Aim 1 and initially the primary outcome of pain. Whereas dyad is the unit of randomization, we assess power based on separate analyses for veterans and partners (one approach used). There are four major intervals of assessment (baseline, 1, 2, and 4 months). For the 228 dyads, we conservatively assume up to 20% missing data over follow-up despite the previous MR trial that had less than 5% attrition. Assuming within-subject (outcome measurement) correlation of .5, the proposed sample size will provide 80% power (2-sided type I error rate of 0.05) to detect a "small-to-medium" effect size of 0.38 (Cohen d). For analyses stratified by site (in addition to evaluation by random effect assessment), the study will provide 80% power to detect a "medium-to-large" effect size of 0.66. By way of comparison, the pilot data presented above describe large effect sizes (d=0.81 for pain symptom reduction). While this effect size was based on a relatively small sample of comparable patients (N<15) and within-subject assessment (paired t test), it provides good assurance that the proposed sample size is sufficient to detect realistic, medium size effects should they exist with the MR program. For secondary outcomes, we will use a type I error rate of 0.01 for multiple comparisons. For the sample and our described assumptions, this will provide 80% power to detect a medium effect size of 0.45 [60].

Sample Size

Qualitative Aim 3 sample size relies on the quality and richness of information obtained [61]. Conceptual saturation is the goal of qualitative research and depends on data to support interpretations. Saturation has been noted to occur within the first 12 interviews [61,62]. Our team has extensive experience in evaluating electronic technologies and has found that recruiting high- and low-volume users provides the richest dataset. To ensure representative data, we will conduct telephone interviews with a purposive (high- vs low-volume users) subsample of up to 42 dyads to achieve representation at each site (14 dyads/site).

Data Collection Procedures

To test Aims 1 and 2, survey data will be collected using Qualtrics, a resource that has demonstrated capacity for remotely and securely collecting participant data in other studies conducted within the VA system. Qualtrics accounts are password-protected, and all data are replicated in real-time. The participants will be assigned a unique personal identification number (PIN) and will receive email messages with a link to prompt participant dyads to access the site and complete data collection. Upon access to Qualtrics, participants will be required to enter their PIN as the first entry into the system. Similarly, access to the online MR content will require entry of the same PIN. The PIN selected by each participant will be maintained in a cross-walk file with their randomly assigned study ID number. The Qualtrics measures will be compiled into a single survey format and collected at baseline, 1, 2, and 4 months. The rationale for this schedule is to formally assess initial, short-term, and sustained effects that may occur with the MR intervention. MR use and satisfaction, pain, tension, and stress items will be conducted weekly for the first 8 weeks of data collection. This more frequent schedule of data collection will permit short-term dose-response analyses (ie, dose of MR utilization) in relation to major symptoms of pain and PTSD. Measures and psychometric properties are illustrated in Table 1 [63-76].

To address Aim 3, telephone interviews (approximately 30 minutes) will be conducted with a subsample of 42 dyads, one month after the onset of the intervention (MR group). Only treatment group members will be recruited for this data collection. Interviews will be conducted with veterans and their partners separately. We will explain the interview purpose, ask permission to audiorecord, and use the interview script to ensure that all topics are covered. Interviewers will solicit respondents' attitudes, opinions, and reports about their preferences and the pros and cons of MR and participating in the practice groups, including their perceptions of usefulness. We will use standard communication techniques to stimulate discussion, with prompts (eg, "tell me more"), summarizing statements, and silence. We have used these methods in several previous studies to gather data effectively.

Table 1. Veteran and partner measures and psychometric properties.

Variable, construct	Measure	Description	Minutes	Time point
Veteran and partner independent variables				
Demographics	Participant survey	15 items assessing age, gender, ethnicity/race, education, income levels, religion, marital status, relationship length, military service/status/grade, deployments, computer/internet use	5	Baseline
Veteran independent variable: covariate				
Other treatments	Concurrent treatments	2 items asking veterans to report concurrent pain and PTSD[a] treatment(s) and CIH[b] modalities to account for dual intervention effects	1	Baseline, Months 1, 2, 4
Veterans and partner independent variable: covariate				
MR[c] program utilization (treatment and control group)	Utilization survey	11 items assessing frequency of use of the MR mind/body and massage practices	3	Weeks 1-8; Months 1, 2, 4
Veteran dependent variables				
Quality of life	Self-Assessment of Change [63,64]	16-item word-pairing scale assessing a variety of shifts in well-being across a broad range of therapeutic modalities and conditions	5	Months 1, 2, 4
Quality of life	Quality of Life Short Form-12 [65]	12 items assessing quality of life using physical status and mental health distress	3	Baseline, Months 1, 2, 4
Pain	Defense and Veterans Pain Rating Scale [66]	Pain Numeric Rating Scale [67], 11-point scale measuring "usual" pain intensity over last week and 4 pain functionality (past month) items	3	
	Pain Outcomes Questionnaire - Short Form VA[d] [68]	19 items assessing pain-related domains, including pain intensity, interference with activities and mobility, negative affect, vitality, pain-related fear; and improbable pain-related symptoms	5	
TBI[e] exposure	OHIO[f] TBI Exposure Screen [69]	8 items designed to elicit self- or proxy-reports of TBI occurring over a person's lifetime; can provide measures of extent of exposure to TBI including current injury	5	Baseline
Pain	Single-item scale	One item assessing pain using a 0-5–point Likert-type scale	1	Baseline, Weeks 1-8; Months 1, 2, 4
Stress and tension	Single-item scale	Two items assessing stress and tension using a 0-5–point Likert-type scale	1	
PTSD & related psychological symptoms	PTSD Checklist [70]	20-item measure of the *Diagnostic and Statistical Manual of Mental Disorder, Fifth Ed* (DSM-5) PTSD symptoms with scales related to stress, anxiety, & emotional numbing	4	Baseline, Months 1, 2, 4
	Depression: Beck Depression Inventory [71]	21 items, a widely used instrument for measuring depression; respondents asked to rate their symptoms and attitudes using a 4-point scale	5	
	Stress: Perceived Stress Scale [72]	10 Likert-scaled items, validated and widely used, to determine perceived stress levels	3	
	Sleep Quality: Pittsburg Sleep Quality Index [73]	19 self-rated questions from which 7 component scores are calculated and summed into a global score to assess sleep quality in the past month	4	
Veterans and partner dependent variables				
Relationship satisfaction	Revised Dyadic Adjustment Scale [74]-Adapted	14-item Likert-scaled instrument is reliable and valid and contains subscales for dyadic consensus, dyadic satisfaction, and dyadic cohesion.	3	Baseline, Months 1, 2, 4
Compassion	Compassion for self and others scales [75,76]	26-item Self-Compassion Scale and 21-item Compassion for Others measures, using 5-point Likert scale	5	
Program satisfaction (treatment only)	MR program satisfaction items	Eleven 10-point Likert-type items assessing satisfaction using MR components, whether they would recommend MR, and massage satisfaction	1	Weeks 1-8; Months 1, 2, 4

[a]PTSD: posttraumatic stress disorder.

[b]CIH: complementary and integrative health.

[c]MR: Mission Reconnect

[d]VA: Veterans Affairs

[e]TBI: traumatic brain injury.

[f]OHIO: Ohio State University TBI Identification Method.

Data Analysis

The intention-to-treat principle will be used for all analyses regardless of the extent of protocol compliance or dropout [77]. Whereas randomization is anticipated to balance both study arms on presenting demographic and clinical characteristics, including within sites and by concomitant receipt of first-line pain or PTSD treatment, continuous variables will be compared by student t tests or Wilcoxon tests (depending on distributional properties); categorical variables will be compared by chi-square analyses.

All outcome measures (primary and secondary) under Aim 1 will be collected at baseline, 1, 2, and 4 months and are continuous variables. Therefore, general linear mixed models will be used with main effects terms for GROUP (MR vs waitlist) and TIME (4 time points) and a GROUP x TIME interaction term (for rate of change). The models will also include site as a random effect. To assess whether the rate of change is curvilinear (ie, the rate of change differs between time points), a quadratic parameter will be tested. Different correlation structures and functional forms of the effects of the MR program will be assessed using the information criteria and a final parsimonious model will be determined for final statistical inference. Initial effects of the MR program (baseline to 1 month) will be evaluated by analysis of covariance. The primary outcome measure for pain will be the total score on the 19-item Pain Outcomes Questionnaire. The primary outcome measure for PTSD will be the total score on the 20-item PTSD Checklist for DSM-5. Subgroup analyses will be explored using similar methods and examination of severity of baseline pain and PTSD scores. These analyses will provide insight into subgroups for whom the MR program may be particularly beneficial. In a similar realm, analyses will be stratified by baseline median level of relationship satisfaction, as derived from total score on the 14-item Revised Dyadic Adjustment Scale. This will permit assessment of the effect of the MR intervention on potential improvement in relationship satisfaction among a cohort of dyads with presumably troubled relationships at entry. Moreover, using the weekly reports (8 weeks) of MR use and satisfaction with the MR program, the MR sample will be split above versus below the median for these two measures and compared separately against the waitlist control condition at 2 months. This approach approximates a "per protocol" analysis in terms of recommended use of the MR program. In addition, especially for Aim 2 outcomes, we seek to examine whether dyads appear to show mutual benefits from the MR program. Therefore, multilevel models will be fit using an over-time dyadic model [78] in which individuals are nested within dyads and time is crossed with dyads (ie, both veteran and partner are assessed at the 4 time points). This analysis accounts for the non-independence due to the correlation between dyad members' general levels on outcomes averaged over time, as well as the time-specific correlation between their outcomes (eg, similarity caused by time-specific events).

Interview transcript data will be managed using descriptive content analytical methods to identify domains and taxonomies related to participants' experiences with MR use [61]. Categories will be compared and contrasted, and relationships among them will be identified. As coding schemas are developed to create domains and taxonomies, data samples will be extracted and coded by research team members and evaluated for interrater reliability and credibility. Descriptive and comparative matrices, which identify the patterns of regularities (shared) and inconsistencies (unique or varied) will then be constructed for the veteran and partner participants. Comparative matrices enable identification of the most relevant, shared, and perhaps representative components, thereby enhancing the potential representation of the findings. Finally, a complex cross-case data display or matrix will be developed to summarize the significant taxonomic outcome structures identified within and between veterans and their partners. This process of descriptive and comparative matrix analysis will allow discernment of the most salient and representative components identified by veterans and their partners.

Missing Data

Missing data will be tabulated by treatment arms and by assessment waves; comparison tests between arms will be conducted to assess potential attrition bias and to examine the missing data mechanism (eg, missing at random). If the missing rate is less than 10%, analyses with list wise deletion (ie, missing values dropped from the analysis) will be performed due to minimal concern over bias. Participants who are lost to follow-up and missing on postrandomization outcome assessments will be included in additional comparison analyses that use multiple imputations of missing data to minimize bias due to differences between those with complete and incomplete data.

Results

Funding for the study began in November 2018, and we are currently in the process of recruitment screening and data randomization for the study. Primary data collection will begin on May 2019 and continue through May 2021. Projected participants per site will be 76 partner dyads, for a total of 456 study participants. Anticipated study results will be published in November 2022. Table 2 shows the projected study schedule.

Table 2. Gantt chart of study benchmarks.

Project activity	Year 1 (2019-2020)				Year 2 (2020-2021)				Year 3 (2021-2022)				Year 4 (2022-2023)			
	Q1 Nov '18	Q2 Feb '19	Q3 May	Q4 Aug	Q1 Nov	Q2 Feb '20	Q3 May	Q4 Aug	Q1 Nov	Q2 Feb '21	Q3 May	Q4 Aug	Q1 Nov	Q2 Feb '22	Q3 May	Q4 Aug
Start-up	X[a]															
Recruitment screening and randomization, Qualtrics development and validation		X														
Recruitment: 3-4 dyads/month per site			X	X	X	X	X	X	X	X						
Primary data collection			X	X	X	X	X	X	X	X	X					
Conduct interviews				X	X	X	X									
Interview transcription				X	X	X	X	X								
Interview data analysis					X	X	X	X	X							
Prepare and stage primary data										X	X	X				
Primary data analysis													X	X		
Data interpretation and triangulation													X	X	X	
Finalize data reports/manuscripts															X	X
Develop materials for dissemination															X	X
Prepare/submit subsequent proposal															X	X
Disseminate materials to audiences															X	X

[a]X: denotes the activity occurred during this time frame.

Discussion

Principal Considerations

The goal of this study protocol is to use evidence-based CIH methods to decrease chronic pain and PTSD symptoms in the veteran-partner dyad. The research goal is to evaluate MR as an approach to manage chronic pain and PTSD symptoms for potential subsequent implementation. This study will possibly provide a model for establishing remote access and sustainable implementation of CIH within VA. If shown to be effective, there is reason to believe it is scalable, feasible, and sustainable and can ramp up nationally. This will also relieve a bottleneck of services by moving from a provider-based delivery to partner delivery.

Strengths and Limitations

This protocol contributes to the science in three distinct ways: (1) a conjoint approach to supporting the veteran through the use of partnered CIH, (2) use of a self-directed evidence-based CIH mobile app for a clinically defined population of veterans and partner, and (3) MR's distinctive use of nonpharmacological interventions to manage chronic pain.

This protocol is an RCT approach powered for generalizability. Mixed methods will illuminate the how and why of the veterans'

and their partners' experience of mobile apps for CIH. The strength of the intervention is that it is remotely delivered thus overcoming geographical restrictions. If shown to be effective, it can be scalable and will relieve potential service bottlenecks by establishing partner-based delivery rather than provider-based delivery.

A limitation in the analytic methods are threats to protocol by using remote access in an electronic model, which may have an impact on recruitment, attrition, and potential issues in the delivery process. However, safeguards are in place, which include staff to help troubleshoot. Using remote data collection can be challenging. By taking this protocol away from paper-based responses and into a model that leverages mobile and remote data collection, lessons may be learned on how best to exercise remote and electronic data collection methods.

The limitations in waitlist controls have been criticized due to ethical concerns. A waitlist control is not ethical when there are other treatment and interventions available. The waitlist can speed recruitment and parcel out expectations, and people may get better anyway. It is a control condition which is better than no control, and it is appropriate at an early stage of intervention development. This protocol allows waitlist controls to gain exposure and provide the study with a comparator group of the

proposed trial. The attractive part of this waitlist is knowing they will get the intervention.

The study scope is focused on providing a remote intervention with veterans who have PTSD and pain symptoms. Though veterans who are symptomatic and not engaged in the health care system may benefit from this intervention, they are not the target of this protocol. It is also possible veterans who are symptomatic but not documented in the VA system may also be missed due to using secondary data source to identify the sample pool. The method of recruitment using secondary sources was based on optimal feasibility for large-scale RCT recruitment. Future research should focus on use of mobile app technology to improve access to CIH in a variety of populations. Additionally, future research should look to the most appropriate electronically captured patient-reported outcome tools. This research did not take the opportunity to measure the partners'

outcomes due to the funding source mechanism. Therefore, future research should look at partner experience. If highly variable responses/outcomes with MR are found, then data that include participant characteristics will guide intervention modifications for this patient group.

Conclusions

This intervention is extremely important and innovative. Large government organizations currently have limited capacity for electronic patient-reported outcomes. Patient-reported outcomes and MR can test different measures electronically by optimizing use of technology and remote delivery of intervention and data collection methods. From the user perspective, MR allows the user to leverage a "when they want, where they want" approach. Optimizing use of technology and remote delivery of intervention and data collection methods will contribute to the field of science.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Funding reviewer comments.

[PDF File (Adobe PDF File), 174KB - resprot_v8i5e13666_app1.pdf]

References

1. Lew HL, Otis JD, Tun C, Kerns RD, Clark ME, Cifu DX. Prevalence of chronic pain, posttraumatic stress disorder, and persistent postconcussive symptoms in OIF/OEF veterans: polytrauma clinical triad. J Rehabil Res Dev 2009;46(6):697-702 [FREE Full text] [Medline: 20104399]
2. Kip KE, Rosenzweig L, Hernandez DF, Shuman A, Diamond DM, Girling SA, et al. Accelerated Resolution Therapy for treatment of pain secondary to symptoms of combat-related posttraumatic stress disorder. Eur J Psychotraumatol 2014;5 [FREE Full text] [doi: 10.3402/ejpt.v5.24066] [Medline: 24959325]
3. Goulet JL, Erdos J, Kancir S, Levin FL, Wright SM, Daniels SM, et al. Measuring performance directly using the veterans health administration electronic medical record: a comparison with external peer review. Med Care 2007 Jan;45(1):73-79 [FREE Full text] [doi: 10.1097/01.mlr.0000244510.09001.e5] [Medline: 17279023]
4. Goulet JL, Kerns RD, Bair M, Becker WC, Brennan P, Burgess DJ, et al. The musculoskeletal diagnosis cohort: examining pain and pain care among veterans. Pain 2016 Dec;157(8):1696-1703 [FREE Full text] [doi: 10.1097/j.pain.0000000000000567] [Medline: 27023420]
5. Seal KH, Cohen G, Waldrop A, Cohen BE, Maguen S, Ren L. Substance use disorders in Iraq and Afghanistan veterans in VA healthcare, 2001-2010: Implications for screening, diagnosis and treatment. Drug Alcohol Depend 2011 Jul 01;116(1-3):93-101. [doi: 10.1016/j.drugalcdep.2010.11.027] [Medline: 21277712]

XSL•FO

RenderX

6.   Seal KH, Shi Y, Cohen G, Cohen BE, Maguen S, Krebs EE, et al. Association of mental health disorders with prescription opioids and high-risk opioid use in US veterans of Iraq and Afghanistan. JAMA 2012 Mar 07;307(9):940-947. [doi: 10.1001/jama.2012.234] [Medline: 22396516]

7.   Ennis N, Sijercic I, Monson CM. Internet-Delivered Early Interventions for Individuals Exposed to Traumatic Events: Systematic Review. J Med Internet Res 2018 Nov 14;20(11):e280 [FREE Full text] [doi: 10.2196/jmir.9795] [Medline: 30429113]

8.   Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]

9.   Hickling EJ, Blanchard EB. Post-traumatic stress disorder and motor vehicle accidents. Journal of Anxiety Disorders 1992 Jul;6(3):285-291. [doi: 10.1016/0887-6185(92)90040-E]

10.   Sharp TJ, Harvey AG. Chronic pain and posttraumatic stress disorder: mutual maintenance? Clin Psychol Rev 2001 Aug;21(6):857-877. [Medline: 11497210]

11.   Department of Veterans Affairs. PTSD: National Center for PTSD.: Veterans Affairs; 2007. Chronic Pain and PTSD: A Guide for Patients URL: https://www.ptsd.va.gov/understand/related/chronic_pain.asp [accessed 2019-04-10] [WebCite Cache ID 77XlNo8fB]

12.   Tsang A, Von Korff M, Lee S, Alonso J, Karam E, Angermeyer MC, et al. Common chronic pain conditions in developed and developing countries: gender and age differences and comorbidity with depression-anxiety disorders. J Pain 2008 Oct;9(10):883-891. [doi: 10.1016/j.jpain.2008.05.005] [Medline: 18602869]

13.   Clark ME, Walker RL, Gironda RJ, Scholten JD. Comparison of pain and emotional symptoms in soldiers with polytrauma: unique aspects of blast exposure. Pain Med 2009 Apr;10(3):447-455. [doi: 10.1111/j.1526-4637.2009.00590.x] [Medline: 19416436]

14.   Smeeding SJW, Bradshaw DH, Kumpfer KL, Trevithick S, Stoddard GJ. Outcome evaluation of the Veterans Affairs Salt Lake City Integrative Health Clinic for Chronic Nonmalignant Pain. Clin J Pain 2011 Feb;27(2):146-155. [doi: 10.1097/AJP.0b013e3181f158e8] [Medline: 20842019]

15.   Shi L, Liu J, Zhao Y. Comparative effectiveness in pain-related outcomes and health care utilizations between veterans with major depressive disorder treated with duloxetine and other antidepressants: a retrospective propensity score-matched comparison. Pain Pract 2012 Jun;12(5):374-381. [doi: 10.1111/j.1533-2500.2011.00495.x] [Medline: 21951787]

16.   Lang KP, Veazey-Morris K, Andrasik F. Exploring the role of insomnia in the relation between PTSD and pain in veterans with polytrauma injuries. J Head Trauma Rehabil 2014;29(1):44-53. [doi: 10.1097/HTR.0b013e31829c85d0] [Medline: 23835878]

17.   Kupersmith J, Lew HL, Ommaya AK, Jaffee M, Koroshetz WJ. Traumatic brain injury research opportunities: results of Department of Veterans Affairs Consensus Conference. J Rehabil Res Dev 2009;46(6):vii-xvi [FREE Full text] [Medline: 20104393]

18.   Bosco MA, Murphy JL, Clark ME. Chronic pain and traumatic brain injury in OEF/OIF service members and Veterans. Headache 2013 Oct;53(9):1518-1522. [doi: 10.1111/head.12172] [Medline: 23848062]

19.   Schnurr PP, Lunney CA. Work-related outcomes among female veterans and service members after treatment of posttraumatic stress disorder. Psychiatr Serv 2012 Nov;63(11):1072-1079. [doi: 10.1176/appi.ps.201100415] [Medline: 22983600]

20.   Westanmo A, Marshall P, Jones E, Burns K, Krebs EE. Opioid Dose Reduction in a VA Health Care System--Implementation of a Primary Care Population-Level Initiative. Pain Med 2015 May;16(5):1019-1026. [doi: 10.1111/pme.12699] [Medline: 25645538]

21.   Walker RL, Clark ME, Sanders SH. The "Postdeployment multi-symptom disorder": An emerging syndrome in need of a new treatment paradigm. Psychological Services 2010;7(3):136-147. [doi: 10.1037/a0019684]

22.   Ohye BY, Brendel RW, Fredman SJ, Bui E, Rauch PK, Allard MD, et al. Three-generation model: A family systems framework for the assessment and treatment of veterans with posttraumatic stress disorder and related conditions. Professional Psychology: Research and Practice 2015 Apr;46(2):97-106. [doi: 10.1037/a0037735]

23.   Capehart B, Bass D. Review: managing posttraumatic stress disorder in combat veterans with comorbid traumatic brain injury. J Rehabil Res Dev 2012;49(5):789-812 [FREE Full text] [Medline: 23015586]

24.   Engel GL. The clinical application of the biopsychosocial model. Am J Psychiatry 1980 May;137(5):535-544. [doi: 10.1176/ajp.137.5.535] [Medline: 7369396]

25.   Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]

26.   Geisser ME, Roth RS, Bachman JE, Eckert TA. The relationship between symptoms of post-traumatic stress disorder and pain, affective disturbance and disability among patients with accident and non-accident related pain. Pain 1996 Aug;66(2-3):207-214. [Medline: 8880842]

27.   Raichle KA, Romano JM, Jensen MP. Partner responses to patient pain and well behaviors and their relationship to patient pain behavior, functioning, and depression. Pain 2011 Jan;152(1):82-88. [doi: 10.1016/j.pain.2010.09.015] [Medline: 20947249]

28. Nelson BS, Wright DW. Understanding and treating post‐traumatic stress disorder symptoms in female partners of veterans with PTSD. Journal of Marital and Family Therapy 1996;22(4):455-467. [doi: 10.1111/j.1752-0606.1996.tb00220.x]

29. Verbosky SJ, Ryan DA. Female partners of Vietnam veterans: stress by proximity. Issues Ment Health Nurs 1988;9(1):95-104. [Medline: 3356550]

30. Keefe FJ, Caldwell DS, Baucom D, Salley A, Robinson E, Timmons K, et al. Spouse-assisted coping skills training in the management of osteoarthritic knee pain. Arthritis Care Res 1996 Aug;9(4):279-291. [Medline: 8997917]

31. Abbasi M, Dehghani M, Keefe FJ, Jafari H, Behtash H, Shams J. Spouse-assisted training in pain coping skills and the outcome of multidisciplinary pain management for chronic low back pain treatment: a 1-year randomized controlled trial. Eur J Pain 2012 Aug;16(7):1033-1043. [doi: 10.1002/j.1532-2149.2011.00097.x] [Medline: 22337646]

32. Bolton RE, Fix GM, VanDeusen Lukas C, Elwy AR, Bokhour BG. Biopsychosocial benefits of movement-based complementary and integrative health therapies for patients with chronic conditions. Chronic Illn 2018 Jan 01. [doi: 10.1177/1742395318782377] [Medline: 29914264]

33. Baucom DH, Shoham V, Mueser KT, Daiuto AD, Stickle TR. Empirically supported couple and family interventions for marital distress and adult mental health problems. J Consult Clin Psychol 1998 Feb;66(1):53-88. [Medline: 9489262]

34. Monson CM, Macdonald A, Brown-Bowers A. Couple/family therapy for posttraumatic stress disorder: review to facilitate interpretation of VA/DOD Clinical Practice Guideline. J Rehabil Res Dev 2012;49(5):717-728 [FREE Full text] [Medline: 23015582]

35. Church D, Brooks AJ. CAM and energy psychology techniques remediate PTSD symptoms in veterans and spouses. Explore (NY) 2014;10(1):24-33. [doi: 10.1016/j.explore.2013.10.006] [Medline: 24439093]

36. Kahn JR, Collinge W, Soltysik R. Post-9/11 Veterans and Their Partners Improve Mental Health Outcomes with a Self-directed Mobile and Web-based Wellness Training Program: A Randomized Controlled Trial. J Med Internet Res 2016 Sep 27;18(9):e255 [FREE Full text] [doi: 10.2196/jmir.5800] [Medline: 27678169]

37. Denneson LM, Corson K, Dobscha SK. Complementary and alternative medicine use among veterans with chronic noncancer pain. J Rehabil Res Dev 2011;48(9):1119-1128 [FREE Full text] [Medline: 22234716]

38. Herman PM, Sorbero ME, Sims-Columbia AC. Complementary and Alternative Medicine Services in the Military Health System. J Altern Complement Med 2017 Nov;23(11):837-843. [doi: 10.1089/acm.2017.0236] [Medline: 29039681]

39. Peterson K, Anderson J, Ferguson L, Mackey K. Evidence Brief: The Comparative Effectiveness of Selected Complementary and Integrative Health (CIH) Interventions for Preventing or Reducing Opioid Use in Adults with Chronic Neck, Low Back, and Large Joint Pain. URL: https://www.hsrd.research.va.gov/publications/esp/cihopioid.cfm [accessed 2019-04-10] [WebCite Cache ID 77X5EPdhT]

40. Haun JN, Graham-Pole J, Shortley B. Children with cancer and blood diseases experience positive physical and psychological effects from massage therapy. Int J Ther Massage Bodywork 2009 Jun 29;2(2):7-14 [FREE Full text] [Medline: 21589728]

41. Cherkin DC, Eisenberg D, Sherman KJ, Barlow W, Kaptchuk TJ, Street J, et al. Randomized trial comparing traditional Chinese medical acupuncture, therapeutic massage, and self-care education for chronic low back pain. Arch Intern Med 2001 Apr 23;161(8):1081-1088. [Medline: 11322842]

42. Moyer CA, Rounds J, Hannum JW. A meta-analysis of massage therapy research. Psychol Bull 2004 Jan;130(1):3-18. [doi: 10.1037/0033-2909.130.1.3] [Medline: 14717648]

43. Tsao JCI. Effectiveness of massage therapy for chronic, non-malignant pain: a review. Evid Based Complement Alternat Med 2007 Jun;4(2):165-179 [FREE Full text] [doi: 10.1093/ecam/nel109] [Medline: 17549233]

44. Evidence-based Synthesis Program (ESP) Center. Evidence Synthesis Program. 2016. Massage for Pain: An Evidence Map URL: https://www.hsrd.research.va.gov/publications/esp/massage-REPORT.pdf [accessed 2019-04-10] [WebCite Cache ID 77XomrbPy]

45. Boyd C, Crawford C, Paat CF, Price A, Xenakis L, Zhang W, Evidence for Massage Therapy (EMT) Working Group. The Impact of Massage Therapy on Function in Pain Populations-A Systematic Review and Meta-Analysis of Randomized Controlled Trials: Part II, Cancer Pain Populations. Pain Med 2016 Dec;17(8):1553-1568 [FREE Full text] [doi: 10.1093/pm/pnw100] [Medline: 27165967]

46. Sharpe PA, Williams HG, Granner ML, Hussey JR. A randomised study of the effects of massage therapy compared to guided relaxation on well-being and stress perception among older adults. Complement Ther Med 2007 Sep;15(3):157-163. [doi: 10.1016/j.ctim.2007.01.004] [Medline: 17709060]

47. Moraska A, Pollini RA, Boulanger K, Brooks MZ, Teitlebaum L. Physiological adjustments to stress measures following massage therapy: a review of the literature. Evid Based Complement Alternat Med 2010 Dec;7(4):409-418 [FREE Full text] [doi: 10.1093/ecam/nen029] [Medline: 18955340]

48. Haun J, Patel N, Schwartz G, Ritenbaugh C. Evaluating the use of gas discharge visualization to measure massage therapy outcomes. J Complement Integr Med 2015 Sep;12(3):231-239 [FREE Full text] [doi: 10.1515/jcim-2014-0014] [Medline: 26087069]

49. Stephenson KR, Simpson TL, Martinez ME, Kearney DJ. Changes in Mindfulness and Posttraumatic Stress Disorder Symptoms Among Veterans Enrolled in Mindfulness-Based Stress Reduction. J Clin Psychol 2017 Dec;73(3):201-217. [doi: 10.1002/jclp.22323] [Medline: 27152480]

50. Garland EL, Manusov EG, Froeliger B, Kelly A, Williams JM, Howard MO. Mindfulness-oriented recovery enhancement for chronic pain and prescription opioid misuse: results from an early-stage randomized controlled trial. J Consult Clin Psychol 2014 Jun;82(3):448-459 [FREE Full text] [doi: 10.1037/a0035798] [Medline: 24491075]

51. Serpa JG, Taylor SL, Tillisch K. Mindfulness-based stress reduction (MBSR) reduces anxiety, depression, and suicidal ideation in veterans. Med Care 2014 Dec;52(12 Suppl 5):S19-S24. [doi: 10.1097/MLR.0000000000000202] [Medline: 25397818]

52. Department of Veterans Afairs; Department of Defense. VA/DoD Clinical Practice Guidelines. 2017. VA/DoD Clinical Practice Guideline for the Management of Posttraumatic Stress Disorder and Acute Stress Disorder URL: https://www.healthquality.va.gov/guidelines/MH/ptsd/VADoDPTSDCPGFinal082917.pdf [accessed 2019-04-10] [WebCite Cache ID 77XpeLT2R]

53. Elwy AR, Johnston JM, Bormann JE, Hull A, Taylor SL. A systematic scoping review of complementary and alternative medicine mind and body practices to improve the health of veterans and military personnel. Med Care 2014 Dec;52(12 Suppl 5):S70-S82. [doi: 10.1097/MLR.0000000000000228] [Medline: 25397827]

54. Kligler B. Research Advisory Committee on Gulf War Veterans' Illnesses. 2017. Update on Complementary and Integrative Health Strategy URL: https://www.va.gov/RAC-GWVI/meetings/apr2017/Kligler.pdf [accessed 2019-04-10] [WebCite Cache ID 77Xq7XBgf]

55. Engel GL. The need for a new medical model: a challenge for biomedicine. Psychodyn Psychiatry 2012 Sep;40(3):377-396. [doi: 10.1521/pdps.2012.40.3.377] [Medline: 23002701]

56. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]

57. Collinge W, Kahn J, Soltysik R. Promoting reintegration of National Guard veterans and their partners using a self-directed program of integrative therapies: a pilot study. Mil Med 2012 Dec;177(12):1477-1485 [FREE Full text] [Medline: 23397692]

58. Gureje O, Von Korff M, Simon GE, Gater R. Persistent pain and well-being: a World Health Organization Study in Primary Care. JAMA 1998 Jul 08;280(2):147-151. [Medline: 9669787]

59. Tian TY, Zlateva I, Anderson DR. Using electronic health records data to identify patients with chronic pain in a primary care setting. J Am Med Inform Assoc 2013 Dec;20(e2):e275-e280 [FREE Full text] [doi: 10.1136/amiajnl-2013-001856] [Medline: 23904323]

60. Cohen J. Statistical Power Analysis For The Behavioral Sciences (2nd Edition). New York, NY: Routledge; 1988.

61. Strauss A, Corbin J. Basics Of Qualitative Research: Second Edition: Techniques And Procedures For Developing Grounded Theory. Thousand Oaks, CA: Sage Publications, Inc; 1998.

62. Sandelowski M. Sample size in qualitative research. Res Nurs Health 1995 Apr;18(2):179-183. [Medline: 7899572]

63. Ritenbaugh C, Nichter M, Nichter MA, Kelly KL, Sims CM, Bell IR, et al. Developing a patient-centered outcome measure for complementary and alternative medicine therapies I: defining content and format. BMC Complement Altern Med 2011 Dec 29;11:135 [FREE Full text] [doi: 10.1186/1472-6882-11-135] [Medline: 22206345]

64. Thompson JJ, Kelly KL, Ritenbaugh C, Hopkins AL, Sims CM, Coons SJ. Developing a patient-centered outcome measure for complementary and alternative medicine therapies II: refining content validity through cognitive interviews. BMC Complement Altern Med 2011 Dec 29;11:136 [FREE Full text] [doi: 10.1186/1472-6882-11-136] [Medline: 22206409]

65. Busija L, Pausenberger E, Haines TP, Haymes S, Buchbinder R, Osborne RH. Adult measures of general health and health-related quality of life: Medical Outcomes Study Short Form 36-Item (SF-36) and Short Form 12-Item (SF-12) Health Surveys, Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), Medical Outcomes Study Short Form 6D (SF-6D), Health Utilities Index Mark 3 (HUI3), Quality of Well-Being Scale (QWB), and Assessment of Quality of Life (AQoL). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11:S383-S412 [FREE Full text] [doi: 10.1002/acr.20541] [Medline: 22588759]

66. Polomano RC, Galloway KT, Kent ML, Brandon-Edwards H, Kwon KN, Morales C, et al. Psychometric Testing of the Defense and Veterans Pain Rating Scale (DVPRS): A New Pain Scale for Military Population. Pain Med 2016 Dec;17(8):1505-1519. [doi: 10.1093/pm/pnw105] [Medline: 27272528]

67. Turk DC, Melzack R. Self-report scales and procedures for assessing pain in adults. In: Handbook Of Pain Assessment, Third Edition. New York, NY: The Guilford Press; 1992.

68. Clark ME, Gironda RJ, Young RW. Development and validation of the Pain Outcomes Questionnaire-VA. J Rehabil Res Dev 2003;40(5):381-395 [FREE Full text] [Medline: 15080223]

69. Corrigan JD, Bogner J. Initial reliability and validity of the Ohio State University TBI Identification Method. J Head Trauma Rehabil 2007;22(6):318-329. [doi: 10.1097/01.HTR.0000300227.67748.77] [Medline: 18025964]

70. Bovin MJ, Marx BP, Weathers FW, Gallagher MW, Rodriguez P, Schnurr PP, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. Psychol Assess 2016 Dec;28(11):1379-1391. [doi: 10.1037/pas0000254] [Medline: 26653052]

71. Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. J Pers Assess 1996 Dec;67(3):588-597. [doi: 10.1207/s15327752jpa6703_13] [Medline: 8991972]

72. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]

73. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [Medline: 2748771]

74. Busby DM, Christensen C, Crane DR, Larson JH. Revision of the Dyadic Adjustment Scale for Use with Distressed Non distressed Couples: Construct Hierarchy Multidimensional Scales. J Marital Fam Ther 1995;21(3):289-309. [doi: 10.1111/j.1752-0606.1995.tb00163.x]

75. Neff KD. The Development and Validation of a Scale to Measure Self-Compassion. Self and Identity 2003 Jul;2(3):223-250. [doi: 10.1080/15298860309027]

76. Sprecher S, Fehr B. Compassionate love for close others and humanity. Journal of Social and Personal Relationships 2016 Jun 30;22(5):629-651. [doi: 10.1177/0265407505056439]

77. Ten Have TR, Normand SLT, Marcus SM, Brown CH, Lavori P, Duan N. Intent-to-Treat vs. Non-Intent-to-Treat Analyses under Treatment Non-Adherence in Mental Health Randomized Trials. Psychiatr Ann 2008 Dec;38(12):772-783 [FREE Full text] [doi: 10.3928/00485713-20081201-10] [Medline: 20717484]

78. Kenny DA, Kashy DA, Cook WL. Dyadic Data Analysis (methodology In The Social Sciences). New York, NY: The Guilford Press; 2019.

Abbreviations

CIH: complementary and integrative health
DOD: Department of Defense
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ICD-9-10: International Classification of Diseases and Related Health Problems, Ninth and Tenth Revisions
ICD-9-CM: International Classification of Diseases and Related Health Problems, Ninth Revision, Clinical Modification
MR: Mission Reconnect
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
VA: Veterans Affairs
VHA: Veterans Health Administration

Protocol

Improving Treatment of Elderly Patients by Interprofessional Education in a Quality Network of Geriatric Medicine: Protocol for Evaluating an Educational Initiative

Daisy Huenefeld[1], MD, MBA; Sibyll Rodde[2], PhD, MSc; Gertrud Bureick[3], PhD, MA; Barbara Elkeles[4], Prof Dr, MD; Joachim Hasebrook[5], Prof Dr, MSc, BSc

[1]Board of the Foundation, St Franziskus Foundation, Muenster, Germany

[2]zeb.healthcare, zeb.rolfes.schierenbeck.associates, Muenster, Germany

[3]BIGi Project, St Franziskus Foundation, Muenster, Germany

[4]Klinik Maria Frieden, Telgte, Germany

[5]zeb.business school, Steinbeis University Berlin, Muenster, Germany

Corresponding Author:
Joachim Hasebrook, Prof Dr, MSc, BSc
zeb.business school
Steinbeis University Berlin
Hammer Str 165
Muenster, 48153
Germany
Phone: 49 15152647546
Email: jhasebrook@zeb-bs.de

Abstract

Background: All statistics on the development of demand for care for multimorbid elderly patients highlight the acute pressure to act to adequately respond to the expected increase in geriatric patient population in the next 15 years. Against this background, great importance must be attached to the improvement of cross-occupational group and cross-sector treatment of these patients. In addition, many professionals in the health care sector often have little knowledge about the special treatment and care needs of the elderly.

Objective: The Quality Network of Geriatric Medicine in north-west Germany is the body responsible for the project; with its member organizations, it provides care for over 400,000 inpatients and is thus one of the largest associations for geriatrics in Germany. The Quality Network conducts binding evaluated qualification measures for staff involved in the treatment and care of multimorbid elderly patients. The training offers are especially intended for staff who have not yet been trained in working with elderly patients. This approach is intended to improve the expertise of various occupational groups on different hierarchy levels, to include patients and their family members in the evaluation process, and to initiate changes within the organizations.

Methods: Various instruments are used in the evaluation of qualification measures: besides written surveys and questionnaires, structured work groups (consensus groups) and interviews are conducted. The evaluation starts before the qualification measures to determine the starting point and then continues during the measure and after its completion. This allows major findings to be integrated directly into the ongoing qualification program. At least 100 trainings on geriatric topics, 80 consensus groups, and 120 patients (and family members) are going to be included in the study.

Results: The evaluation of the educational initiative is funded by the State of Northrhine-Westfalia (Germany; LZG TG 71 001 / 2015 and LZG TG 71 002 / 2015). The results of the study will be published after review and approval by the state authorities – presumably by the end of 2019. The before and after comparison of the treatment-related outcomes at the beginning and near the completion of the educational initiative gives insights into how transfer-oriented education can improve the treatment of elderly patients across sector lines for inpatients as well as outpatients. The evaluation of the implementation of educational content in day-to-day work and occupational groups is to facilitate recommendations about economically sensible use of educational resources and about further adjustments to the training content.

XSL•FO
RenderX

Conclusions: The evaluation develops the foundation for targeted and needs-oriented qualification measures as well as transfer in cross-sector, multiprofessional networks. Instruments and results will be published and provided to other health care networks and institutions. The Quality Network will implement the results of the evaluation process in its member institutions.

International Registered Report Identifier (IRRID): DERR1-10.2196/11067

Introduction

The average proportion of people older than 65 years in the Organisation for Economic Co-operation and Development countries increased from 9% in 1969 to 15% in 2010 and by the year 2050 will have reached approximately 27% [1]; at present, one-fifth (20.7%) of the German population is older than 65 years [2]. Already in the year 2020, which is approaching fast, 2 out of every 3 hospital beds will be occupied by patients older than 60 years—an age group that often has at least one chronic disease [3,4]. For medical care, this means that the symptoms that the admission diagnosis relates to should not be treated in isolation but that other preexisting diseases of patients also need to be treated appropriately [5]. This increasingly leads to problems in the medical care for elderly patients [6]. In addition, risks that arise from falls, malnutrition, and polypharmacy are often not given sufficient attention. Consequences can include poor treatment results with (partially) restricted quality of life or even mortality [5,6].

Globally, the next 20 years will see a duplication of the prevalence of dementia-type illnesses [7]. Especially the very old often suffer from cognitive impairments or dementia, which also increases the vulnerability of this patient group. This is however often not known upon admission and can have an adverse effect on the (course of) treatment. The specific vulnerability of elderly people demands specialized expert treatment.

Need for Qualification

The quality of elderly medical care depends to a major degree on the quality and quantity of motivated, well-trained professionals [8]. Three-fourths of the more than 1800 ward and department heads surveyed throughout Germany responded that they regard training of their staff in dementia topics as necessary [9]. Other studies show that the specialist diagnostic knowledge of physicians to distinguish between dementia and delirium often does not meet the necessary requirements [10].

Organizational processes in hospitals need to match the needs of this vulnerable patient group, especially patients suffering from dementia [11]. Therefore, the geriatric knowledge of all parties who come into contact with elderly people (physicians, nursing staff, therapists, social workers, and administration) needs to be improved [12].

The improvement of communicative conditions and skills of all care providers leads to a greater quality of care results, improved patient satisfaction, and higher work satisfaction of staff [13]. Furthermore, the exchange between colleagues can contribute to the positive development of collaboration across sector borders. In this way, it would be possible to prepare and define the postinpatient care environment already during an inpatient stay [14].

Qualification measures for the involved occupational groups are generally offered on a voluntary basis and often used by interested staff or organizations with a geriatric focus. However, institutions that would benefit the most from trainings in geriatric topics often do not take part. There is thus a risk that educational measures will not reach those care providers and institutions that have not yet recognized the importance of the demographic change [15].

Experiences with a work shadowing program at the St. Franziskus Hospital in Münster show that exclusively training 1 occupational group is insufficient because real changes to the care situation can only be achieved through interaction between different occupations [16].

The Qualitätsverbund Geriatrie Nord-West-Deutschland (in English: Quality Network of Geriatric Medicine in north-west Germany) is a group of more than 65 inpatient and outpatient institutions including hospitals with and without geriatric departments, inpatient geriatric rehabilitation facilities, elderly care institutions, outpatient care services, networks of doctor's offices, and outpatient rehabilitation providers.

Objectives

The main objective of this project is improved care for elderly people in different inpatient and outpatient institutions of the Quality Network. The project involves systematically evaluating knowledge transfer and learnings. In addition, necessary structural changes, for example in work procedures, for improvement of the care situation for elderly patients are to be introduced and established for the near future. This project is a study to evaluate geriatric training measures and thus not a clinical trial. The project is subsidized by the Landeszentrum Gesundheit Nordrhein-Westfalen (North Rhine-Westphalian Center for Health). Participating institutions incur no costs for the project.

The guiding research questions are as follows:

- Do all parties who come into contact with elderly people attend courses for geriatric topics? Are the participants satisfied with the content of the training? Do they suggest improvements?

- Is the transfer of expertise in their everyday working environments successful? Do the participants recommend that others take part in the training session?
- How many consensus groups involving people from different occupations in the institutions work toward targeted measures for improved care of elderly patients in their place of work? Is the implementation of measures/actions effective?
- Do patients notice any positive change in view of the treatment at the end of the project?
- Do physicians and caregivers apply geriatric screenings and assessments more often at the end of the project?

Methods

Subjects

On the network's initiative, the level and need for education of the staff were analyzed in 26 institutions before the actual start of the project. Data are available for 2200 people.

At least 1000 people who took part in a total of 100 trainings on geriatric topics were included in the recent education evaluation. This sample includes people from all hierarchy levels and occupational groups who come into contact with elderly patients (see Figure 1). In addition, 125 consensus group meetings (with approximately 4-8 participants) are to be

evaluated so far so that approximately. 500 to 1000 more people are involved in this evaluation step.

Participation in the study was voluntary. All participants were informed in writing about the aims and process of the study before anonymous data collection. Participants agreed with a second survey 6 months later (paper or Web-based), and for this reason, many provided their email address to the study leaders. The Web-based questionnaire was presented by a professional tool that guaranteed data protection, data security, and anonymity. As the study's purpose was to evaluate training measures, the prior permission from an ethics commission was not required.

The patient sample includes people who are 75 years old or older and who are being treated for 3 days or longer as an inpatient for accident or abdominal surgery. At the beginning and near the end of the data collection in the project, patients and their family members are surveyed about inpatient treatment and the quality of life and health. Patients and their families are informed in writing and also told about the aims of the project and provide their written permission to take part in the survey. Participation is voluntary. There are plans to survey 60 patients and family members at the start (as a prestudy) and another 60 patients and family members at the end (poststudy) of data collection. In total, approximately 1800 people are included in the project.

Figure 1. Overview of occupational groups involved in the educational initiative.

Design

The educational initiative aims at all staff members of the institutions who are included in the care and treatment of elderly, multimorbid patients. This includes nursing staff, physicians, therapists, auxiliary staff, medical assistants, social workers, medical students, trainees, as well as people from administration and management.

The educational measures are evaluated at various measuring times (before and after) and with regard to various contents and target groups.

As the *baseline*, the level and need for education of staff members regarding geriatric topics are assessed. All occupational groups that have contact with elderly patients (see Figure 1) are included in the survey. This should identify the training topics that are particularly important or missing from the training offer. This assessment took place before the actual project start on the Quality Network's own initiative. In total, 26 institutions took part with more than 8200 respondents.

The evaluation mainly comprises existing training offers about 20 diverse subject areas with a geriatric focus. The educational measures are on topics such as *Dealing with patients with dementia*, *Using geriatric screenings and assessments*, or *Nutrition for the elderly*. Depending on the topic, the trainings are intended for caregivers, physicians, other health professionals, and various management staff. The questionnaires for the evaluation were developed at the start of the project, together with academic experts, based on the standards for transfer-oriented education [17]. *Directly before* and *directly after* geriatric trainings, the participants are asked about their expectations and their satisfaction regarding the course. The regular feedback from course participants should indicate whether prior knowledge and experiences are considered to adapt the training offer according to needs in the medium term. *After 6 months,* the participants are asked how well they have transferred the training content into their day-to-day work. Both directly after the course and 6 months later, the participants specify whether they would recommend that others take part in the training session. Ideas from the trainings should be specifically implemented in the network institutions and adapted to the specific requirements in inpatient and outpatient settings. Work groups/quality groups involving people from different occupations in the institutions and using the nominal group technique [18] work toward targeted measures for improved care of elderly patients at their place of work. The consensus groups are evaluated *alongside* their activities and *3 months after completion*. To do so, meeting minutes, questionnaires, and interview data are evaluated qualitatively and quantitatively [8,19].

To determine a patient-related outcome, assessments from patients and family members/caregivers as well as various aspects of the treatment are gathered. *At the start* and *near the end of the data collection*, patients and family members assess the treatment. In this context, it is recorded whether and which specific geriatric screenings and/or assessments were conducted during the treatment. A meta-analysis about comprehensive geriatric assessments showed that use of assessments led to improved patient-related outcomes as well as reduced costs [20]. Moreover, applying assessments helps to identify vulnerable patients and adapt therapies to their specific needs better than without assessments [21,22].

The questionnaire for patients includes the items for the German language Short Form (SF)-12 for recording health-related living quality [23,24], and it also includes items relating to the course of treatment. The content of the family member questionnaire corresponds to the patients' questionnaire. The items of each questionnaire were developed together with clinical experts, with the exception of the SF-12 items. An overview of the measuring times and evaluation instruments is depicted in Figure 2.

Material

Gathering Data on Level and Need for Education

The questionnaire used before project start for recording the level of education and need for education about geriatric topics was divided into 4 chapters, each dealing with 1 theme. The first chapter included 3 items. The staff members specified whether further training about geriatric topics is offered in their organization and whether they themselves have taken part in internal or external geriatric trainings. The questions could be answered with "Yes," "No," or "No response." The second chapter included a list of 28 geriatric training topics such as "Communicating with patients who have dementia" or "Promoting mobility." The staff members judged their needs for training about these topics (possible answers were "Yes," "No," or "No response") and selected any trainings that they had already taken part in. The third chapter focused on assessing prior trainings. In total, 19 items related to aspects of preparation and conducting of prior trainings as well as the implementation of content into day-to-day work. Respondents could choose from "Applies in all cases," "Applies partially," "Hardly applies," "Does not apply," and "No response." The final question was based on an assessment of the likelihood of recommendation and was phrased as follows: "I recommend participating in trainings on geriatric topics to my colleague." The possible answers were the same as above. The last part of the survey was for demographic details.

Figure 2. Overview of measuring times and evaluation instruments.
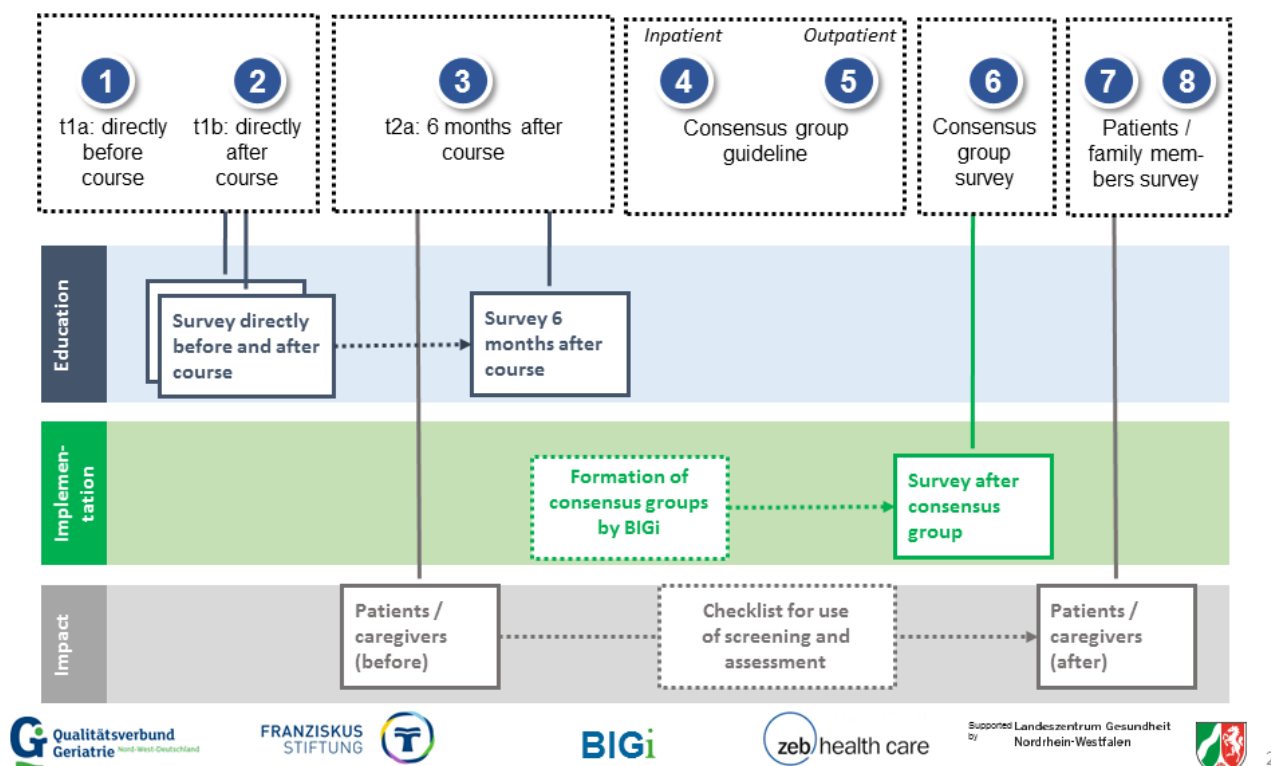
Evaluation of Geriatric Trainings

Findings from economic education research point out that success factors for educational measures are to be recorded using a phase-oriented perspective (preparation of the measure, conduct, transfer of learnings; [25,26]). This phase-oriented approach was considered in the development of the questionnaires for evaluating geriatric trainings.

The questionnaire filled out by participants *immediately before* the training includes 12 statements that can be answered with the following alternative responses: "Applies in all cases," "Applies partially," "Hardly applies," "Does not apply," and "No response." The first 2 items relate to the motivation to take part in the course (personal interest and recommendation from employer). The remaining statements mention options for personal preparation and expectations in the course and regarding implementation of the content into day-to-day work. For example, the items include the following: "I have received a lot of information about the contents of this course in advance," "It is important to me that this course provides sufficient time to consolidate and practice," and "It is important to me that there are sufficient opportunities after this course to put the learnings into practice." Respondents can also note their personal wishes regarding the respective courses in 3 free textboxes. The questions are as follows: "What is particularly important to you in this course?," "Future courses must...," and "In future courses, I do not want to...." Finally, demographic data are collected.

The survey filled out *immediately after* the training includes 10 statements with the possible responses: "Applies in all cases,"

"Applies partially," "Hardly applies," "Does not apply," and "No response." The items relate to the assessment of training content, the behavior of the trainer, and the support of the employer. They are phrased as follows: "The contents of the course matched completely with my needs in day-to-day work," "The course trainer recognized problems and needs of participants and addressed them in the course," and "My organization ensured my smooth participation in the course." The tenth statement measures the likelihood of recommendation and is phrased as follows: "All in all, I would recommend this training to a colleague." Here also, 3 free text boxes are provided for participants to note their positive and negative impressions about the training as well as recommendations for improvement. Finally, demographic details are requested.

The survey on *implementing training content into day-to-day work* is sent *6 months after the end of the course* as either a paper or Web-based version. The 6 statements have the aforementioned answer categories and relate to the practical use of the training content and the perceived support and appreciation from colleagues and/or line managers in implementing the training content into day-to-day work. The items include, for example, "The contents of the training help me in practical dealings with elderly patients" and "My knowledge from the training and my efforts in the implementation of improvement measures were appreciated in my organization." The final statement is to measure the likelihood of recommendation. In 2 free textboxes, the respondents can note whether and which training content they can implement in their day-to-day work. The final section is a collection of demographic data.

Figure 3. Agenda for a consensus group meeting.
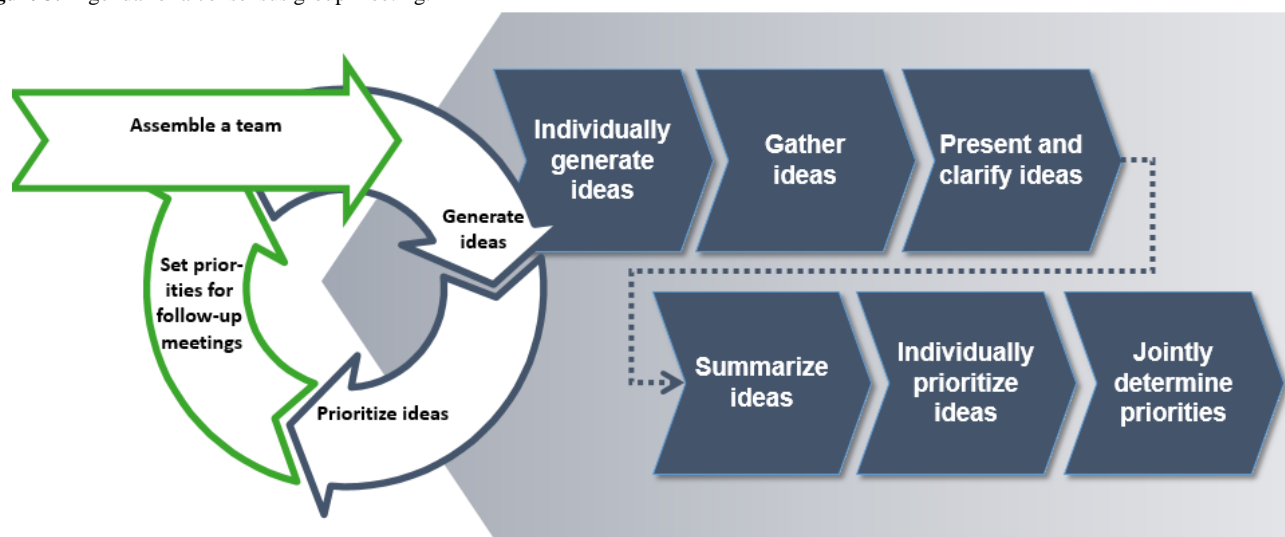
Conducting and Evaluating Consensus Groups

A guideline (including information letters and worksheets) for preparing and conducting consensus groups was developed for practical use as part of the project. It serves as a handbook for facilitators who give the meetings organizational and content structure. Appropriate adjustments were made to consider the diverse requirements in inpatient and outpatient settings. The agenda for a consensus group meeting is presented in Figure 3.

Meeting minutes document the characteristics, such as duration, topic, preferred ideas, occupational groups, and gender of the facilitator. At the end of the consensus group meeting, the participants and the facilitator assess the group work in writing based on 4 statements and the possible responses: "Applies in all cases," "Applies partially," "Hardly applies," "Does not apply," and "No response." The items are as follows: "Suggestions for improved care for elderly, multimorbid patients were developed and assessed together," "The cooperation of various occupational groups was helpful for developing improvement suggestions for the benefit of elderly, multimorbid patients," and "The fact that the meeting(s) were supported by a facilitator was helpful for the content-related work of the group." The final question is based on the likelihood of recommendation. In 3 free textboxes, participants can note what they found to be helpful or unhelpful for the work in the consensus group and any ideas they might have for improving similar work groups. Then 3 months after the completion of the consensus groups, institution managers are interviewed in a 25-min, structured phone call about whether and to what degree the initiated change suggestions were able to be implemented in the institution.

Assessment of the Treatment-Related Outcome

The treatment-related outcome is assessed from the following 2 perspectives: (1) use of geriatric screening and assessment instruments during the course of treatment, and (2) assessment of inpatient treatment by patients and family members.

Patient and Family Member Survey

Surveys for patients and family members relate to each other in terms of content and mentioned aspects of inpatient treatment

and later discharge, which are particularly relevant for elderly, multimorbid patients. For example, there are questions on whether the patients receive sufficient help in eating and drinking as well as explanation about outpatient support. All items are phrased as questions and adapted to the respective readers, for example, "Were your pre-existing conditions/accompanying conditions considered by the nursing staff during your treatment?" (patient survey) compared with "Were the pre-existing conditions/accompanying conditions of your family member considered by the nursing staff during treatment?" (family member survey). The response categories are "Yes, completely," "Somewhat," "No," "Was not necessary," and "I don't know."

An assessment of the state of health before the current hospital stay is requested in both surveys. The patient survey includes the SF-12 survey as a standardized measurement instrument about the state of health [23]. In addition, demographic data about the patients is gathered.

Further Measures Including Frequency of Geriatric Screenings/Assessments

A check list is used to record whether any screenings and assessments, which are recommended for elderly medical care [27], were conducted during inpatient treatment. By using the screening instruments, important diagnostic information for the treatment can be gathered, for example, regarding the general level of functioning, the nutritional state, or the risk of delirium of the patient [28], as these aspects are considered to be of high relevance for an optimal outcome especially after surgery [29,19].

Furthermore, it is recorded whether the social services were involved in the preparations for discharge and/or family members were informed about current medication or nutritional recommendations, as sufficient information seems to have a positive influence in the course of treatment [30]. In addition, demographic patient data are gathered. Within the project, neither individual patient data nor any screening or assessments results have been stored. Rather, it has only been recorded whether screening and assessment instruments have been used and whether the main aspects—mobility, nutrition and

cognition—have been covered because screening and assessments have proven to show positive impact on output and outcome of geriatric treatment [20-22].

Procedure

To prepare for the evaluation, at the start of the project, the following 3 work groups were formed, which included project managers as well as experts from the fields of geriatrics, surgery, general medicine, inpatient and outpatient care, therapy, further training, quality management, and administration:

- The first work group discussed and agreed upon the evaluation tools, such as surveys, and the procedure how to collect data from the geriatric training courses in the participating institutions.
- The second work group addressed the conducting and evaluation of consensus groups/quality groups.
- The third work group focused on recording treatment-related outcomes. The assessment of the treatment by patients and family members was regarded as fundamental. Furthermore, the recording of certain geriatric screenings/assessments as well as discharge management were seen as important.

Data collection began in June 2016 and was completed by the end of 2018. Data analysis and final reporting are currently in progress.

Data Collection on Level of Education

On the Quality Network's initiative, the level and need for education of staff members regarding geriatric topics were assessed in advance. Members of various occupational groups who are involved in the care for multimorbid, elderly patients received a survey about the level and need for education. The results of these surveys were assessed specifically per institution. At the same time, each participating institution received a benchmark report to be able to compare the results of their own organization with those of others. The latest survey results are used in revising existing educational measures and are considered in the development of new ones.

Evaluation of Geriatric Trainings

Training participants complete a survey relating to their expectations at the start of the course and another about their level of satisfaction with the course at the end. The surveys are handed out, collected, and submitted to the project manager by the trainer. The education experts at the institutions receive a quantitative and qualitative analysis for the various courses. Then, 6 months after the end of the training, course participants specify to what degree they have been able to implement the education content into their day-to-day work. This survey on the subsequent assessment is sent to the participants as a paper of Web-based version.

Conducting and Evaluating Consensus Groups

Generally, 4 to 8 staff members from various occupational groups, fields, and hierarchy levels take part in a consensus group/quality group (up to 4 times for approximately 30-60 min). Relating to a specific question about improved care for elderly patients, ideas for measures should be developed in the group, and the implementation of these measures should be initiated. For conducting and evaluating, various work materials

including guidelines have been developed. A facilitator can support the group in finding ideas and developing a common suggestion. Using meeting minutes, specific course characteristics such as the topic of the work meeting and the preferred suggestion should be documented. In a survey at the end of the consensus group, the group work will be assessed, and meeting minutes will be created. Finally, the facilitator will submit the suggestion to the institution's management and ask for a review and approval. Upon approval, the details about specific measures will be recorded in writing and their implementation into day-to-day work will begin.

Then, 3 months after the completion of the consensus groups/quality groups, representatives of the institutions will be interviewed in a structured interview about whether and to what degree the initiated change suggestions were able to be implemented.

Treatment-Related Outcome

The patient sample includes people who are 75 years old or older and who are being treated for 3 days or longer as an inpatient for accident or abdominal surgery. At the beginning and near the end of the data collection in the project, patients and their family members are surveyed about the completed treatment and the quality of life and health. Surveys for patients and family members relate to each other in terms of content and mention aspects of inpatient treatment and later discharge, which are particularly relevant for elderly, multimorbid patients, among other topics. An assessment of the state of health before the current hospital stay is requested in both surveys. The patient survey includes the SF-12 survey as a standardized measurement instrument about the state of health [26].

To assess the outcome, a checklist is used to record whether certain screenings and assessments that are recommended for elderly medical care [26] were conducted during inpatient treatment. By using these instruments, significant diagnostic information for the treatment can be gathered, for example, regarding the general level of functioning, the nutritional state, or the risk of delirium of the patient [27]. Furthermore, it is analyzed whether the social services were involved in the preparations for dismissal and/or family members were informed about current medication. Patient-related information about the courses of treatment is not part of the educational evaluation.

Results

We expect academically sound conclusions about how the patient-related medical outcome is improved through the transfer of educational and accompanying measures in the intersector care of elderly patients. Furthermore, particularly changes at the Quality Network's institutions actively involved in education and those who offer and carry out many courses should be compared with the changes of results in less active institutions. We expect that this comparison will provide detailed conclusions about the necessary adjustments to standard procedures, process instructions, organizational guidelines, educational plans, and training guidelines.

From the evaluation of how educational content has been implemented after visiting a course and through the formation

of consensus groups/quality groups, we expect conclusions about the economically viable use of educational measures in improving patient care. The before and after comparison at institutions with high transfer rates, that is, specific implementations of learnings through measures that are developed in consensus groups/quality groups, is also expected to provide starting points for optimizing the use of medical resources and preventing cost-driving revolving door effects.

Discussion

The individual competence of individual members of the treatment team for geriatric patients can be viewed as a source of collective competence development and vice versa. The knowledge to be gained for improving elderly medical care is meant to facilitate finding solution patterns for new tasks and shifting requirements and/or transferring knowledge and competence to the organization, the group, and the Quality Network in cooperation with other experts and thus to promote organizational learning [31].

First Results

Initial results and experiences clearly show that the offers of the interprofessional educational initiative in geriatrics are met with serious interest from all occupational groups and that the various groups are all taking part. At the same time, the survey results about the level and need for education confirm the results of other studies in this field [9,16]. The survey results also give the responsible people in participating institutions the possibility to address the individual training needs of their staff members. They can then work on ensuring that any "blind spots" in geriatric fields are closed and that their institutions are best prepared for demographic changes.

Through joint trainings over the course of the project, the exchange between colleagues in inpatient and outpatient settings can improve so that it is not only various occupational groups that benefit but also patients and family members. Due to the broad participation in the project, it is already becoming apparent that the participating institutions have implemented numerous improvements to the care of geriatric patients.

Side-Effects of Training Evaluation

The project has already contributed to important innovations for both the member organizations of the Quality Network of Geriatric Medicine and also for the educational work for health care of elderly, multimorbid patients, which are as follows:

- The educational measures and their evaluation include various occupational groups, hierarchy and management levels, as well as sectors in inpatient and outpatient settings.
- Unlike the usual initiatives, the evaluation not only includes conducting the course but also the initial conditions as well as the implemenfation of learnings at the institutions.
- The evaluation includes gaining knowledge and educational success, but it also focuses on relevant outcome variables to identify treatment improvements for elderly patients and their family members. New methods are developed and tested for this purpose.
- The evaluation is itself part of the change and improvement process in which adaptive questioning and testing instruments are used, which contribute to the various improvements in the institutions. It remains to be seen whether the methods developed in the project such as the evaluation surveys and the process model for the consensus groups will be adjusted by the institutions for their purposes and adopted.
- The evaluation of various measures in 1 network is the first of its kind. This comprehensive evaluation can provide stimulus for the entire interconnected health care sector.

The overall links between training needs, training assessment, implementation efforts, and measurable results can only be assessed once the project has finished. The findings will be considered in the development and implementation of later professional educational work and its evaluation. The interprofessional educational initiative not only consists of various training courses but also their evaluation. One important element is the exchange between professionals, which also acts as promotion for the initiative itself. The competitive edge, which results from the professional range of treatment and care options in the Quality Network of Geriatric Medicine in north-west Germany, is not to be underestimated. Upon completion of the project, the full results will be made available to the public so that other institutions can also benefit from the results.

Conflicts of Interest

None declared.

References

1. Organisation of Economic Co-operation and Development. 2019 Feb 22. Health at a Glance 2017 URL: http://www.oecd.org/health/health-systems/health-at-a-glance-19991312.htm [accessed 2019-04-15] [WebCite Cache ID 77ennHtN8]

XSL•FO
RenderX

2. Weyerer S. Epidemiology and demographic change. In: Pantel J, Schröder J, Bollheimer C, Sieber C, Kruse A, editors. [Practical Handbook for Geriatric Medicine Geriatrics-Gerontopsychiatry-Gerontology]. Stuttgart, Germany: Kohlhammer; 2014:34-49.

3. Deloitte HC. Ärzteblatt. 2014. [Auditors expect a lack of hospital beds from 2020 onwards] URL: https://www.aerzteblatt.de/nachrichten/58645/Wirtschaftspruefer-erwarten-Bettenmangel-im-Krankenhaus-ab-2020 [accessed 2019-04-15] [WebCite Cache ID 77eolPUDb]

4. Santoni G, Angleman S, Welmer A, Mangialasche F, Marengoni A, Fratiglioni L. Correction: age-related variation in health status after age 60. PLoS One 2015;10(6):e0130024 [FREE Full text] [doi: 10.1371/journal.pone.0130024] [Medline: 26039712]

5. Elkeles B, Schandner O, Socha U, Tunn R. [The Elderly Patient: Prevention and Treatment More Frequent Complications]. München, Germany: Hans Marseille Verlag; 2012.

6. Eifert B. Landesseniorenvertretung. 2016. The main topic: Political influence now! URL: https://lsv-nrw.de/wp-content/uploads/2016/07/NRW_Ausgabe_96.pdf [accessed 2019-04-16] [WebCite Cache ID 77gDQFp3r]

7. Karakaya T, Fußer F, Pantel J. [Dementia and mild cognitive impairment]. In: Pantel J, Schroeder J, Bollheimer C, Sieber C, Kruse A, editors. [Practical Handbook for Geriatric Medicine]. Stuttgart, Germany: Kohlhammer; 2014:299-330.

8. Kolb G, Andersen-Ranberg K, Cruz-Jentoft A, O´Neill D, Topinkovae E, Michel JP. Geriatric care in Europe – the EUGMS Survey part I: Belgium, Czech Republic, Denmark, Germany, Ireland, Spain, Switzerland, United Kingdom. Eur Geriatr Med 2011 Oct;2(5):290-295. [doi: 10.1016/j.eurger.2011.06.007]

9. Isfort M, Klostermann J, Gehlen D, Siegling B. Deutsches Institut für angewandte Pflegeforschung. Köln, Germany; 2014. [Dementia in the hospital: care thermometer 2014] URL: https://www.dip.de/fileadmin/data/pdf/projekte/Isfort_Demenz_im_Krankenhaus_SP_8_2014_740-749.pdf [accessed 2019-04-18] [WebCite Cache ID 77jv1eJun]

10. Griffiths A, Knight A, Harwood R, Gladman J. Preparation to care for confused older patients in general hospitals: a study of UK health professionals. Age Ageing 2014 Jul;43(4):521-527 [FREE Full text] [doi: 10.1093/ageing/aft171] [Medline: 24165310]

11. Büter K, Motzek T, Dietz B, Hofrichter L, Junge M, Kopf D, et al. [Dementia-friendly hospital wards: expert recommendations for planning and design]. Z Gerontol Geriatr 2017 Jan;50(1):67-72. [doi: 10.1007/s00391-016-1079-0] [Medline: 27325444]

12. Brandenburg H. [Interdisciplinarity and cooperation]. In: Pantel J, Schroeder J, Bollheimer C, Sieber C, Kruse A, editors. [Practical Handbook for Geriatric Medicine]. Stuttgart, Germany: Kohlhammer; 2014:65-75.

13. Baller G, Schaller B. [Communication in the Hospital: Successfully Communicate With Patients, Physician Colleagues and Hospital Staff]. Berlin, Germany: Springer; 2017.

14. Schilder M. Geriatrics (care case-oriented learning and teaching). Stuttgart, Germany: Kohlhammer; 2014.

15. Gurlit S, Möllmann H. [Preoperative risk identification using the Identification of Seniors at Risk?: suitability as sole screening tool for inpatient aged risk patients]. Z Gerontol Geriatr 2018 Jun;51(4):388-393. [doi: 10.1007/s00391-018-1401-0] [Medline: 29796867]

16. Gurlit S, Thiesemann R, Wolff B, Brommer J, Gogol N. Caring for people with dementia in general hospitals: an education curriculum from the Alzheimer's Society of Lower Saxony, Germany. Z Gerontol Geriatr 2013 Apr;46(3):222-225. [doi: 10.1007/s00391-013-0479-7] [Medline: 23474868]

17. Zurwehme A. [Controlling in further education institutions - structure and results of an empirical study]. [Journal of Planning & Corporate Management] 2008;18(4):445-469. [doi: 10.1007/s00187-007-0034-z]

18. Delbecq AL, Van de Ven AH, Gustafson DH. Guidelines for Conducting NGT Meetings. In: Group Techniques for Program Planning: A Guide to Nominal Group and Delphi processes. Glenview, IL: Foresman; 1975.

19. Paillaud E, Soubeyran P, Caillet P, Cudennec T, Brain E, Terret C, G-CODE collaborators. Multidisciplinary development of the Geriatric Core Dataset for clinical research in older patients with cancer: a French initiative with international survey. Eur J Cancer 2018 Nov;103:61-68. [doi: 10.1016/j.ejca.2018.07.137] [Medline: 30212804]

20. Eamer G, Saravana-Bawan B, van der Westhuizen B, Chambers T, Ohinmaa A, Khadaroo RG. Economic evaluations of comprehensive geriatric assessment in surgical patients: a systematic review. J Surg Res 2017 Dec;218:9-17. [doi: 10.1016/j.jss.2017.03.041] [Medline: 28985882]

21. van Loon IN, Wouters TR, Boereboom FT, Bots ML, Verhaar MC, Hamaker ME. The relevance of geriatric impairments in patients starting dialysis: a systematic review. Clin J Am Soc Nephrol 2016 Dec 7;11(7):1245-1259 [FREE Full text] [doi: 10.2215/CJN.06660615] [Medline: 27117581]

22. Rodin MB, Mohile SG. A practical approach to geriatric assessment in oncology. J Clin Oncol 2007 May 10;25(14):1936-1944. [doi: 10.1200/JCO.2006.10.2954] [Medline: 17488994]

23. Bullinger M, Kirchberger I. [SF-36 Health questionnaire]. Göttingen, Germany: Hogrefe; 1998.

24. Gandek B, Ware J, Aaronson N, Apolone G, Bjorner J, Brazier J, et al. Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol 1998 Nov;51(11):1171-1178. [Medline: 9817135]

25. Goetze W, Zurwehme A. [Practical Science or Scientific Practice. For the development of tailor-made control instruments for further education providers]. [Continuing Education] 2016(1):40-43.

26. Zurwehme A. Does Management Control Influence Organizational Performance in Service Units. The Case of Training Institutions. Dresden: Dresden University of Technology; 2007.

27. Olotu-Steffen C, Gurlit S, Kiefmann R. [Preoperative evaluation and arrangements: the elderly patient]. Anasthesiol Intensivmed Notfallmed Schmerzther 2017 May;52(5):342-355. [doi: 10.1055/s-0042-109286] [Medline: 28561149]

28. Freund H. [Geriatric Assessment and Test Procedure. Basic Terms Guidance Paths (2nd edition)]. Stuttgart, Germany: Kohlhammer; 2014.

29. Engelter S, Kressig R. [Special features of modern rehabilitation of elderly people]. Praxis (Bern 1994) 2018 Feb;107(4):193-196. [doi: 10.1024/1661-8157/a002913] [Medline: 29439633]

30. Korczak D, Steinhauser G, Kuczera C. Deutsches Institut für Medizinische Dokumentation und Information. Berlin; 2012. [Effectiveness of outpatient and inpatient geriatric rehabilitation in patients with secondary dementia dementia] URL: https://portal.dimdi.de/de/hta/hta_berichte/hta331_bericht_de.pdf [accessed 2019-04-16] [WebCite Cache ID 77gHSxB7S]

31. Wilkens U, Kettler H, Schmette M. [Interaction between levels of individual and collective competence. Theory access and modeling]. In: Schreyögg G G, Conrad P, editors. [Management Research, Volume 16: Management of Competencies]. Wiesbaden, Germany: Gabler; 2006:121-161.

Abbreviations

SF: Short Form

Protocol

Implementing Systematic Screening and Structured Care for Distressed Callers Using Cancer Council's Telephone Services: Protocol for a Randomized Stepped-Wedge Trial

Elizabeth A Fradgley[1,2,3], PhD; Anna Boltong[4], PhD; Lorna O'Brien[5], RN; Allison W Boyes[2,3,6], PhD; Katherine Lane[7], RN; Annette Beattie[8]; Tara Clinton-McHarg[1,2,9], PhD; Paul B Jacobsen[10], PhD; Christopher Doran[11], PhD; Daniel Barker[3], PhD; Della Roach[1,2,3], BioMed Hon I; Jo Taylor[1,2,3], MPH; Christine L Paul[1,2,3], PhD

[1]Priority Research Centre for Cancer Research Innovation and Translation, University of Newcastle, Callaghan, Australia

[2]Priority Research Centre for Health Behaviour, University of Newcastle, Callaghan, Australia

[3]School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

[4]Strategy and Support, Cancer Council Victoria, Melbourne, Australia

[5]St John of God Hospital, Richmond, Australia

[6]Hunter Medical Research Institute, Callaghan, Australia

[7]Cancer Information and Support Services, Cancer Council Victoria, Melbourne, Australia

[8]Cancer Information and Support Services, Cancer Council NSW, Woolloomooloo, Australia

[9]School of Psychology, University of Newcastle, Callaghan, Australia

[10]Healthcare Delivery Research Program, National Cancer Institute, Bethesda, MD, United States

[11]Centre for Indigenous Health Equity Research, School of Health, Medical and Applied Sciences, University of Central Queensland, Brisbane, Australia

Corresponding Author:
Elizabeth A Fradgley, PhD
Priority Research Centre for Cancer Research Innovation and Translation
University of Newcastle
Level 4
HMRI Building
Callaghan, 2308
Australia
Phone: 61 0240420642
Email: elizabeth.fradgley@newcastle.edu.au

Abstract

Background: Structured distress management, comprised a 2-stage screening and referral model, can direct supportive care resources toward individuals who are most likely to benefit. This structured approach has yet to be trialed in Australian community-based services such as Cancer Council New South Wales (NSW) and Victoria Cancer Information and Support (CIS) 13 11 20 lines who care for a large community of cancer patients and caregivers.

Objective: The aim of this study was to evaluate the effectiveness of structured screening and referral in (1) increasing the proportion of distressed CIS callers who accept supportive care referrals and (2) reducing distress levels at 6-month follow-up.

Methods: In this stepped-wedge trial, Cancer Council NSW and Victoria CIS consultants are randomized to deliver structured care during inbound 13 11 20 calls in accordance with 3 intervention periods. Eligible callers are patients or caregivers who score 4 or more on the Distress Thermometer; NSW or Victorian residents; aged 18 years or older; and English proficient. Study data are collected via computer-assisted telephone interviews (CATIs) at 3- and 6-month follow-up and CIS record audit. CATIs include demographic and service use items and the General Health Questionnaire (GHQ-28) to assess distress. An economic analysis of the structured care model will be completed.

Results: The structured care model was developed by guideline review and identification of service characteristics to guide mapping decisions; place-card methodology; and clinical vignettes with think-aloud methodology to confirm referral appropriateness. The model includes an additional screening tool (Patient Health Questionnaire-4) and a referral model with 16-20 CIS services. Descriptive statistics will be used to assess referral uptake rates. Differences between GHQ-28 scores for structured and usual care callers will be tested using a generalized linear mixed model with fixed effects for intervention and each time period. The

trial will recruit 1512 callers. The sample size will provide the study with approximately 80% power to detect a difference of 0.3 SD in the mean score of the GHQ-28 at an alpha level of .05 and assuming an intra-cluster correlation of .04. A random sample of recorded calls will be reviewed to assess intervention fidelity and contamination. To date, 1835 distressed callers have been invited to participate with 60.71% (1114/1835) enrolled in the study. A total of 692 participants have completed 6-month CATIs. Recruitment is anticipated to end in late 2019.

Conclusions: This trial is among the first to rigorously test the outcomes of a community-based structured approach to distress management. The model is evidence-informed, practice-ready, and trialed in a real-world setting. The study outcomes will advance the understanding of distress management internationally for both patients and caregivers.

Trial Registration: Australian New Zealand Clinical Trial Registry ACTRN12617000352303; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=372105&isReview=true (Archived by WebCite on http://www.webcitation.org/78AW0Ba09)

International Registered Report Identifier (IRRID): DERR1-10.2196/12473

Introduction

Identifying and Responding to Cancer-Related Distress

Cancer-related distress is defined as a multifactorial unpleasant emotional experience of a psychological, social, or spiritual nature which may interfere with the ability to cope effectively with the disease, its symptoms, and treatment [1]. A recent literature review of symptom prevalence suggests distress is experienced by approximately 40% of cancer patients [2]. Australian longitudinal data have also demonstrated that up to 31% of caregivers experience borderline or clinical anxiety [3].

Distress and psychological morbidity among people affected by cancer are associated with decreased social functioning, more intense physical symptoms, cognitive impairment, poor adherence to treatment, and reduced length of life [4,5]. Internationally, there are distress screening and management guidelines available; Australian examples include Cancer Australia's Clinical Guidance for Responding to Suffering in Adults with Cancer [6] and the Clinical pathway for the screening, assessment and management of anxiety and depression in adult cancer patients [7]. Guidelines recommend using a brief distress screening tool, such as the distress thermometer (DT) or Edmonton Symptom Assessment Scale.

Evidence from clinical settings suggests timely identification when paired with structured management of psychological distress can improve medical management and reduce distress [8-10]. This evidence must be cautiously interpreted as other studies demonstrate inconsistent or minimal improvement in patient outcomes following screening [11]. The debate surrounding the utility of distress screening is ongoing and complex [12,13]. However, any program that involves tokenistic distress screening without a feasible referral pathway is unlikely to influence patient outcomes or experiences. Furthermore, screening models may be improved by focusing on patients' adaptive or maladaptive reactions to their distress in addition to severity [14]. This trial will contribute further data to the debate by involving new settings in which structured distress management has not yet been trialed.

Incorporating and Evaluating Distress Screening Practices in Telephone-Based Services

The International Psycho-Oncology Society emphasized that "Distress should be recognized, monitored, documented and treated promptly at all stages of disease and *in all settings*." [15]. One setting in which structured care might be implemented routinely is the community-based Australian Cancer Council Cancer Information and Support (CIS) telephone service. The CIS model of care operates in numerous countries [16] including the United Kingdom [17] and Australia [18]. Under the Australian CIS model, health professionals provide emotional, practical, and informational telephone support to both patients and caregivers [18]. Strengths of the model include its ability to assist individuals who are unable to receive traditional face-to-face supportive care owing to geographical isolation or poor physical health [17]. In 2017, the Australian CIS 13 11 20 telephone service received 46,000 calls nationally [19], of which, the New South Wales (NSW) CIS received 12,225 calls [20] and Victorian CIS service received 11,429 calls [21].

In total, 2 Australian studies established the acceptability and feasibility of implementing distress screening and tiered care in the CIS context [22,23]. These exploratory studies revealed that, despite screening acceptability, a low proportion of callers take up the referrals that are offered. Poor referral uptake (approximately 20% to 25%) has also been reported in hospital-based services [24,25]. As additional psychosocial care is associated with improved emotional well-being and quality of life, it is critical to maximize the proportion of distressed patients and caregivers acting upon these referrals [26].

This stepped-wedge trial rigorously tests the uptake, likely impact, and costs of a structured care approach to distress screening and management across the NSW and Victorian CIS telephone services. The trial compares the effectiveness of a distress screening model using the DT only (ie, usual care) against a 2-staged distress screening model incorporating the DT, Patient Health Questionnaire (PHQ-4), and a referral model (ie, structured care model). Effectiveness of the structured care model is gauged by distressed callers' referral uptake rates and 6-month distress levels measured by the General Health

Questionnaire (GHQ-28). Usual and structured care are delivered by CIS consultants during inbound calls from distressed cancer patients and caregivers.

This trial was prospectively registered with the Australian New Zealand Clinical Trials Registry (ACTRN12617000352303) on March 8, 2017.

Aims

To identify the following:

1. The proportion of distressed people affected by cancer who call the CIS service and take up an offer of referral under the structured care model.
2. Whether the structured care model reduces the level of distress at 6-month follow-up among people affected by cancer when compared with usual care.
3. The relative costs and benefits of structured versus usual care models at 6-month follow-up from the service provider perspective.

Hypotheses

At 6-month follow-up:

1. There will be at least a 20% increase in the proportion of distressed callers who are offered a referral under the structured care model and accept the offer.
2. Distressed callers who receive structured care referrals will report lower scores on the GHQ-28 (0.3 SD lower) at 6 months when compared with those who receive usual care.

3. Structured care will incur higher service delivery costs per distressed caller than usual care. These higher costs will be considered appropriate by the service provider and consumers to reduce caller distress.

Methods

Study Design

The stepped-wedge trial is conducted with the Cancer Council NSW and Victoria CIS 13 11 20 line. The structured care model is sequentially rolled out with CIS consultants randomly allocated to transition to structured care over 3 intervention periods (Table 1). There is a 1-month transition period between each intervention period. Transition periods are included in reporting guidelines for stepped-wedge trials [27]; the transition periods allow for consultants to trial the new structured care call content (PHQ-4 and referral model), discuss and iteratively refine new content with other structured care consultants and receive feedback from the research team.

On the basis of previous CIS call volumes, the trial is approximately 24 months with the opportunity to adjust timeframes in accordance with CIS caller recruitment rates and other internal requirements (eg, implementation of a new electronic medical record system). The study is reported according to SPIRIT recommendations [28].

Table 1. Consultant allocation to intervention periods with caller sample size.

Consultant[a]	Intervention period (n=504)				
	1	Transition	2	Transition	3
A	Usual care	→	Structured care	__b	Structured care
B	Usual care	→	Structured care	__b	Structured care
C	Usual care	→	Structured care	__b	Structured care
D	Usual care	→	Structured care	__b	Structured care
E	Usual care	__b	Usual care	→	Structured care
F	Usual care	__b	Usual care	→	Structured care
G	Usual care	__b	Usual care	→	Structured care
H	Usual care	__b	Usual care	→	Structured care

[a]Each Cancer Information and Support (CIS) service has at least 4 consultants participating in the study.

[b]Not applicable. No transition required.

Caller Eligibility and Recruitment

Caller Eligibility

Eligible participants are inbound callers to the 13 11 20 CIS who reside in NSW or Victoria; are aged 18 years or older; have been diagnosed with cancer or support someone with cancer; have DT scores of 4 or more; and have provided consent to telephone follow-up. A meta-analysis of use of the DT with patients diagnosed with cancer found a sensitivity of 81% and

specificity of 72% to detect distress using a cut-off score of 4 [29].

Caller Recruitment

At the end of the inbound call, eligible individuals are invited by consultants to participate and are asked for permission to pass their contact details to the researchers. After receiving the contact details, the research team post study packages to potential participants with 3 consent options: written consent via return post; electronic consent via online form or email; or verbal consent with the research team. Individuals who do not

return a consent form within 10 days are contacted by telephone. Basic data on callers who either declined to provide their contact or declined participation after receiving a study package will be analyzed to ascertain any consent bias. See Figure 1 for a brief overview of caller recruitment, call content, and data collection time points.

Figure 1. Brief description of call content, recruitment process, and computer assisted telephone interviews (CATIs) time points.
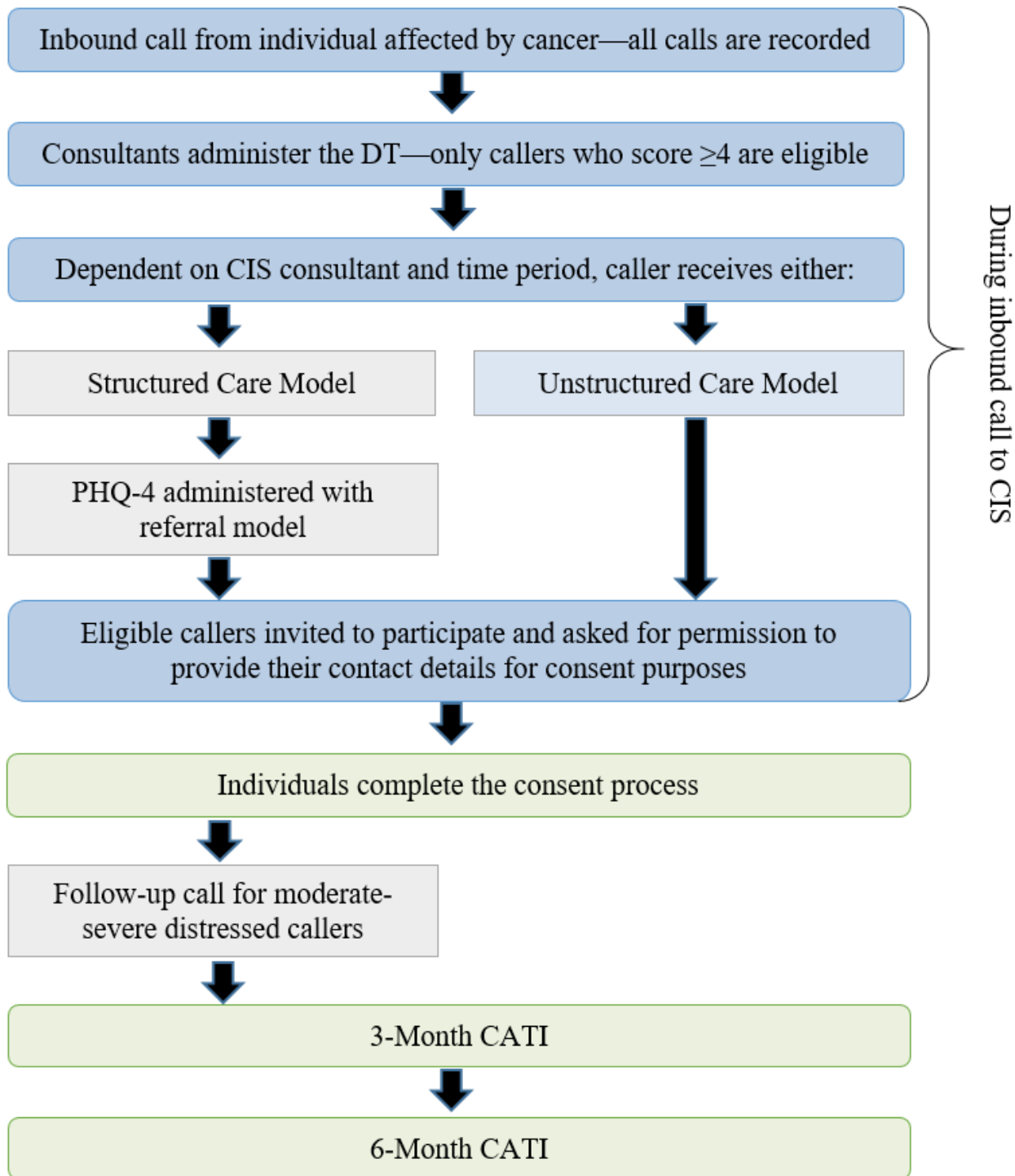
Consultant Training and Randomization

Consultant Qualifications

CIS consultants are qualified oncology or psychosocial professionals such as specialist oncology nurses, psychologists or counsellors, and social workers. To account for the multidisciplinary backgrounds of CIS consultants, each consultant receives training in therapeutic communication skills and routinely participate in clinical supervision and professional development workshops. As part of the training, new consultants

also complete clinical shadowing and receive simulated calls as a quality assurance exercise.

Intervention Training

To participate in this study, consultants participated in 2 3-hour face-to-face group workshops: (1) usual care training focused on administering the DT and recruiting callers and (2) structured care training focused on integrating the PHQ-4 and referral model into calls. Each interactive workshop is followed by a group videoconference to troubleshoot issues and reinforce new behaviors. Booster training is provided at 3- to 6-month intervals depending on the need.

Consultant Randomization

Consultants are randomized to deliver either usual care or structured care during inbound calls with all consultants delivering structured care (ie, PHQ-4 with the referral model) at study completion (Table 1). A randomization table was created by computer software (ie, computerized sequence generation) and using stratified block randomization by state. Owing to the nature of the intervention, consultants are not blinded to the allocation and callers are not informed of consultants' allocation.

Unstructured Care Calls (Usual Care Condition)

All CIS inbound calls begin with an opening such as "How can I help you today?" The conversation is caller-directed, in that the consultant seeks to understand the reason for the call and establish a level of rapport with the caller using responsive listening and expressions of empathy. Although the ensuing care is not formally structured, all consultants use the same client record management (CRM) system which provides a common framework and resources to guide the scope and progress of calls. Inbound calls can last from 10 to 45 minutes.

Under usual care, all callers are screened using the DT. Callers may be offered any of the following internal CIS services: (1) online peer-based support and information; (2) face-to-face support groups where people affected by cancer support each other; (3) one-to-one telephone support from a person who has recovered from a similar experience; (4) telephone support group meetings of 3 to 7 members and qualified facilitators twice a month; (5) referral to social work, legal, financial, or transport assistance; (6) cancer survivor programs such as a personalized diet and exercise programs; (7) information resources; or (8) referral to a psychologist coordinated via CIS. These referrals can be introduced at any point throughout the call.

Callers who speak to unstructured care consultants are not denied any services, and callers may request specific services. The usual care group is not assessed for service suitability in the systematic and structured fashion which will be the case for the intervention group.

Structured Care Calls (Intervention Condition)

A structured care call begins with the same openings as in usual care, is caller-directed, and callers are also screened with the DT. To deliver structured care, consultants administer an additional screening tool and apply a referral model based on screening results (Multimedia Appendix 1). There is no specific time point at which the consultants must introduce additional screening tools and referrals during the call, so the discussion remains caller-directed and conversational. Similar to callers who speak to unstructured care consultants, those who receive structured care are not denied any services and callers may request specific services.

Structured Care Screening Tools

The first screening stage is conducted using the DT. Those who score 4 or more subsequently complete the PHQ-4. The PHQ-4 has been used in cancer samples previously and has been shown to be accurate and reliable [30]. The PHQ-4 is brief and well-suited to the conversational style of CIS consultations. To report severity of distress, the PHQ-4 consists of 2 items regarding depressed mood (PHQ-2) and 2 items from the General Anxiety Disorder [(GAD-2)] tool relating to anxious mood. Using a Likert scale, respondents indicate how often they experienced the symptom in the previous 2 weeks. PHQ-4 scores range from 0 to 12 with higher scores indicating greater distress severity. The PHQ-4 score will be used to determine which services are most appropriate for each caller based on the structured care referral model.

Structured Care Referral Model

To develop a referral model that is evidence-informed and a pragmatic combination of distress screening, tiered care, and stepped care, the team held a 2-day workshop to map internal CIS services to PHQ-4 distress scores. The iterative model development process included group review of existing guidelines; identification of service characteristics to guide mapping decisions (eg, frequency and health professional delivering the intervention); place-card methodology to arrange services by PHQ-4 scores; and application of clinical vignettes with think-aloud methodology to confirm appropriateness of referral decisions. The referral model was further refined following a pilot test with approximately 40 distressed callers. Examples of changes to the model include reformatting to a pyramid shape with 3 tiers; additional detail on timing of an outbound follow-up call; and description of universal care options such as CIS information brochures.

The final referral model includes 16 to 20 CIS services across 3 levels of distress and is state-specific, given different service availability and branding (Multimedia Appendix 1). The referral model also includes an additional follow-up call to repeat screening and further support those individuals with elevated or unchanged distress scores. If accepted, the timing of the call is determined by caller preferences to accommodate pivotal moments in the cancer journey such as treatment commencement.

Intervention Fidelity and Contamination

All calls to the Cancer Council are recorded as a part of standard operating procedures and the research team will review a random sample of recorded phone calls. Using the CIS CRM Systems, the research team will audit the proportion of calls in which the PHQ-4 scores were recorded to determine fidelity. Contamination between the 2 groups, for example, use of the PHQ-4 items, will also be evaluated by auditing usual care calls in addition to structured care calls for fidelity. Contamination

between usual care consultants and structured care consultants is minimized by strategic design of the CRM system so that the PHQ-4 is not easily viewable by usual care consultants and separated training sessions; and, as a part of training, consultants were informed about the trial design and the need to reduce contamination.

Study Outcomes

Study outcomes are assessed through 6-month computer assisted telephone interviews (CATIs) with participants; the CIS CRM Systems; and review of audio recordings. The data collection time periods were designed to increase patient recall of referrals in 2 shorter 3-month periods. CATIs are an appropriate method of data collection from participants recruited through a telephone-based support service.

Referral uptake is the primary outcome and is measured as the proportion of participants who report being provided with a referral and report at 6-month follow-up that an action has been taken to progress the referral (ie, an appointment or telephone interaction). Referral uptake was selected as the primary outcome as it can reflect the appropriateness of the type of referral offered to callers under the structured care model. The hypothesized increase of 20% in referral uptake was selected based on internal CIS data in the absence of comparable studies in telephone-based supportive care service uptake. The study-specific questions are tailored to the CIS services; questions were pilot-tested with a consumer advisory panel and reviewed after the first 20 participants.

Distress is measured by the GHQ-28 [31] that is a widely used self-report measure of general psychological distress. The measure uses 28 items to assess perception of health in terms of ability to play a useful part in life; make decisions; overcome difficulties; enjoy normal activities; face problems; and feel confident, worthwhile, and happy [31]. The GHQ-28 has excellent internal consistency, diagnostic accuracy, and test-retest reliability [32,33]. The measure has been used in the Australian community [34] and with patients with cancer [35].

Moderators

Sociodemographic and disease-related characteristics include cancer type, stage of disease, age, gender, postcode, marital status, education, health care card recipient, private health insurance, household income, previous psychological treatment, and/or morbidity and other assistance received since study enrolment. Caller type (patient or caregiver), DT score at inbound CIS call, and reason for contacting the CIS (eg, information, emotional, or practical support) will also be accounted for in study analyses. Subgroup analyses will be conducted for caregivers. Basic information regarding the consultants, such as the number of years at the CIS, will also be incorporated into analyses.

The Health Education Impact Questionnaire (HeiQ) is a 42-item tool for assessing the efficacy and impact of health education and self-management programs for people with chronic diseases [36]. The HeiQ has demonstrated reliability and validity, including with oncology samples, and has been used previously by Cancer Councils to evaluate programs [37,38].

Acceptability

The consultants' experience in using the structured care model, and the perceived value of this approach, will be explored in qualitative interviews at conclusion of the study. The semistructured interviews will include questions about the value of the additional questions and referral model at eliciting and managing emotional distress and the impact of these additions in building rapport and maintaining a caller-directed approach. The interview will be recorded, partially transcribed, and analyzed according to content analysis facilitated through NVivo qualitative data analysis software (QSR International Pty Ltd. Version 10, 2014). The Consolidated Criteria for Reporting Qualitative Research will be used [39].

Costs

The length and number of calls made or received by participants is automatically recorded. Standard hourly rates for consultants will be used to calculate the cost of service delivery. Service provider costs owing to the uptake of referrals in both groups are tracked through the Cancer Council CRM, and standard hourly costs will be applied to provide a full assessment of the cost implications for each of the care models. Standard hourly costs will be derived directly from CIS CRM data.

Participant outcome data and service provider cost data will be submitted to a series of discussion sessions involving consumers and relevant service leads at conclusion of the study. These sessions will explore whether the identified consumer benefits are perceived to be commensurate with the additional service provider costs incurred. The discussion sessions will be guided by the Nominal Group Technique [40]: (1) Service leads will receive a prediscussion report with information such as cost per caller screened under usual and structured care, odds of referral uptake, and average change in distress scores; (2) The roundtable discussion will be led by a group facilitator who will ask for individual feedback recorded in a *round-robin* format; (3) The group will then collectively discuss the feedback until consensus is reached on the cost and benefits of intervention. This discussion will be audio-recorded; and 4) Participants will have the opportunity to provide further comment directly to the research team.

Statistical Methods and Sample Size

Descriptive statistics will assess referral uptake rates. Differences between GHQ-28 scores for the structured and usual care groups will be tested using a generalized linear mixed model with fixed effects for the intervention and each time period after baseline. To account for the fact that outcomes are measured at the participant level while randomization is at the consultant level, a normally distributed random intercept for consultants will be included in the model. The parameter of interest from these models will be the estimated coefficient for the intervention term. On the basis of 8 consultants participating, the trial aims to recruit a total of 1512 distressed callers across the 3 steps. The sample size provides the study with approximately 80% power to detect a difference of 0.3 SD in the mean score of the GHQ-28 at an alpha level of .05 and assuming an intracluster correlation of .04. An effect size of .3 was selected based on previous trials of psychoeducation and

XSL•FO

RenderX

telephone-administered interventions with patients with cancer [41].

Ethics

The study has been approved by the University of Newcastle Human Research Ethics Committee (NHMRC Committee Code: EC00144; Reference No. H-2016-0180); the Cancer Council of NSW (NHMRC Committee Code: EC00345; Reference No. 304); and the Cancer Council Victoria (NHMRC Committee Code: EC00203; Reference No. 1605).

Results

Progress

As of April 2019, 1835 eligible CIS callers have been invited to participate in the study by CIS consultants. A total of 1114 (60.70%, 1114/1835) individuals consented to participate; 372 (20.27%, 372/1835) declined to participate; and 180 (9.81%, 180/1835) did not respond to the invitation to participate after postal and telephone reminders. The consent status of the remaining 79 individuals is not yet known, and the individuals are currently receiving follow-up reminders through phone and email prompts.

Of the 1114 consenting participants, 692 have now completed the 6-month CATI. From the start of the study, 182 enrolled participants did not complete their 6-month CATI—this represents a lost-to-follow-up of 20.8% (182/874). Just over half of the current sample (56.01%, 624/1114) are patients. The remaining 44.00% (490/1114) of participants are individuals supporting someone with a cancer diagnosis, in remission or bereaved.

Timeline

The trial entered the final intervention stage in April 2019, with all participating CIS consultants now delivering structured care. Study recruitment will continue for 6 months after this transition, with a further 6 months required to complete follow-up assessments. As such, trial outcome data are anticipated to be available in early 2020.

Discussion

Evidence-Informed Distress Screening by Telephone-Based Services Can Fill an Important Gap in Supportive Care

Distress among patients with cancer and caregivers is recognized as an important and challenging issue and is yet to be managed in a widespread, consistent, and effective fashion [11]. The Cancer Council CIS services hold a unique coordination role within community-based cancer care by facilitating access to additional psychosocial support programs outside of hospital-based settings. The CIS services also have tremendous potential to support caregivers and survivors who will experience clinically significant distress [2,3]. Furthermore, the CIS can provide a safety net by identifying the many distressed patients with cancer who *fall-through the cracks* in Australian cancer services [42,43]. For example, a previous Australian study suggested that a third of cancer services do not routinely screen outpatients for distress [42].

Study Contribution to the Literature on Distress Screening and Management

The intention for any community-based telephone counselling service, such as the CIS services, to implement distress screening and various forms of structured care requires thorough evaluation of the benefits to callers. These benefits must be considered alongside the additional staff time and potential changes to the way staff interact with callers. For example, it is unknown if additional screening may affect the type and intensity of support offered to distressed callers, particularly as emotional well-being may not have been the motivating reason for contacting the service. A key strength of this study is the diverse sample of patients and caregivers recruited to mimic the large and heterogeneous community supported by the Cancer Council CIS services and other international telephone counselling services [16-18]. As evidence suggests that caregivers will experience distress levels similar to those of patients [3], the study will provide an invaluable opportunity to specifically report on the experience of distressed caregivers and their use of supportive care services.

As a critical factor for the translation of research into practice, this study has a strong focus on the resource implications of implementing a structured care model. The final phase of the proposed study involves facilitated discussion with service leaders and managers to examine the psychosocial outcome data alongside the cost data. This process will assist the end users in assessing the resource implications of implementing structured care, which will be important for the translation and sustainability of a structured care approach. The assessor comments have been provided in Multimedia Appendix 2.

This trial is the first to rigorously test the outcomes of a community-based (rather than clinical) structured approach to distress management. The model is evidence-informed, practice-ready, and trialed in a real-world setting. The outcomes of this trial will advance the understanding of distress management internationally. The proposed trial is also one of the first to deliver a harmonized multi-state intervention across state-based CIS borders. As the CIS service reported approximately 46,000 calls nationally in 2017, it is evident that an effective and consistent distress management model has tremendous potential to improve the psychosocial care for a large number of patients with cancer and caregivers.

of the trial. CP and DB will oversee data analysis, whereas all authors will be involved in interpretation of results. All authors have contributed to, read, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Example of referral model.

[PNG File, 829KB - resprot_v8i5e12473_app1.png]

Multimedia Appendix 2

Assessor comments as part of National Health and Medical Research Council peer-review process.

[PDF File (Adobe PDF File), 150KB - resprot_v8i5e12473_app2.pdf]

References

1. National Comprehensive Cancer Network. 2014. NCCN clinical practice guidelines in oncology: distress management (V2) URL: https://www.nccn.org/professionals/physician_gls/pdf/distress.pdf [accessed 2019-05-06] [WebCite Cache ID 789uj8f56]

2. Reilly C, Bruner D, Mitchell S, Minasian L, Basch E, Dueck A, et al. A literature synthesis of symptom prevalence and severity in persons receiving active cancer treatment. Support Care Cancer 2013 Jun;21(6):1525-1550 [FREE Full text] [doi: 10.1007/s00520-012-1688-0] [Medline: 23314601]

3. Lambert S, Girgis A, Lecathelinais C, Stacey F. Walking a mile in their shoes: anxiety and depression among partners and caregivers of cancer survivors at 6 and 12 months post-diagnosis. Support Care Cancer 2013 Jan;21(1):75-85. [doi: 10.1007/s00520-012-1495-7] [Medline: 22661096]

4. Snyder C, Garrett-Mayer E, Brahmer J, Carducci M, Pili R, Stearns V, et al. Symptoms, supportive care needs, and function in cancer patients: how are they related? Qual Life Res 2008 Jun;17(5):665-677 [FREE Full text] [doi: 10.1007/s11136-008-9331-8] [Medline: 18493865]

5. Russ T, Stamatakis E, Hamer M, Starr J, Kivimäki M, Batty G. Association between psychological distress and mortality: individual participant pooled analysis of 10 prospective cohort studies. Br Med J 2012 Jul 31;345:e4933 [FREE Full text] [doi: 10.1136/bmj.e4933] [Medline: 22849956]

6. Cancer Australia. Australia: Australian Government; 2014. Clinical guidance for responding to suffering in adults with cancer URL: https://guidelines.canceraustralia.gov.au/guidelines/guideline_22.pdf [accessed 2019-05-01] [WebCite Cache ID 785HodXEJ]

7. Butow P, Price M, Shaw J, Turner J, Clayton J, Grimison P, et al. Clinical pathway for the screening, assessment and management of anxiety and depression in adult cancer patients: Australian guidelines. Psychooncology 2015 Sep;24(9):987-1001. [doi: 10.1002/pon.3920] [Medline: 26268799]

8. Carlson L, Groff S, Maciejewski O, Bultz B. Screening for distress in lung and breast cancer outpatients: a randomized controlled trial. J Clin Oncol 2010 Nov 20;28(33):4884-4891. [doi: 10.1200/JCO.2009.27.3698] [Medline: 20940193]

9. Boyes A, Newell S, Girgis A, McElduff P, Sanson-Fisher R. Does routine assessment and real-time feedback improve cancer patients' psychosocial well-being? Eur J Cancer Care (Engl) 2006 May;15(2):163-171. [doi: 10.1111/j.1365-2354.2005.00633.x] [Medline: 16643264]

10. Mitchell A. Screening for cancer-related distress: when is implementation successful and when is it unsuccessful? Acta Oncol 2013 Feb;52(2):216-224. [doi: 10.3109/0284186X.2012.745949] [Medline: 23320770]

11. Braeken A, Kempen G, Eekers D, Houben R, Van Gils F, Ambergen T, et al. Psychosocial screening effects on health-related outcomes in patients receiving radiotherapy. A cluster randomised controlled trial. Psychooncology 2013 Dec;22(12):2736-2746. [doi: 10.1002/pon.3340] [Medline: 23824561]

12. Dekker J, Beekman A, Boenink A, Bomhof-Roordink H, Braamse A, Collette E, et al. Comment on 'psychological distress in patients with cancer: is screening the effective solution?'. Br J Cancer 2013 Jun 25;108(12):2628-2630 [FREE Full text] [doi: 10.1038/bjc.2013.286] [Medline: 23756863]

13. Fielding R, Lam W. Unresolved problems with distress screening. Br J Cancer 2013 May 14;108(9):1922-1923 [FREE Full text] [doi: 10.1038/bjc.2013.137] [Medline: 23549039]

14. Dekker J, Braamse A, Schuurhuizen C, Beekman A, van Linde M, Sprangers M, et al. Distress in patients with cancer - on the need to distinguish between adaptive and maladaptive emotional responses. Acta Oncol 2017 Dec;56(7):1026-1029. [doi: 10.1080/0284186X.2017.1280848] [Medline: 28145789]

15. Lazenby M. The international endorsement of US distress screening and psychosocial guidelines in oncology: a model for dissemination. J Natl Compr Canc Netw 2014 Feb;12(2):221-227. [Medline: 24586084]

16. Morra M, Thomsen C, Vezina A, Akkerman D, Bright M, Dickens C, et al. The international cancer information service: a worldwide resource. J Cancer Educ 2007;22(1 Suppl):S61-S69. [doi: 10.1080/08858190701348448] [Medline: 17572003]

17. Marcus A, Garrett K, Kulchak-Rahm A, Barnes D, Dortch W, Juno S. Telephone counseling in psychosocial oncology: a report from the Cancer Information and Counseling Line. Patient Educ Couns 2002 Apr;46(4):267-275. [Medline: 11932126]

18. Jefford M, Kirke B, Grogan S, Yeoman G, Boyes A. Australia's Cancer Helpline--an audit of utility and caller profile. Aust Fam Physician 2005 May;34(5):393-394 [FREE Full text] [Medline: 15887947]

19. Cancer Council Australia. 2017. Annual Report 2016/2017 URL: https://www.cancercouncil.com.au/wp-content/uploads/2017/11/17064_CAN062_AR16-17_LOW-RES.pdf [accessed 2019-05-01] [WebCite Cache ID 78546eUaj]

20. Cancer Council. About Cancer Council NSW URL: https://www.cancercouncil.com.au/about-us/

21. Cancer Council Victoria. About us URL: https://www.cancervic.org.au/about [accessed 2019-05-01] [WebCite Cache ID 7854BOpYf]

22. Hawkes A, Hughes K, Hutchison S, Chambers S. Feasibility of brief psychological distress screening by a community-based telephone helpline for cancer patients and carers. BMC Cancer 2010 Jan 12;10:14 [FREE Full text] [doi: 10.1186/1471-2407-10-14] [Medline: 20067645]

23. Hughes K, Sargeant H, Hawkes A. Acceptability of the distress thermometer and problem list to community-based telephone cancer helpline operators, and to cancer patients and carers. BMC Cancer 2011 Jan 31;11:46 [FREE Full text] [doi: 10.1186/1471-2407-11-46] [Medline: 21281476]

24. Funk R, Cisneros C, Williams R, Kendall J, Hamann H. What happens after distress screening? Patterns of supportive care service utilization among oncology patients identified through a systematic screening protocol. Support Care Cancer 2016 Dec;24(7):2861-2868. [doi: 10.1007/s00520-016-3099-0] [Medline: 26838023]

25. Lambert S, Kelly B, Boyes A, Cameron A, Adams C, Proietto A, et al. Insights into preferences for psycho-oncology services among women with gynecologic cancer following distress screening. J Natl Compr Canc Netw 2014 Jun;12(6):899-906. [Medline: 24925200]

26. Faller H, Schuler M, Richard M, Heckl U, Weis J, Küffner R. Effects of psycho-oncologic interventions on emotional distress and quality of life in adult patients with cancer: systematic review and meta-analysis. J Clin Oncol 2013 Feb 20;31(6):782-793. [doi: 10.1200/JCO.2011.40.8922] [Medline: 23319686]

27. Hemming K, Taljaard M, McKenzie J, Hooper R, Copas A, Thompson J, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. Br Med J 2018 Nov 09;363:k1614 [FREE Full text] [doi: 10.1136/bmj.k1614] [Medline: 30413417]

28. Chan A, Tetzlaff J, Altman D, Laupacis A, Gøtzsche P, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013 Feb 05;158(3):200-207 [FREE Full text] [doi: 10.7326/0003-4819-158-3-201302050-00583] [Medline: 23295957]

29. Ma X, Zhang J, Zhong W, Shu C, Wang F, Wen J, et al. The diagnostic role of a short screening tool--the distress thermometer: a meta-analysis. Support Care Cancer 2014 Jul;22(7):1741-1755. [doi: 10.1007/s00520-014-2143-1] [Medline: 24510195]

30. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med 2007 Nov;22(11):1596-1602 [FREE Full text] [doi: 10.1007/s11606-007-0333-y] [Medline: 17874169]

31. Goldberg D, Hillier V. A scaled version of the General Health Questionnaire. Psychol Med 1979 Feb;9(1):139-145. [Medline: 424481]

32. Vodermaier A, Linden W, Siu C. Screening for emotional distress in cancer patients: a systematic review of assessment instruments. J Natl Cancer Inst 2009 Nov 04;101(21):1464-1488 [FREE Full text] [doi: 10.1093/jnci/djp336] [Medline: 19826136]

33. Meader N, Mitchell A, Chew-Graham C, Goldberg D, Rizzo M, Bird V, et al. Case identification of depression in patients with chronic physical health problems: a diagnostic accuracy meta-analysis of 113 studies. Br J Gen Pract 2011 Dec;61(593):e808-e820 [FREE Full text] [doi: 10.3399/bjgp11X613151] [Medline: 22137418]

34. Taylor A, Wilson D, Dal Grande E, Ben-Tovim D, Elzinga R, Goldney R, et al. Mental health status of the South Australian population. Aust N Z J Public Health 2000 Feb;24(1):29-34. [Medline: 10777975]

35. Petersen R, Quinlivan J. Preventing anxiety and depression in gynaecological cancer: a randomised controlled trial. BJOG 2002 Apr;109(4):386-394 [FREE Full text] [Medline: 12013159]

36. Osborne R, Elsworth G, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]

37. Osborne R, Batterham R, Livingston J. The evaluation of chronic disease self-management support across settings: the international experience of the health education impact questionnaire quality monitoring system. Nurs Clin North Am 2011 Sep;46(3):255-270. [doi: 10.1016/j.cnur.2011.05.010] [Medline: 21791261]

38. Maunsell E, Lauzier S, Brunet J, Pelletier S, Osborne R, Campbell H. Health-related empowerment in cancer: validity of scales from the Health Education Impact Questionnaire. Cancer 2014 Oct 15;120(20):3228-3236 [FREE Full text] [doi: 10.1002/cncr.28847] [Medline: 24988944]

39. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

40. Gallagher M, Hares T, Spencer J, Bradshaw C, Webb I. The nominal group technique: a research tool for general practice? Fam Pract 1993 Mar;10(1):76-81. [Medline: 8477899]

41. Galway K, Black A, Cantwell M, Cardwell C, Mills M, Donnelly M. Psychosocial interventions to improve quality of life and emotional wellbeing for recently diagnosed cancer patients. Cochrane Database Syst Rev 2012 Nov 14;11:CD007064 [FREE Full text] [doi: 10.1002/14651858.CD007064.pub2] [Medline: 23152241]

42. Fradgley E, Byrnes E, Mccarter K, Britton B, Rankin N, Carter G. A report card on Australian cancer services' distress screening and management practices: A pass mark or time to go back to school? Psycho-oncology 2018:07030-05774.

43. Zucca A, Sanson-Fisher R, Waller A, Carey M, Fradgley E, Regan T. Medical oncology patients: are they offered help and does it provide relief? J Pain Symptom Manag 2015;50(4):436-444. [doi: 10.1016/j.jpainsymman.2015.04.018]

Abbreviations

CATI: computer assisted telephone interview
CIS: Cancer Information and Support
CRM: client record management
DT: Distress Thermometer
GHQ: General Health Questionnaire
HeiQ: Health Education Impact Questionnaire
NHMRC: Newcastle Human Research Ethics Committee
NSW: New South Wales
PHQ: Patient Health Questionnaire

Protocol

Evaluating the Feasibility and Impact of a Yoga Intervention on Cognition, Physical Function, Physical Activity, and Affective Outcomes in People Living With HIV: Protocol for a Randomized Pilot Trial

Adria Quigley[1], BSc, MSc; Kelly K O'Brien[2,3,4], BSc, BScPT, PhD; Marie-Josée Brouillette[5], MD, FRCPC; Marilyn MacKay-Lyons[1,6,7], BSc, MSc, PhD

[1]Department of Physiotherapy, Dalhousie University, Halifax, NS, Canada

[2]Department of Physical Therapy, University of Toronto, Toronto, ON, Canada

[3]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

[4]Rehabilitation Sciences Institute, University of Toronto, Toronto, ON, Canada

[5]Department of Psychiatry, McGill University, Montreal, QC, Canada

[6]Department of Medicine, Dalhousie University, Halifax, NS, Canada

[7]Nova Scotia Health Authority, Halifax, NS, Canada

Corresponding Author:
Adria Quigley, BSc, MSc
Department of Physiotherapy
Dalhousie University
Forrest Building, 5869 University Avenue
PO Box 15000
Halifax, NS, B3H 4R2
Canada
Phone: 1 782 234 2121
Fax: 1 (902) 494 1941
Email: adriaquigley@gmail.com

Abstract

Background: Despite lower mortality rates due to combination antiretroviral therapy, people living with HIV (PLWH) are grappling with increasingly complex health issues, including cognitive impairments in areas such as memory, attention, processing speed, and motor function. Yoga has been shown to be an effective form of exercise and mindfulness-based stress reduction for many clinical populations. However, no randomized trials have evaluated the impact of yoga on cognitive and physical function among PLWH.

Objective: The aim of this pilot randomized trial was to determine the feasibility of a yoga intervention to lay the groundwork for a full-scale, multisite, community-based trial for PLWH. Specific objectives are to (1) assess the feasibility of study protocol and procedures, (2) compare cognition in the yoga group with the usual care control group after 12 weeks of the intervention in PLWH, and (3) compare the effects of the 12-week yoga intervention versus control on balance, walking speed, physical activity, mental health, medication adherence, and quality of life among PLWH.

Methods: We propose a pilot randomized trial with 2 parallel groups (yoga versus control). We will recruit 25 PLWH (>35 years) from community and health organizations in Halifax, Canada. After baseline assessment with blinded assessors, participants will be randomly assigned to the yoga or control group, using a random computer generator. Participants in the yoga group will engage in supervised 60-min group-based yoga sessions 3 times a week for 12 weeks at a yoga studio. Participants in the control group will maintain their current physical activity levels throughout the study.

Results: As per the Consolidated Standards of Reporting Trials extension for pilot studies, means of all outcomes, mean change, and 95% CIs will be calculated for each group separately. Two-tailed independent *t* tests and Fisher exact tests will be used to compare groups at baseline. We will analyze quantitative postintervention questionnaire responses using Chi-square tests, and open-ended responses will be analyzed thematically. Intention-to-treat and per-protocol analyses will be used to analyze secondary variables. Changes in outcome variables will be examined between groups and within groups. Effect sizes will be reported for

XSL•FO
RenderX

each outcome. A priori adherence and satisfaction criteria will be met if participants attend >70% of the yoga sessions and if >70% of the participants are satisfied with the intervention as determined by a postparticipation questionnaire. Study enrollment began in January 2018, with results expected for October 2019.

Conclusions: This pilot randomized trial will be the first to investigate the feasibility and effect of a yoga intervention on cognitive and physical outcomes among PLWH. This work will inform the feasibility of further investigations in terms of capacity building, participant recruitment and retention, and assessment and intervention protocols.

Trial Registration: ClinicalTrials.gov NCT03071562; https://clinicaltrials.gov/ct2/show/NCT03071562 (Archived by WebCite at http://www.webcitation.org/785sfhWkw)

International Registered Report Identifier (IRRID): DERR1-10.2196/13818

KEYWORDS

HIV; AIDS; yoga; cognition

Introduction

Cognitive Impairment in People Living With HIV

Despite lower mortality rates due to combination antiretroviral therapy (cART), people living with HIV (PLWH) are grappling with increasingly complex health issues [1], including cognitive impairments in areas such as memory, attention, processing speed, and motor function [2]. Even with the widespread use of cART, 30% to 60% of PLWH experience cognitive impairment [2,3]. Given that the number of people with HIV-associated cognitive impairment is expected to increase 5- to 10-fold by the year 2030 [4], and the incidence of HIV infection is increasing among older adults [5], this issue has become a public health concern [6]. Aging and HIV appear to have combined deleterious effects on both brain structure and function, and some investigators have hypothesized that these effects could be synergistic [7,8]. As such, the combined effect of age and cognitive impairment in HIV has become a concern over the past decade, especially as PLWH now have a life expectancy that rivals that of their HIV-negative counterparts [9]. Proposed mechanisms for cognitive dysfunction include direct attacks of the virus on brain tissue and indirect processes such as local or systemic inflammation [10]. Glial cells, possible reservoirs for the virus, release proinflammatory cytokines and toxins associated with cognitive disorders and neuron degeneration [11]. Protein gp120 damages neurons by causing calcium overload and reducing brain-derived neurotrophic factor, the central growth factor involved in neurogenesis [12].

HIV-associated cognitive impairment has a profound impact on activities of daily living [13], social function [14], quality of life [15], employment [16], and adherence to pharmacological [17] and nonpharmacological treatment [18]. Despite the fact that ~95% adherence to cART is required for adequate viral suppression, 66% of participants in a HIV clinical trial *simply forgot* to take their medications [19]. Pharmacological adherence is a major priority, given that cART is the mainstay of proper HIV management. A study of 267 adults with HIV revealed that those with cognitive impairment performed worse on functional laboratory measures of shopping, cooking, finances, medication management, and work-related skills than those with normal cognition [20]. Furthermore, the authors discovered that poor executive function, learning, attention, working

memory, and verbal abilities strongly predicted functional performance [20]. Authors of another study revealed that symptomatic cognitive impairment was associated with significantly worse scores in 8 domains of the Medical Outcomes Survey for HIV (MOS-HIV) [21]. PLWH with cognitive impairment are less likely to be employed [22], have a difficult time returning to work after disability [23], and have difficulties adapting to the demands of work [20].

Gait and Balance Impairments Among People Living With HIV

Although the cognitive aspects of HIV-associated neurocognitive disorder such as memory, attention, and processing speed have been studied in great detail, the motor aspects have not received much attention. There is evidence of a shared pathology between cognitive and motor functions; a large study of 1549 PLWH revealed a significant relationship between slowed gait and worsening cognitive function [24]. Balance and gait impairments are common among PLWH [25], and they are associated with frailty, higher rates of falls, and increased mortality [26]. Decreased gait speed is linked to higher fall risk, even in those taking cART with undetectable viral loads [26]. A recent systematic review and meta-analysis of 16 cross-sectional studies and 1 prospective cohort study conducted by Berner and colleagues (2017) evaluated the available literature on gait and balance dysfunction in PLWH [27]; a total of 3 [28-30] of 8 studies [25,26,28-33] that examined gait speed reported slowing of fast gait speeds among PLWH compared with controls.

Balance performance tests also reveal balance impairments among PLWH. Using the Single Leg Stance Time Test, Bauer and colleagues (2011) [25] revealed a significant decrease in nonpreferred single leg stance time among obese PLWH compared with seronegative controls in their sample of 86 seropositive and 121 seronegative individuals. Sullivan and colleagues (2011) [34] had similar findings in their sample of 40 female and male PLWH, but they found no differences between groups in tandem stance time. Using the Single Leg Stance Time Test with eyes closed in their sample of 308 PLWH, Tanon and colleagues (2017) [35] determined that 87% of participants demonstrated balance impairments. Performance on the Heel-To-Toe Walk Test with eyes closed [34], the Limits of Stability Test [25,31], and the 360-Degree-Turn Test (among

PLWH with obesity only) [25] may also be impaired. Notably, PLWH appear to perform well on the Berg Balance Scale [26,27,36], which indicates that more challenging dynamic balance assessments are required to identify impairments in this population.

Exercise and Cognitive Function in People Living With HIV

Quigley and colleagues (2018) recently published a scoping review to map the available evidence regarding physical activity and cognitive outcomes (both objective and self-reported) among PLWH [37]. The scoping review included 16 studies: 5 randomized controlled trials (RCTs) [38-42], 3 pre-post single group observational studies [43-45], and 8 cross-sectional studies [46-53], with a total of 1701 PLWH [37]. The noninterventional research indicated a strong association between physical activity levels and cognitive performance as measured by a cognitive battery in PLWH; all 8 cross-sectional studies demonstrated positive associations [46-53]. However, only 2 of the 8 interventional studies—an RCT [41] of aerobic and resistance exercise and a single cohort study involving Tai Chi [43]—revealed positive outcomes regarding cognition in PLWH. McDermott and colleagues [42] conducted the only RCT to directly examine the effect of exercise on an objective measure of cognition in PLWH. Their 16-week aerobic exercise intervention, 3 times per week at 40% to 75% of heart rate reserve neither had an effect on Montreal Cognitive Assessment scores nor had an effect on Trails A and B scores [42]. However, the sample size comprised 11 participants, and the Montreal Cognitive Assessment may not be sensitive to cognitive impairment in PLWH [54]. Clearly, confirmatory evidence of the effect of exercise on cognition in this population is lacking.

The Effect of Yoga on Cognitive and Physical Function

Yoga has emerged as an effective form of exercise and mindfulness-based stress reduction across many clinical populations [55]. It is an ancient practice combining postures, mindfulness, spirituality, and breath control to enhance flexibility, strength, and balance, and it is increasingly being recognized as a mainstream intervention to promote a more preventative and holistic health care approach [56,57]. Findings of a meta-analysis of 15 RCTs suggest that yoga interventions lasting 1 to 6 months are associated with enhanced overall cognitive function (Hedges g=0.33), attention and processing speed (Hedges g=0.299), executive function (Hedges g=0.27), and memory (Hedges g=0.18) in people with and without chronic diseases [58]. In fact, it appears that acute bouts of yoga may be superior to aerobic exercise for improving inhibition and working memory, as determined by a repeated-measures study of 30 healthy younger women [59]. There are numerous mechanisms thought to underlie cognitive improvements with yoga interventions. It is possible that yoga may contribute to dominance of the parasympathetic nervous system [60,61] while downregulating the sympathetic nervous system and the hypothalamic-pituitary-adrenal axis [61]. A systematic review of 25 RCTs conducted with healthy and chronic disease populations revealed that those who participated in yoga improved their cortisol levels, heart rate, and blood pressure relative to controls [62]. There is also evidence that yoga and

other types of mind-body exercise (including Tai Chi) are associated with improved mood; a meta-analysis of 40 interventional studies revealed that Tai Chi has positive effects on both anxiety and depression [63]. Improvements in the stress response with mind-body exercise may contribute to improved cognitive performance [64]; an RCT of 118 older adults revealed that yoga participants had an attenuated cortisol response and improved executive function relative to the control group following an 8-week yoga intervention [64]. Of note, self-reported mood stress and cortisol levels predicted executive function performance [64]. Other potential mechanisms associated with yoga interventions include the learning of novel tasks, which is associated with changes in brain structure and function [58], sustained attention [65], activation of the default mode network (including learning and consolidation functions) [66], and improved meta-cognition (one's conscious awareness of his or her cognitive processes), which is closely related to executive function [67].

Yoga is also an effective treatment for impaired balance in people with [68-71] and without physical impairments [72,73] because of its positive effects on strength [74], mobility [69], balance self-efficacy [70,71], and visuospatial memory [75]. A 2016 systematic review and meta-analysis of 6 RCTs confirmed that healthy older adults and individuals with various health conditions, such as stroke, Parkinson's disease, and knee osteoarthritis, reap yoga-induced benefits to postural stability and mobility [76]. The investigators suggested that health care professionals should recommend yoga to older adults as a safe and effective intervention for balance and mobility limitations [76]. There is considerably less research evaluating the effect of yoga on balance, quality of life, and depression in PLWH. In their case-series study of 3 PLWH, Kietrys and colleagues (2018) observed improvements in several gait parameters (including double-limb support time, step length, stride length, stride velocity, and walking velocity) and balance (as measured by the Multidirectional Reach Test) in 2 of the 3 participants following a 4-week yoga intervention [77]. There is some RCT evidence for the benefits of yoga on quality of life [78] and depression [79] in PLWH; however, the former study did not involve yoga postures, and the latter intervention was only a month in total duration. To date, no RCTs have evaluated the impact of yoga on cognitive and physical performance among PLWH.

Purpose and Objectives

The purpose of this pilot RCT is to determine the feasibility of a yoga intervention to lay the groundwork for a full-scale, multisite, and community-based trial with PLWH. Specific objectives are to (1) assess the feasibility of the study protocol and procedures, (2) compare cognitive function in PLWH in a yoga intervention group with a usual care control group among PLWH after 12 weeks of the intervention, and (3) compare the effects of the 12-week yoga intervention versus control on balance, walking speed, physical activity, mental health, medication adherence, and quality of life in PLWH.

Methods

Design

We propose a pilot randomized trial with 2 parallel groups, comparing the yoga group with a usual care control group using quantitative methods of data collection. Figure 1 outlines the sequencing of the study protocol. The conceptual framework for pilot and feasibility studies created by Eldridge and

colleagues [80] and the Consolidated Standards of Reporting Trials (CONSORT) 26-item checklist for randomized pilot and feasibility studies will be employed to ensure methods are properly defined and reported [81]. The study is guided by a community advisory committee comprising 7 members of the HIV community and 3 representatives from local HIV organizations. Our research team held consultations with the community advisory committee to assist with study design and recruitment strategy.

Figure 1. Consolidated Standards of Reporting Trials flow diagram.
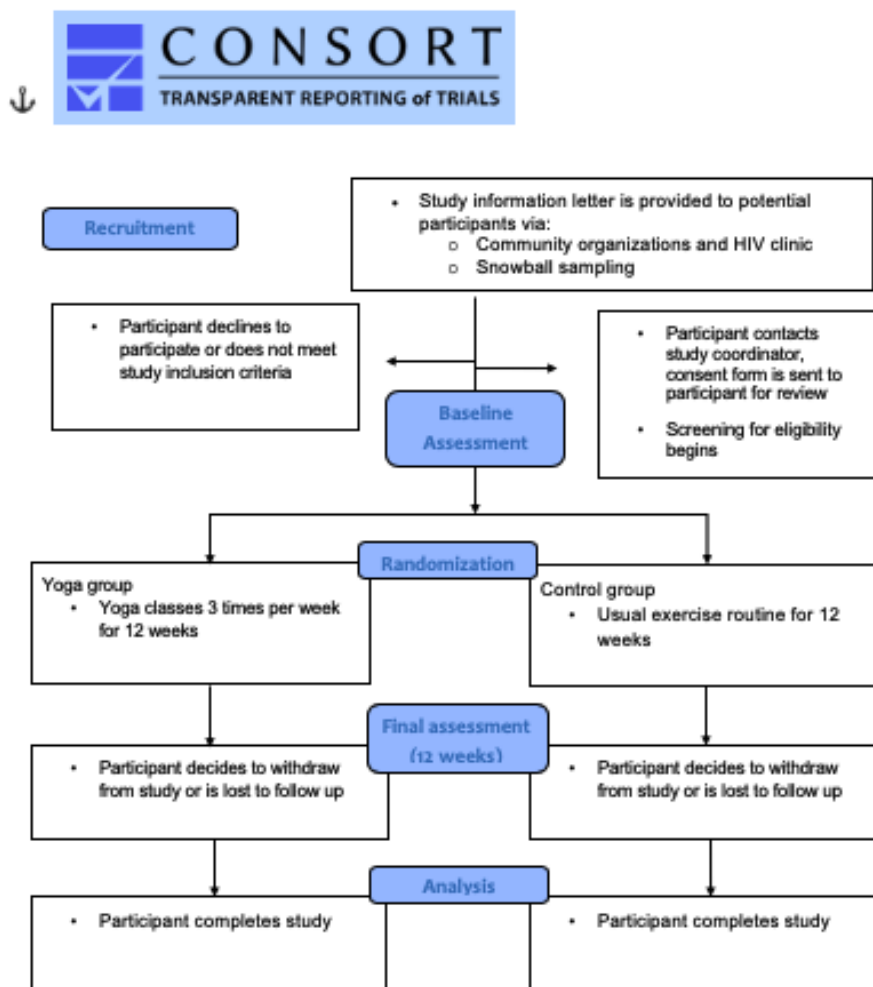
Participants

We will identify cognitive concerns on the Communicating Cognitive Concerns Questionnaire (C3Q) with a cut-off of 35 points or less [82]; in addition, we will include a maximum total of 25 PLWH who are aged 35 years or older of any gender, are English speaking, live within 50 km of the study site, are able to provide informed consent, and are deemed medically stable as assessed by the Physical Activity Readiness Questionnaire Plus [83]. Study exclusion will include regular participation in a yoga program during the 6 months before study commencement.

Recruitment

Recruitment will occur via newsletters and posters at community organizations and health centers in Halifax, Nova Scotia.

Furthermore, staff at the local HIV Clinic have agreed to approach eligible individuals and provide them with a study information brochure. To obtain a sample that is diverse in terms of ethnicity, gender, and severity of HIV disease, we will also employ snowball sampling techniques, whereby potential participants will be asked to identify other potential participants. All interested individuals will contact the study coordinator. The coordinator will explain the general purpose and procedures of the study, risks and potential benefits, time commitment, and responsibilities of the participants. Each potential participant will be informed that health care services will not be affected by study participation or withdrawal. A copy of the consent form will be provided and reviewed, and all the questions will be answered to the potential participant's satisfaction. Potential participants who remain interested in enrolling in the study will

be asked to sign the consent form approved by the local Research Ethics Board (REB).

Randomization

After baseline assessment, an individual not directly involved in the study will randomly assign participants in a 1:1 ratio to the yoga or control group using a random computer generator. Group assignment of each participant will be concealed in individual opaque envelopes that will remain sealed until after completion of the baseline assessment. The number of participants screened and randomized to each group will be recorded, as per the CONSORT extension for randomized pilot trials [81].

Ethical Considerations

The study protocol was approved by the REB (protocol reference #1022158). The procedures will be followed in accordance with institutional ethical standards and the Helsinki Declaration. The trial was registered on ClinicalTrials.gov. Proposed amendments to the protocol will be submitted for review to the REB. For ethical reasons, we cannot ask participants to avoid making medication changes; any changes participants make to their medications will be documented. Unanticipated or adverse events will be reported immediately to the REB. Participant confidentiality and autonomy will be maintained throughout

the study, and data will be anonymized and secured. Study data will be stored in a locked office at Dalhousie University. Electronic data will be stored in encrypted form and will exclusively be accessed by the research team. Restricting access to data on-site until the data have been appropriately coded and deidentified will mitigate the risk of residual disclosure. All data will be destroyed after 7 years. Decisions to stop participating will be respected. To offset participants' personal and travel costs, we will provide bus tickets for assessments and yoga sessions, and we will provide parking reimbursement, snacks, and honoraria for the assessments.

Intervention Protocols

Yoga Group

Groups of 4 to 5 participants will engage in 60-min group-based Hatha-style yoga sessions 3 times per week for 12 weeks under the supervision of a yoga-certified physiotherapist at a local yoga studio. Classes will begin with a 15-min warm-up, which includes seated meditation, breathing exercises, shoulder and neck stretches, back mobility exercises, and sun salutations. Then, participants will perform 10 min of standing and 15 min of balance poses, followed by 10 min of abdominal work and back-bends. The class will finish with 5 min of final rest (savasana). The yoga protocol can be seen in Table 1.

Table 1. Yoga protocol.

Warm-up (15 min)	Standing poses (10 min)	Balance poses (15 min)	Abdominals and back bends (10 min)	Cool down (10 min)
Seated meditation; alternate nostril breathing; bellows breath; shoulder/neck stretches; cat-cow forward fold; sun salutations	Warrior 1; warrior 2; triangle; extended side angle; reverse warrior; high lunge with twist	Tree pose; eagle pose; standing holding knee; modified warrior 3 (chair support); half moon	Bird-dog; side plank; bridge; cobra; sphinx	Twist; cobbler's pose; hip stretches; corpse pose; side-lying; seated om

Yoga Protocol

Yoga mats, blocks, chairs, and straps will be provided to the participants. Postures will be modified for people with balance impairments or neuropathies. If participants are unable to get down to the floor or balance without support, postures will be performed with the use of a chair or other props. As Indigenous people are overrepresented in the HIV epidemic in Canada (they represented 11.3% of all new infections in 2016) [84], the sample population should reflect the cultural diversity within the catchment area of the study. Every month, a smudging ceremony with an Elder representing the Indigenous people will take place for 5 to 10 min before class commencement. The rationale for performing the smudging ceremony is that it is commonly associated with yoga practices [85]; in fact, a recent survey of 360 yoga practitioners identified spirituality as a common reason for starting and maintaining their yoga practice [86].

Attendance Policy

Of the total of 36 sessions (3 classes a week for 12 weeks), each participant will be encouraged to attend 70% of classes. Consideration will be given to withdrawing a participant from the study if the participant cancels or does not attend more than 6 sessions for reasons other than illness. In the event of a

reversible illness that results in the participant being absent for more than 6 sessions, the participant will be withdrawn from the study and offered to be reenrolled in the yoga group after an 8-week washout period. If a session is cancelled, a make-up session will be scheduled.

Control Group

The control group will be asked to continue with its regular exercise routine, and the group will be asked to not make any changes during the study. Interested participants in the control group will be offered the opportunity to attend ongoing yoga classes as frequently as they would like, following study completion.

Assessment Protocol

As per the CONSORT extension for pilot RCTs, the number of participants screened for eligibility, randomly assigned, received intended treatment, and assessed for each objective will be recorded [81]. Study data will be collected in the Physiotherapy department at Dalhousie University and managed using Research Electronic Data Capture (REDCap) software (REDCap Inc) [87]. The authors will provide access to the study's REDCap (REDCap Inc) data upon request. Table 2 outlines the outcome variables and measurement tools.

XSL•FO

RenderX

Table 2. Outcomes and measurement tools.

Outcome	Measurement tool	Objective measure or self-report
Yoga readiness (previous experience with yoga, injuries and ability level)	Yoga readiness questionnaire (yoga group only)	Self-report
Cognitive performance	Communicating Cognitive Concerns questionnaire; Brief Cognitive Ability Measure	Self-report; objective
Motor function (balance, walking speed)	Community Mobility and Balance Scale; 10-meter walk test	Objective; objective
Mental health	Hospital Anxiety and Depression Scale	Self-report
Quality of life	Medical Outcomes Survey	Self-report
Medication adherence	Simplified Medication Adherence Questionnaire	Self-report
Physical activity	Rapid Assessment of Physical Activity; Fitbit Flex 2—total distance walked (km) and steps taken per day	Self-report; objective
Participant satisfaction, safety, comfort, fatigue, and benefits	Postparticipation survey (yoga group only)	Self-report

Outcome Variables and Measurement Tools

Demographic Information

We will administer a 13-item paper-based self-reported questionnaire asking about age, sex, gender, ethnicity, education level, employment, income, comorbidities, year diagnosed with HIV, viral load (if known), CD4 count (if known), medications, comorbidities, and physical activity (how often the participant was physically active in the previous week) at baseline to describe the sample and assess group comparability. Participants randomized to the yoga group will be asked to fill out a yoga-readiness questionnaire we created to provide the yoga instructor with safety and injury information.

Primary Measures

Many domains of feasibility will be assessed by both participants and study personnel using monitoring processes and a 13-item paper-based post-intervention questionnaire, which includes both questions on a Likert scale ranging from strongly disagree to strongly agree and open-ended questions (see Multimedia Appendix 1):

1. Project coordination (team building, communication and meetings, collaboration, consensus building, troubleshooting, scheduling, protocol consistency, and timelines). Any issues with (or changes to) the study protocol or scheduling will be documented.
2. Participant issues (recruitment, comfort, satisfaction, safety, attendance, time commitment, attrition, and reasons for ineligibility drop out/declining to participate), as assessed by the postintervention questionnaire and documentation by the study coordinator.
3. Assessment protocol elements (time and personnel requirements, usefulness of outcome variables, participant burden, and feasibility) will be recorded by the study coordinator.
4. Intervention protocols (time, equipment, and personnel requirements) will be recorded by the study coordinator.
5. Data quality (completeness, intra/interparticipant variability, interpretability, and trends) will be checked by the study coordinator. Per the CONSORT checklist, our a priori

adherence and satisfaction criteria will be met if participants attend 70% of the yoga sessions and if 70% of the participants are satisfied with the yoga intervention as per the postparticipation questionnaire.

Secondary and Tertiary Measures

Cognition, physical performance (balance, walking speed), physical activity, and affective (mental health, quality of life, and medication adherence) evaluations will be administered at baseline and postintervention (12 weeks) by a trained assessor, blinded to the group assignment. The rationale for blinding the assessor is to reduce bias in scoring during the assessment sessions. The estimated length of time for the assessment sessions is 2 hours per participant. We will measure *cognitive function* using the Brief Cognitive Ability Measure (B-CAM), a computerized cognitive test developed for PLWH, using Rasch measurement theory and analysis that takes 30 min to administer [88,89]. The B-CAM provides a measure of global cognition that is calibrated—the intervals between logits are equal, meaning the data are continuous [88,90].

Cognitive domains tested with the B-CAM include visual detection (reaction time), Flanker task (response inhibition) [91], memory (learning and recall of 8 words), Shape 2-back (working memory) [92], Corsi block-tapping forward and back tests (visuospatial memory) [93], verbal fluency (letters F-A-S in English) mini Trail Making Test B (executive function) [94], and the Tower of London test (planning) [95]. The scoring of the B-CAM ranges from 0 to 24, with higher values indicating better global cognition [90]. To reduce the likelihood of practice effects, different versions of the B-CAM are performed at baseline and final assessments [90]. Group-based trajectory analysis has revealed that no practice effects were found at the item level [90].

Self-reported cognition will also be assessed using the C3Q, an 18-item paper-based questionnaire that was developed to estimate the presence and frequency of memory, attention, executive function, visuospatial, speech and language, behavior and emotion, and cognitive challenges among PLWH [82]. The frequency of such challenges are recorded by the participant on

a 3-point scale: frequently (almost every day), sometimes (once a week), or rarely (once a month) [82].

Balance will be measured using the Community Balance and Mobility (CB&M) test, a high-level balance assessment of tasks performed in the community, developed for people with traumatic brain injury [96]. It is a valid and reliable measure of dynamic postural control in people with traumatic brain injury [96,97] and older community-dwelling individuals [98], and it is not as susceptible to ceiling effects as the Berg Balance Scale [98,99]. *Walking speed* will be measured using the 10-meter walk test because of the association of gait speed with cognitive performance in PLWH [24], its previous use in the HIV literature [26], and its ability to predict survival in older adults [100]. *Depression* will be assessed using the Hospital Anxiety and Depression Scale, a paper-based self-report questionnaire [101], which has very good to excellent internal consistency, test-retest reliability and convergent validity, and acceptable discriminant validity in PLWH [102]. Quality of life will be assessed using MOS-HIV, a paper-based questionnaire that comprises 10 domains (physical function, social function, role function, cognitive function, pain, mental health, energy, health distress, quality of life, and overall health), with good to high internal consistency and construct validity in PLWH [103]. *Physical activity* will be assessed using the Rapid Assessment of Physical Activity, a 9-item paper-based questionnaire that measures moderate and vigorous physical activity, including strength and flexibility within the last week [104]. It was validated in older adults [104], and it has been used in studies with people with HIV [105]. Objective levels of physical activity (total distance walked, and number of steps taken per day) will be measured using accelerometers (Fitbit flex 2) [106]. Accelerometer data will be electronically synced and downloaded after weeks 1 and 12 and stored in an encrypted file. Participants will also be asked about *Medication adherence* (specifically cART), measured with the paper-based Simplified Medication Adherence Questionnaire (SMAQ), which has 72% sensitivity, 91% specificity, and a likelihood ratio of 7.94 for nonadherent patients [107].

Participant Safety

Participants will be monitored throughout the yoga sessions and the assessments. If a participant presents with any medical or safety concerns, the supervising physiotherapist will provide the appropriate first aid or injury treatment; then, the supervising physiotherapist will refer the participant to the participant's family physician for follow-up. Any harms or unanticipated effects will be recorded as per the CONSORT checklist [81]. Owing to the low-risk nature of the study, we do not anticipate any additional safety or medical issues associated with the yoga interventions.

Results

Data Analysis

All questionnaires and measures will be assessed for missing data. The data will be analyzed to determine if the assumptions for parametric tests are met. Descriptive statistics will be used to characterize the participants. As per the CONSORT extension for pilot studies, means of all outcomes, mean change, and 95%

CIs will be calculated for each group separately. We will also follow the Sex and Gender Equity in Research guidelines [108] by disaggregating data by sex and gender. Participant dropouts will also be reported disaggregated by sex.

Independent *t* tests and Fisher exact tests will be used to compare the 2 groups at baseline. If the 2 groups differ at baseline, that variable will be included in the analysis as a covariate. We will analyze quantitative postintervention questionnaire responses using Chi-square tests, and open-ended responses will be analyzed thematically. Intention-to-treat and per-protocol analyses will be used in the analysis of the secondary variables. Changes in outcome variables will be examined between groups and within groups. Floor and ceiling effects will be calculated for the CB&M test. Effect sizes will be reported for each outcome. Alpha level will be set at .05, using 2-tailed for all inferences, and data will be analyzed with SPSS Version 25 (SPSS Inc). As this is a pilot study, sample size calculations are not recommended [81]. This pilot study will not be adequately powered to conclusively state the influence of the intervention on study outcomes, but if trends are promising, a future, more adequately powered trial will be planned. This pilot study will provide preliminary data for future sample size calculations.

Study enrollment began in January 2018, with results expected in October 2019.

Dissemination

Study results will be disseminated to PLWH, researchers, health care providers, community-based organizations, stakeholders, and policy makers. Knowledge translation will take place via peer-reviewed journals, podium and poster presentations at conferences and forums, newsletters, and presentations at community-based organizations.

Discussion

Study Strengths

This pilot implementation trial will be the first to investigate the effect and feasibility of a yoga intervention on cognitive and physical outcomes in PLWH. Not only will the study generate preliminary data about the effects of yoga on cognitive and physical function, but it will also inform the feasibility and utility of further investigation in terms of team capacity building, recruitment and retention strategies, and assessment of intervention protocols. The focus of the project is clearly aligned with a key research priority of the Canada-International HIV and Rehabilitation Research Collaborative, which is to determine the effectiveness of rehabilitation interventions and service delivery models [109].

Our research addresses HIV beyond a biological perspective to reduce not only physical limitations but also the social impact of HIV. By targeting an inexpensive nonpharmacological intervention, we hope to identify feasible community-based strategies that may contribute to slowing the health-related consequences of HIV while improving quality of life for PLWH.

Anticipated Challenges and Limitations

Potential challenges will include recruitment and retention of participants over the course of the 12-week intervention. With

approximately 500 PLWH living in the local area [110], we anticipate that by involving community leaders and end users from the outset of conceptualization and planning and conducting the study in a familiar community setting, we will successfully recruit 25 PLWH. Although attrition is of concern in exercise studies requiring multiple visits, a 2015 study on yoga and meditation reported an overall attendance rate of 89% among PLWH [111].

Study limitations include a lack of mechanism to confirm HIV diagnoses for participants not recruited from the HIV clinic and limited study inclusion to individuals who speak and understand English, which may reduce the generalizability of our findings. Participants were also not asked about substance abuse or specific comorbidities, such as peripheral neuropathy, which may affect cognitive and physical performance.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Postparticipation questionnaire.

[PDF File (Adobe PDF File), 40KB - resprot_v8i5e13818_app1.pdf]

Multimedia Appendix 2

Canadian Institutes of Health Research peer-review reports.

[PDF File (Adobe PDF File), 1008KB - resprot_v8i5e13818_app2.pdf]

References

1. Guaraldi G, Orlando G, Zona S, Menozzi M, Carli F, Garlassi E, et al. Premature age-related comorbidities among HIV-infected persons compared with the general population. Clin Infect Dis 2011 Dec;53(11):1120-1126. [doi: 10.1093/cid/cir627] [Medline: 21998278]
2. Grant I. Neurocognitive disturbances in HIV. Int Rev Psychiatry 2008 Feb;20(1):33-47. [doi: 10.1080/09540260701877894] [Medline: 18240061]
3. Goodkin K. Psychiatric Aspects of HIV Spectrum Disease. Focus (Madison) 2009;7(3):303-310. [doi: 10.1176/foc.7.3.foc303]
4. Cysique L, Bain M, Brew B, Murray J. The burden of HIV-associated neurocognitive impairment in Australia and its estimates for the future. Sex Health 2011 Dec;8(4):541-550. [doi: 10.1071/SH11003] [Medline: 22127041]
5. Public Health Agency of Canada. 2010. HIV/AIDS Epi Updates: HIV/AIDS among older Canadians URL: https://tinyurl.com/yylmemo9 [accessed 2019-05-06] [WebCite Cache ID 78B1kQbIn]
6. Vance DE, McDougall GJ, Wilson N, Debiasi MO, Cody SL. Cognitive consequences of aging with HIV: implications for neuroplasticity and rehabilitation. Top Geriatr Rehabil 2014 Jan;30(1):35-45 [FREE Full text] [doi: 10.1097/TGR.0000000000000002] [Medline: 24817785]
7. Hardy D, Vance D. The neuropsychology of HIV/AIDS in older adults. Neuropsychol Rev 2009 Jun;19(2):263-272. [doi: 10.1007/s11065-009-9087-0] [Medline: 19247836]
8. Cohen RA, Seider TR, Navia B. HIV effects on age-associated neurocognitive dysfunction: premature cognitive aging or neurodegenerative disease? Alzheimers Res Ther 2015;7(1):37 [FREE Full text] [doi: 10.1186/s13195-015-0123-4] [Medline: 25848401]
9. Sabin CA. Do people with HIV infection have a normal life expectancy in the era of combination antiretroviral therapy? BMC Med 2013 Nov 27;11:251 [FREE Full text] [doi: 10.1186/1741-7015-11-251] [Medline: 24283830]
10. Gongvatana A, Harezlak J, Buchthal S, Daar E, Schifitto G, Campbell T, et al. Progressive cerebral injury in the setting of chronic HIV infection and antiretroviral therapy. J Neurovirol 2013 Jun;19(3):209-218 [FREE Full text] [doi: 10.1007/s13365-013-0162-1] [Medline: 23613008]
11. McArthur JC, Steiner J, Sacktor N, Nath A. Human immunodeficiency virus-associated neurocognitive disorders: mind the gap. Ann Neurol 2010 Jun;67(6):699-714. [doi: 10.1002/ana.22053] [Medline: 20517932]
12. Mattson MP. Exercise and the brain: a slap on the HAND. J Neurovirol 2013 Oct;19(5):407-409 [FREE Full text] [doi: 10.1007/s13365-013-0208-4] [Medline: 24072548]

13. Alfahad TB, Nath A. Update on HIV-associated neurocognitive disorders. Curr Neurol Neurosci Rep 2013 Oct;13(10):387 [FREE Full text] [doi: 10.1007/s11910-013-0387-7] [Medline: 23954972]

14. Hopcroft L, Bester L, Clement D, Quigley A, Sachdeva M, Rourke SB, et al. "My body's a 50 year-old but my brain is definitely an 85 year-old": exploring the experiences of men ageing with HIV-associated neurocognitive challenges. J Int AIDS Soc 2013 Jul 23;16:18506 [FREE Full text] [doi: 10.7448/IAS.16.1.18506] [Medline: 23883539]

15. Tozzi V, Balestra P, Murri R, Galgani S, Bellagamba R, Narciso P, et al. Neurocognitive impairment influences quality of life in HIV-infected patients receiving HAART. Int J STD AIDS 2004 Apr;15(4):254-259. [doi: 10.1258/095646204773557794] [Medline: 15075020]

16. Gorman AA, Foley JM, Ettenhofer ML, Hinkin CH, van Gorp WG. Functional consequences of HIV-associated neuropsychological impairment. Neuropsychol Rev 2009 Jun;19(2):186-203 [FREE Full text] [doi: 10.1007/s11065-009-9095-0] [Medline: 19472057]

17. Becker BW, Thames AD, Woo E, Castellon SA, Hinkin CH. Longitudinal change in cognitive function and medication adherence in HIV-infected adults. AIDS Behav 2011 Nov;15(8):1888-1894 [FREE Full text] [doi: 10.1007/s10461-011-9924-z] [Medline: 21437726]

18. Hinkin CH, Castellon SA, Durvasula RS, Hardy DJ, Lam MN, Mason KI, et al. Medication adherence among HIV+ adults: effects of cognitive dysfunction and regimen complexity. Neurology 2002 Dec 24;59(12):1944-1950 [FREE Full text] [Medline: 12499488]

19. Chesney M, Morin M, Sherr L. Adherence to HIV combination therapy. Soc Sci Med 2000 Jun;50(11):1599-1605. [Medline: 10795966]

20. Heaton R, Marcotte T, Mindt MR, Sadek J, Moore DJ, Bentley H, HNRC Group. The impact of HIV-associated neuropsychological impairment on everyday functioning. J Int Neuropsychol Soc 2004 May;10(3):317-331. [doi: 10.1017/S1355617704102130] [Medline: 15147590]

21. Simioni S, Cavassini M, Annoni J, Rimbault Abraham A, Bourquin I, Schiffer V, et al. Cognitive dysfunction in HIV patients despite long-standing suppression of viremia. AIDS 2010 Jun 01;24(9):1243-1250. [doi: 10.1097/QAD.0b013e3283354a7b] [Medline: 19996937]

22. Woods SP, Weber E, Weisz BM, Twamley EW, Grant I, HIV Neurobehavioral Research Programs Group. Prospective memory deficits are associated with unemployment in persons living with HIV infection. Rehabil Psychol 2011 Feb;56(1):77-84 [FREE Full text] [doi: 10.1037/a0022753] [Medline: 21401289]

23. van Gorp WG, Rabkin J, Ferrando S, Mintz J, Ryan E, Borkowski T, et al. Neuropsychiatric predictors of return to work in HIV/AIDS. J Int Neuropsychol Soc 2007 Jan;13(1):80-89. [doi: 10.1017/S1355617707070117] [Medline: 17166306]

24. Robertson K, Parsons T, Sidtis J, Hanlon Inman T, Robertson WT, Hall CD, et al. Timed Gait test: normative data for the assessment of the AIDS dementia complex. J Clin Exp Neuropsychol 2006 Oct;28(7):1053-1064. [doi: 10.1080/13803390500205684] [Medline: 16840235]

25. Bauer LO, Wu Z, Wolfson LI. An obese body mass increases the adverse effects of HIV/AIDS on balance and gait. Phys Ther 2011 Jul;91(7):1063-1071 [FREE Full text] [doi: 10.2522/ptj.20100292] [Medline: 21527386]

26. Richert L, Brault M, Mercié P, Dauchy FA, Bruyand M, Greib C, Groupe d'Epidémiologie Clinique du SIDA en Aquitaine (GECSA). Decline in locomotor functions over time in HIV-infected patients. AIDS 2014 Jun 19;28(10):1441-1449. [doi: 10.1097/QAD.0000000000000246] [Medline: 24566096]

27. Berner K, Morris L, Baumeister J, Louw Q. Objective impairments of gait and balance in adults living with HIV-1 infection: a systematic review and meta-analysis of observational studies. BMC Musculoskelet Disord 2017 Aug 01;18(1):325 [FREE Full text] [doi: 10.1186/s12891-017-1682-2] [Medline: 28764704]

28. Erlandson KM, Allshouse AA, Jankowski CM, Duong S, MaWhinney S, Kohrt WM, et al. Risk factors for falls in HIV-infected persons. J Acquir Immune Defic Syndr 2012 Dec 01;61(4):484-489 [FREE Full text] [doi: 10.1097/QAI.0b013e3182716e38] [Medline: 23143526]

29. Simmonds M, Novy D, Sandoval R. The differential influence of pain and fatigue on physical performance and health status in ambulatory patients with human immunodeficiency virus. Clin J Pain 2005;21(3):200-206. [Medline: 15818071]

30. Beans J, Stevenson T, Katzel LI, Sorkin JD, Warner AL, Gottlieb SS, et al. Ambulatory Function in Men with and without HIV Infection: Association with Cardiorespiratory Fitness. J AIDS Clin Res 2013;s9:003. [doi: 10.4172/2155-6113.S9-003]

31. Bauer L, Ceballos N, Shanley J, Wolfson L. Sensorimotor dysfunction in HIV/AIDS: effects of antiretroviral treatment and comorbid psychiatric disorders. AIDS 2005 Mar 25;19(5):495-502 [FREE Full text] [Medline: 15764855]

32. Erlandson K, Allshouse A, Jankowski C, Duong S, Mawhinney S, Kohrt WM, et al. Comparison of functional status instruments in HIV-infected adults on effective antiretroviral therapy. HIV Clin Trials 2012;13(6):324-334 [FREE Full text] [doi: 10.1310/hct1306-324] [Medline: 23195670]

33. Erlandson KM, Allshouse AA, Jankowski CM, Mawhinney S, Kohrt WM, Campbell TB. Relationship of physical function and quality of life among persons aging with HIV infection. AIDS 2014 Aug 24;28(13):1939-1943 [FREE Full text] [doi: 10.1097/QAD.0000000000000384] [Medline: 24992000]

34. Sullivan EV, Rosenbloom MJ, Rohlfing T, Kemper CA, Deresinski S, Pfefferbaum A. Pontocerebellar contribution to postural instability and psychomotor slowing in HIV infection without dementia. Brain Imaging Behav 2011 Mar;5(1):12-24 [FREE Full text] [doi: 10.1007/s11682-010-9107-y] [Medline: 20872291]

35.   Tanon A, Diallo Z, Lenaud S, Niangoran S, Peres K, Eholié SP, et al. Aging with HIV infection and locomotors disorders: experience of the infectious and tropical diseases unit, Abidjan, Côte d'Ivoire. J AIDS Clin Res 2017;8:680 [FREE Full text] [doi: 10.4172/2155-6113.1000680]

36.   Richert L, Dehail P, Mercié P, Dauchy FA, Bruyand M, Greib C, Groupe d'Epidémiologie Clinique du SIDA en Aquitaine (GECSA). High frequency of poor locomotor performance in HIV-infected patients. AIDS 2011 Mar 27;25(6):797-805. [doi: 10.1097/QAD.0b013e3283455dff] [Medline: 21330905]

37.   Quigley A, O'Brien K, Parker R, MacKay-Lyons M. Exercise and cognitive function in people living with HIV: a scoping review. Disabil Rehabil 2019 Jun;41(12):1384-1395. [doi: 10.1080/09638288.2018.1432079] [Medline: 29376434]

38.   Baigis J, Korniewicz D, Chase G, Butz A, Jacobson D, Wu A. Effectiveness of a home-based exercise intervention for HIV-infected adults: a randomized trial. J Assoc Nurses AIDS Care 2002;13(2):33-45. [doi: 10.1016/S1055-3290(06)60199-4] [Medline: 11936063]

39.   Galantino M, Shepard K, Krafft L, Laperriere A, Ducette J, Sorbello A, et al. The effect of group aerobic exercise and t'ai chi on functional outcomes and quality of life for persons living with acquired immunodeficiency syndrome. J Altern Complement Med 2005 Dec;11(6):1085-1092. [doi: 10.1089/acm.2005.11.1085] [Medline: 16398601]

40.   Gillespie J. The Ohio State University. 1997. Health promotion and quality of life in HIV-1 infected individuals URL: https://etd.ohiolink.edu/!etd.send_file?accession=osu1487948158627446&disposition=inline

41.   Fillipas S, Oldmeadow L, Bailey M, Cherry C. A six-month, supervised, aerobic and resistance exercise program improves self-efficacy in people with human immunodeficiency virus: a randomised controlled trial. Aust J Physiother 2006;52(3):185-190 [FREE Full text] [Medline: 16942453]

42.   McDermott A, Zaporojan L, McNamara P, Doherty CP, Redmond J, Forde C, et al. The effects of a 16-week aerobic exercise programme on cognitive function in people living with HIV. AIDS Care 2017 Dec;29(6):667-674. [doi: 10.1080/09540121.2016.1263723] [Medline: 27892704]

43.   Robins J, McCain N, Gray D, Elswick R, Walter J, McDade E. Research on psychoneuroimmunology: tai chi as a stress management approach for individuals with HIV disease. Appl Nurs Res 2006 Feb;19(1):2-9 [FREE Full text] [doi: 10.1016/j.apnr.2005.03.002] [Medline: 16455435]

44.   Brown D, Claffey A, Harding R. Evaluation of a physiotherapy-led group rehabilitation intervention for adults living with HIV: referrals, adherence and outcomes. AIDS Care 2016 Dec;28(12):1495-1505. [doi: 10.1080/09540121.2016.1191611] [Medline: 27264319]

45.   Schlabe S, Vogel M, Boesecke C, Schwarze-Zander C, Rockstroh JK, Körner C, et al. Moderate endurance training (marathon-training) - effects on immunologic and metabolic parameters in HIV-infected patients: the 42 KM cologne project. BMC Infect Dis 2017 Dec 08;17(1):550 [FREE Full text] [doi: 10.1186/s12879-017-2651-y] [Medline: 28789630]

46.   Fazeli P, Marquine M, Dufour C, Henry BL, Montoya J, Gouaux B, HNRP Group. Physical activity is associated with better neurocognitive and everyday functioning among older adults with HIV disease. AIDS Behav 2015 Aug;19(8):1470-1477 [FREE Full text] [doi: 10.1007/s10461-015-1024-z] [Medline: 25731660]

47.   Dufour CA, Marquine MJ, Fazeli PL, Henry BL, Ellis RJ, Grant I, HNRP Group. Physical exercise is associated with less neurocognitive impairment among HIV-infected adults. J Neurovirol 2013 Oct;19(5):410-417 [FREE Full text] [doi: 10.1007/s13365-013-0184-8] [Medline: 23934585]

48.   Mapstone M, Hilton T, Yang H, Guido JJ, Luque AE, Hall WJ, et al. Poor aerobic fitness may contribute to cognitive decline in HIV-infected older adults. Aging Dis 2013;4(6):311-319 [FREE Full text] [doi: 10.14336/AD.2013.0400311] [Medline: 24307964]

49.   Monroe A, Zhang L, Jacobson L, Plankey MW, Brown TT, Miller EN, et al. The association between physical activity and cognition in men with and without HIV infection. HIV Med 2017 Dec;18(8):555-563 [FREE Full text] [doi: 10.1111/hiv.12490] [Medline: 28294530]

50.   Ortega M, Baker L, Vaida F, Paul R, Basco B, Ances BM. Physical activity affects brain integrity in HIV+ individuals. J Int Neuropsychol Soc 2015 Nov;21(10):880-889 [FREE Full text] [doi: 10.1017/S1355617715000879] [Medline: 26581799]

51.   Honn V, Para M, Whitacre C, Bornstein R. Effect of exercise on neuropsychological performance in asymptomatic HIV infection. AIDS Behav 1999;3(1):67-74. [doi: 10.1023/A:1025471503738]

52.   Fazeli P, Woods S, Heaton R, Umlauf A, Gouaux B, Rosario D, HNRP Group. An active lifestyle is associated with better neurocognitive functioning in adults living with HIV infection. J Neurovirol 2014 Jun;20(3):233-242 [FREE Full text] [doi: 10.1007/s13365-014-0240-z] [Medline: 24554483]

53.   Dufour C, Marquine M, Fazeli P, Umlauf A, Henry BL, Zlatar Z, HIV Neurobehavioral Research Program Group. A longitudinal analysis of the impact of physical activity on neurocognitive functioning among HIV-infected adults. AIDS Behav 2018 Dec;22(5):1562-1572 [FREE Full text] [doi: 10.1007/s10461-016-1643-z] [Medline: 27990580]

54.   Janssen MA, Bosch M, Koopmans PP, Kessels RP. Validity of the Montreal Cognitive Assessment and the HIV Dementia Scale in the assessment of cognitive impairment in HIV-1 infected patients. J Neurovirol 2015 Aug;21(4):383-390 [FREE Full text] [doi: 10.1007/s13365-015-0324-4] [Medline: 25678141]

55.   Field T. Yoga research review. Complement Ther Clin Pract 2016 Aug;24:145-161. [doi: 10.1016/j.ctcp.2016.06.005] [Medline: 27502816]

56. Ross A, Williams L, Pappas-Sandonas M, Touchton-Leonard K, Fogel D. Incorporating yoga therapy into primary care: the Casey Health Institute. Int J Yoga Therap 2015;25(1):43-49. [doi: 10.17761/1531-2054-25.1.43] [Medline: 26667288]

57. Slocum-Gori S, Howard A, Balneaves L, Kazanjian A. Investigating the perceived feasibility of integrative medicine in a conventional oncology setting: yoga therapy as a treatment for breast cancer survivors. Integr Cancer Ther 2013 Mar;12(2):103-112 [FREE Full text] [doi: 10.1177/1534735412443851] [Medline: 22710259]

58. Gothe N, McAuley E. Yoga and cognition: a meta-analysis of chronic and acute effects. Psychosom Med 2015 Sep;77(7):784-797. [doi: 10.1097/PSY.0000000000000218] [Medline: 26186435]

59. Gothe N, Pontifex M, Hillman C, McAuley E. The acute effects of yoga on executive function. J Phys Act Health 2013 May;10(4):488-495. [Medline: 22820158]

60. Innes KE, Bourguignon C, Taylor AG. Risk indices associated with the insulin resistance syndrome, cardiovascular disease, and possible protection with yoga: a systematic review. J Am Board Fam Pract 2005;18(6):491-519 [FREE Full text] [Medline: 16322413]

61. Ross A, Thomas S. The health benefits of yoga and exercise: a review of comparison studies. J Altern Complement Med 2010 Jan;16(1):3-12. [doi: 10.1089/acm.2009.0044] [Medline: 20105062]

62. Pascoe MC, Bauer IE. A systematic review of randomised control trials on the effects of yoga on stress measures and mood. J Psychiatr Res 2015 Sep;68:270-282. [doi: 10.1016/j.jpsychires.2015.07.013] [Medline: 26228429]

63. Wang C, Bannuru R, Ramel J, Kupelnick B, Scott T, Schmid CH. Tai Chi on psychological well-being: systematic review and meta-analysis. BMC Complement Altern Med 2010 May 21;10:23 [FREE Full text] [doi: 10.1186/1472-6882-10-23] [Medline: 20492638]

64. Gothe NP, Keswani RK, McAuley E. Yoga practice improves executive function by attenuating stress levels. Biol Psychol 2016 Dec;121(Pt A):109-116. [doi: 10.1016/j.biopsycho.2016.10.010] [Medline: 27794449]

65. Gothe N, Kramer A, McAuley E. Hatha yoga practice improves attention and processing speed in older adults: results from an 8-week randomized control trial. J Altern Complement Med 2017 Jan;23(1):35-40. [doi: 10.1089/acm.2016.0185] [Medline: 27809558]

66. Simon R, Engström M. The default mode network as a biomarker for monitoring the therapeutic effects of meditation. Front Psychol 2015;6:776. [doi: 10.3389/fpsyg.2015.00776] [Medline: 26106351]

67. Fernandez-Duque D, Baird JA, Posner MI. Executive attention and metacognitive regulation. Conscious Cogn 2000 Jun;9(2 Pt 1):288-307. [doi: 10.1006/ccog.2000.0447] [Medline: 10924249]

68. Schmid A, Van Puymbroeck M, Altenburger P, Schalk NL, Dierks TA, Miller KK, et al. Poststroke balance improves with yoga: a pilot study. Stroke 2012 Sep;43(9):2402-2407. [doi: 10.1161/STROKEAHA.112.658211] [Medline: 22836351]

69. Guner S, Inanici F. Yoga therapy and ambulatory multiple sclerosis assessment of gait analysis parameters, fatigue and balance. J Bodyw Mov Ther 2015 Jan;19(1):72-81. [doi: 10.1016/j.jbmt.2014.04.004] [Medline: 25603746]

70. Schmid A, Van Puymbroeck M, Miller K, Schalk N. Group yoga intervention leads to improved balance and balance self-efficacy after stroke. BMC Complement Altern Med 2012;12(Suppl 1):P222 [FREE Full text] [doi: 10.1186/1472-6882-12-S1-P222]

71. Schmid A, Van Puymbroeck M, Altenburger P, Dierks TA, Miller KK, Damush TM, et al. Balance and balance self-efficacy are associated with activity and participation after stroke: a cross-sectional study in people with chronic stroke. Arch Phys Med Rehabil 2012 Jun;93(6):1101-1107. [doi: 10.1016/j.apmr.2012.01.020] [Medline: 22502804]

72. Schmid A, Van Puymbroeck M, Koceja D. Effect of a 12-week yoga intervention on fear of falling and balance in older adults: a pilot study. Arch Phys Med Rehabil 2010 Apr;91(4):576-583. [doi: 10.1016/j.apmr.2009.12.018] [Medline: 20382290]

73. Tiedemann A, O'Rourke S, Sesto R, Sherrington C. A 12-week Iyengar yoga program improved balance and mobility in older community-dwelling people: a pilot randomized controlled trial. J Gerontol A Biol Sci Med Sci 2013 Sep;68(9):1068-1075. [doi: 10.1093/gerona/glt087] [Medline: 23825035]

74. Halder K, Chatterjee A, Pal R, Tomer OS, Saha M. Age related differences of selected Hatha yoga practices on anthropometric characteristics, muscular strength and flexibility of healthy individuals. Int J Yoga 2015 Jan;8(1):37-46 [FREE Full text] [doi: 10.4103/0973-6131.146057] [Medline: 25558132]

75. Eyre H, Acevedo B, Yang H, Siddarth P, Van Dyk K, Ercoli L, et al. Changes in neural connectivity and memory following a yoga intervention for older adults: a pilot study. J Alzheimers Dis 2016;52(2):673-684 [FREE Full text] [doi: 10.3233/JAD-150653] [Medline: 27060939]

76. Youkhana S, Dean CM, Wolff M, Sherrington C, Tiedemann A. Yoga-based exercise improves balance and mobility in people aged 60 and over: a systematic review and meta-analysis. Age Ageing 2016 Jan;45(1):21-29. [doi: 10.1093/ageing/afv175] [Medline: 26707903]

77. Kietrys DM, Galantino ML, Cohen ET, Parrott JS, Gould-Fogerite S, O Brien KK. Yoga for persons with HIV-related distal sensory polyneuropathy. Rehabil Oncol 2018;36(2):123-131. [doi: 10.1097/01.REO.0000000000000089]

78. Mawar N, Katendra T, Bagul R, Bembalkar S, Vedamurthachar A, Tripathy S, et al. Sudarshan Kriya yoga improves quality of life in healthy people living with HIV (PLHIV): results from an open label randomized clinical trial. Indian J Med Res 2015;141(1):90. [doi: 10.4103/0971-5916.154509]

XSL•FO

RenderX

79. Naoroibam R, Metri KG, Bhargav H, Nagaratna R, Nagendra HR. Effect of Integrated Yoga (IY) on psychological states and CD4 counts of HIV-1 infected patients: a randomized controlled pilot study. Int J Yoga 2016;9(1):57-61 [FREE Full text] [doi: 10.4103/0973-6131.171723] [Medline: 26865772]

80. Eldridge S, Lancaster G, Campbell M, Thabane L, Hopewell S, Coleman CL, et al. Defining feasibility and pilot studies in preparation for randomised controlled trials: development of a conceptual framework. PLoS One 2016;11(3):e0150205 [FREE Full text] [doi: 10.1371/journal.pone.0150205] [Medline: 26978655]

81. Eldridge S, Chan C, Campbell M, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Pilot Feasibility Stud 2016;2:64 [FREE Full text] [doi: 10.1186/s40814-016-0105-8] [Medline: 27965879]

82. Askari S, Fellows L, Brouillette M, Moriello C, Duracinsky M, Mayo NE. Development of an item pool reflecting cognitive concerns expressed by people with HIV. Am J Occup Ther 2018 Mar;72(2):1-9. [doi: 10.5014/ajot.2018.023945] [Medline: 29426385]

83. Bredin SS, Gledhill N, Jamnik VK, Warburton DE. PAR-Q+ and ePARmed-X+: new risk stratification and physical activity clearance strategy for physicians and patients alike. Can Fam Physician 2013 Mar;59(3):273-277 [FREE Full text] [Medline: 23486800]

84. Public Health Agency of Canada. Public Health Agency of Canada. 2016. Summary: estimates of HIV Incidence, Prevalence and Canada's Progress on Meeting the 90-90-90 HIV Targets URL: https://tinyurl.com/y2x5p449 [accessed 2019-01-05] [WebCite Cache ID 77f477hTd]

85. Kripalu Center for Yoga & Health. The Sacred Art of Smudging URL: https://kripalu.org/resources/sacred-art-smudging [accessed 2019-02-25] [WebCite Cache ID 76SWb5h2O]

86. Park C, Riley K, Bedesin E, Stewart V. Why practice yoga? Practitioners' motivations for adopting and maintaining yoga practice. J Health Psychol 2016 Dec;21(6):887-896. [doi: 10.1177/1359105314541314] [Medline: 25030795]

87. Vanderbilt. REDCap Software. URL: https://www.project-redcap.org/software/ [accessed 2019-01-05] [WebCite Cache ID 76SWXeG2Z]

88. Mayo NE, Brouillette M, Fellows LK, Positive Brain Health Now Investigators. Understanding and optimizing brain health in HIV now: protocol for a longitudinal cohort study with multiple randomized controlled trials. BMC Neurol 2016 Jan 14;16:8 [FREE Full text] [doi: 10.1186/s12883-016-0527-1] [Medline: 26762403]

89. Koski L, Brouillette M, Lalonde R, Hello B, Wong E, Tsuchida A, et al. Computerized testing augments pencil-and-paper tasks in measuring HIV-associated mild cognitive impairment(*). HIV Med 2011 Sep;12(8):472-480 [FREE Full text] [doi: 10.1111/j.1468-1293.2010.00910.x] [Medline: 21395965]

90. Brouillette M, Fellows LK, Finch L, Thomas R, Mayo NE. Properties of a brief assessment tool for longitudinal measurement of cognition in people living with HIV. PLoS One 2019;14(3):e0213908 [FREE Full text] [doi: 10.1371/journal.pone.0213908] [Medline: 30908501]

91. Eriksen BA, Eriksen CW. Effects of noise letters upon the identification of a target letter in a nonsearch task. Percept Psychophys 1974;16(1):143-149. [doi: 10.3758/BF03203267]

92. Kane MJ, Conway AR, Miura TK, Colflesh GJ. Working memory, attention control, and the N-back task: a question of construct validity. J Exp Psychol Learn Mem Cogn 2007 May;33(3):615-622. [doi: 10.1037/0278-7393.33.3.615] [Medline: 17470009]

93. Robbins TW, James M, Owen AM, Sahakian BJ, Lawrence AD, McInnes L, et al. A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: implications for theories of executive functioning and cognitive aging. Cambridge Neuropsychological Test Automated Battery. J Int Neuropsychol Soc 1998 Sep;4(5):474-490. [Medline: 9745237]

94. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]

95. Phillips LH, Wynn VE, McPherson S, Gilhooly KJ. Mental planning and the Tower of London task. Q J Exp Psychol A 2001 May;54(2):579-597. [doi: 10.1080/713755977] [Medline: 11394063]

96. Knorr S, Brouwer B, Garland S. Validity of the Community Balance and Mobility Scale in community-dwelling persons after stroke. Arch Phys Med Rehabil 2010 Jun;91(6):890-896. [doi: 10.1016/j.apmr.2010.02.010] [Medline: 20510980]

97. Howe J, Inness E, Venturini A, Williams J, Verrier M. The Community Balance and Mobility Scale--a balance measure for individuals with traumatic brain injury. Clin Rehabil 2006 Oct;20(10):885-895. [doi: 10.1177/0269215506072183] [Medline: 17008340]

98. Balasubramanian CK. The community balance and mobility scale alleviates the ceiling effects observed in the currently used gait and balance assessments for the community-dwelling older adults. J Geriatr Phys Ther 2015;38(2):78-89. [doi: 10.1519/JPT.0000000000000024] [Medline: 24949849]

99. Inness E, Howe J, Niechwiej-Szwedo E, Jaglal SB, McIlroy WE, Verrier MC. Measuring balance and mobility after traumatic brain injury: validation of the Community Balance and Mobility Scale (CB&M). Physiother Can 2011;63(2):199-208 [FREE Full text] [doi: 10.3138/ptc.2009-45] [Medline: 22379260]

100. Studenski S, Perera S, Patel K, Rosano C, Faulkner K, Inzitari M, et al. Gait speed and survival in older adults. J Am Med Assoc 2011 Jan 5;305(1):50-58 [FREE Full text] [doi: 10.1001/jama.2010.1923] [Medline: 21205966]

101. Zigmond A, Snaith R. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]

102. Savard J, Laberge B, Gauthier J, Ivers H, Bergeron M. Evaluating anxiety and depression in HIV-infected patients. J Pers Assess 1998 Dec;71(3):349-367. [doi: 10.1207/s15327752jpa7103_5] [Medline: 9933941]

103. Skevington S, O'Connell K. Measuring Quality of Life in HIV and AIDS: A Review of the Recent Literature. Psychol Health 2003;18(3):331-350 [FREE Full text] [doi: 10.1080/0887044031000084030]

104. Topolski TD, LoGerfo J, Patrick DL, Williams B, Walwick J, Patrick MB. The Rapid Assessment of Physical Activity (RAPA) among older adults. Prev Chronic Dis 2006 Oct;3(4):A118 [FREE Full text] [Medline: 16978493]

105. O'Brien K, Bayoumi A, Solomon P, Tang A, Murzin K, Carusone SC, et al. Evaluating a community-based exercise intervention with adults living with HIV: protocol for an interrupted time series study. BMJ Open 2016;6(10) [FREE Full text] [doi: 10.1136/bmjopen-2016-013618]

106. Fitbit. 2018. Fitbit Flex 2TM Fitness Wristband URL: https://www.fitbit.com/au/flex2 [accessed 2019-02-25] [WebCite Cache ID 76SWBBSUu]

107. Knobel H, Alonso J, Casado JL, Collazos J, González J, Ruiz I, et al. Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study. AIDS 2002 Mar 8;16(4):605-613. [Medline: 11873004]

108. Heidari S, Babor TF, De Castro P, Tort S, Curno M. Sex and Gender Equity in Research: rationale for the SAGER guidelines and recommended use. Res Integr Peer Rev 2016 May;1:2 [FREE Full text] [doi: 10.1186/s41073-016-0007-6] [Medline: 29451543]

109. O'Brien K, Ibáñez-Carrasco F, Solomon P, Harding R, Cattaneo J, Chegwidden W, et al. Advancing research and practice in HIV and rehabilitation: a framework of research priorities in HIV, disability and rehabilitation. BMC Infect Dis 2014 Dec 31;14:724 [FREE Full text] [doi: 10.1186/s12879-014-0724-8] [Medline: 25551619]

110. Nova Scotia Health and Wellness. 2011. Surveillance Report On HIV/AIDS In NOVA SCOTIA: 1983 to 2011 URL: https://novascotia.ca/dhw/populationhealth/documents/HIV-AIDS-Surveillance-Report.pdf [accessed 2019-01-05] [WebCite Cache ID 76SWhNmGd]

111. Agarwal R, Kumar A, Lewis J. A pilot feasibility and acceptability study of yoga/meditation on the quality of life and markers of stress in persons living with HIV who also use crack cocaine. J Altern Complement Med 2015 Mar;21(3):152-158 [FREE Full text] [doi: 10.1089/acm.2014.0112] [Medline: 25695849]

Abbreviations

B-CAM: Brief Cognitive Ability Measure
C3Q: Communicating Cognitive Concerns Questionnaire
cART: combination antiretroviral therapy
CB&M: Community Balance and Mobility
CONSORT: Consolidated Standards of Reporting Trials
MOS-HIV: Medical Outcomes Survey for HIV
PLWH: people living with HIV
RCT: randomized controlled trial
REB: Research Ethics Board
REDCap: Research Electronic Data Capture

Protocol

Feasibility of a Sleep Self-Management Intervention in Pregnancy Using a Personalized Health Monitoring Device: Protocol for a Pilot Randomized Controlled Trial

Marquis Hawkins[1,2], PhD; Favorite Iradukunda[3], BS, MS; Mary Paterno[3], PhD

[1]Department of Biostatistics and Epidemiology, University of Massachusetts Amherst, Amherst, MA, United States
[2]Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA, United States
[3]College of Nursing, University of Massachusetts Amherst, Amherst, MA, United States

Corresponding Author:
Marquis Hawkins, PhD
Department of Epidemiology
University of Pittsburgh
130 DeSoto Street
5138 Public Health
Pittsburgh, PA,
United States
Phone: 1 412 383 1931
Email: mah400@pitt.edu

Abstract

Background: Sleep disruptions are common during pregnancy and associated with increased risk of adverse maternal outcomes such as preeclampsia, gestational diabetes, prolonged labor, and cesarean birth. Given the morbidity associated with poor sleep, cost-effective approaches to improving sleep that can be disseminated in community or clinical settings are needed. Personal health monitor (PHM) devices offer an opportunity to promote behavior change, but their acceptability and efficacy at improving sleep in pregnant women are unknown.

Objective: The goal of the paper is to describe the protocol for an ongoing pilot randomized controlled trial that aims to establish the feasibility, acceptability, and preliminary efficacy of using a PHM device (Shine 2, Misfit) to promote sleep during pregnancy.

Methods: The proposed pilot study is a 12-week, parallel arm, randomized controlled trial. Pregnant women, at 24 weeks gestation, will be randomized at a 1:1 ratio to a 12-week sleep education plus PHM device group or a sleep education alone comparison group. The primary outcomes will be measures of feasibility (ie, recruitment, enrollment, adherence) and acceptability (ie, participant satisfaction). The secondary outcomes will be self-reported sleep quality and duration, excessive daytime sleepiness, fatigue, and depressive symptoms.

Results: Recruitment for this study began in September 2017 and ended in March 2018. Data collection for the primary and secondary aims was completed in August 2018. We anticipate that the data analysis for primary and secondary aims will be completed by December 2019. The results from this trial will inform the development of a larger National Institutes of Health grant application to test the efficacy of an enhanced version of the sleep intervention that we plan to submit in the year 2020.

Conclusions: This study will be the first to apply a PHM device as a tool for promoting self-management of sleep among pregnant women. PHM devices have the potential to facilitate behavioral interventions because they include theory-driven, self-regulatory techniques such as behavioral self-monitoring. The results of the study will inform the development of a sleep health intervention for pregnant women.

Trial Registration: ClinicalTrials.gov NCT03783663; https://clinicaltrials.gov/ct2/show/NCT03783663 (Archived by WebCite at http://www.webcitation.org/779Ou8hon)

International Registered Report Identifier (IRRID): DERR1-10.2196/12455

KEYWORDS

eHealth; pregnancy; personal health monitoring; behavior; maternal health

XSL·FO
RenderX

Introduction

Sleep disruptions are common during pregnancy [1] and are associated with an increased risk of adverse maternal/fetal outcomes such as hyperglycemia [2,3], gestational diabetes [4], preeclampsia [5], cesarean delivery [6], longer labor [7], and delivering a low birthweight infant [8]. Additionally, women with poor sleep in pregnancy experience increased depressive symptoms during pregnancy [9] and postpartum depression [10], which increase their risk of depression later in life [11].

While sleep disruptions among pregnant women are often due to physical (eg, pain, nocturia, growing fetus) and hormonal changes [12], lifestyle factors such as sleep hygiene (ie, behavioral and environmental habits that promote or disrupt sleep) also contribute to sleep disruptions. For example, in a convenience sample of 197 pregnant women in their third trimester of pregnancy in Taiwan, poor sleepers were shown to have worse sleep hygiene than good sleepers [13]. Moreover, other health behaviors, such as physical activity and diet, are associated with better sleep duration in pregnancy [14-17]. These studies demonstrate that modifiable factors contribute to sleep disruptions during pregnancy, suggesting that interventions could make a positive impact on sleep in this group. A review of seven nonpharmacological (eg, acupuncture, physical activity, massage) sleep interventions during pregnancy showed trends of improving sleep; however, the studies have generally been of low quality and none focused on sleep hygiene [18]. Given the scarcity of effective approaches to promote sleep among pregnant women, there is a need to develop interventions that can be easily disseminated in community or clinical settings.

Personal health monitors (PHMs) have grown in popularity, with an estimated 1 in 8 consumers now owning a PHM [19]. The popularity of PHM use may provide an opportunity to promote positive health behaviors, such as sleep health, on a large scale. Many PHM devices include theory-driven behavior change techniques such as self-monitoring, goal setting, review of behavioral goals, rewards, and facilitation of comparisons with peers [20]. Several researchers have incorporated the use of PHMs into standard interventions to promote physical activity and weight loss [21-26]. However, there is limited evidence on the efficacy of wearable devices to promote sleep.

Here we describe the protocol for an ongoing pilot randomized controlled trial [NCT03783663] that aims to establish the feasibility, acceptability, and preliminary efficacy of using a PHM device (Shine 2, Misfit) to promote sleep during pregnancy. Specifically, the primary aim of the study will be to (1) establish the feasibility and acceptability of conducting a 12-week intervention for sleep self-management among pregnant women using a PHM device and (2) determine the feasibility of collecting data on sleep and physical activity using a PHM device. The secondary aim will be to determine the preliminary efficacy of the trial on improving self-reported sleep quality and nocturnal sleep duration and decreasing sleep disturbances, excessive daytime sleepiness, fatigue, and depressive symptoms.

Methods

Study Design

This pilot study is a 12-week, parallel arm, randomized controlled trial. Participants will be randomized at a 1:1 ratio to a 12-week sleep education plus PHM device group or a sleep education alone comparison group at approximately 24 weeks gestation of pregnancy (Figure 1). This time period was chosen to capture data during the third trimester of pregnancy when sleep may worsen [1] while allowing sufficient time for study completion prior to birth. We will use blocked randomization to ensure equal group size due to the small pilot sample size. The study has been approved by the institutional review board of the University of Massachusetts Amherst. All participants will be required to provide informed written consent prior to enrollment.

Figure 1. Participant flowchart.
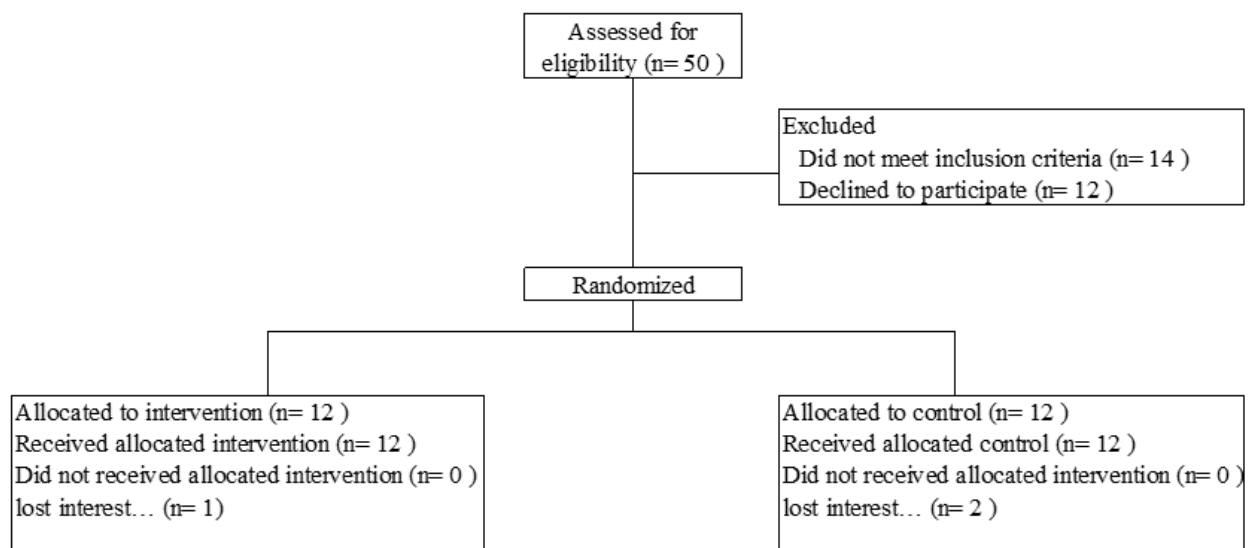
Procedures

At the time of recruitment (t_0) (14 to 24 weeks gestation), participants will complete a screening survey to determine their eligibility. At 24 weeks gestation (t_1), eligible participants will enroll in the study and complete a baseline survey to assess demographic characteristics and secondary outcome measures. Both researchers and participants will be blinded to group assignment at the baseline assessment and during the educational intervention. All participants will receive an individual 30-minute educational intervention on sleep hygiene during pregnancy delivered by a study registered nurse at enrollment. The focus on the sleep education will be to provide tips on how to achieve a better night's sleep during pregnancy such as managing pregnancy weight gain with aerobic activity and keeping a consistent schedule. Participants will also receive a sleep hygiene brochure, adapted from a Massachusetts Lung and Allergy sleep clinic sleep hygiene brochure to include information specific to pregnancy. We piloted the development of the brochure through an online survey that asked about the usefulness of the information and overall aesthetics. We used similar recruitment approaches (ie, Facebook) and eligibility criteria to identify women in the brochure development pilot as will be used for the pilot intervention to ensure the feedback is appropriate for the target population. Based on the participant feedback, we made modifications to the brochure's graphical design; the information content was deemed useful. After completing the baseline survey and sleep hygiene education, participants and researchers will be unblinded to the group assignment (t_2). Assignments will be made prior to the start of recruitment by randomizing study identification (ID) numbers into intervention or control group using Web-based randomization software [27]. The results of the randomization process will be placed into individually sealed envelopes labeled with study ID numbers. At t_2, the study nurse will open the envelope for the ID number assigned to the participant and reveal group assignment. Those in the intervention group will then receive the intervention.

At 36 weeks gestation (t_3), all participants will complete an in-person follow-up survey to assess the same secondary outcome measures as at t_1. Participants will also complete a brief qualitative interview to determine their satisfaction with the sleep hygiene education and, for those in the intervention group, the self-monitoring intervention. Participants will complete a final follow-up phone survey approximately 2 to 4 weeks after delivery (t_4).

Study Population

We will recruit pregnant women who reside within approximately 50 miles of the University of Massachusetts Amherst campus. To be eligible to participate in the study, at the time of recruitment (t_0) women must (1) be aged 18 years or older, (2) be 14 to <24 completed weeks gestation, (3) have no known maternal or fetal complications, (4) have a mobile phone compatible with the study PHM device, (5) have internet access, (6) be English speaking, and (7) be receiving prenatal care. We will exclude women with preexisting diabetes mellitus, hypertension, or a diagnosed sleep disorder, as some research

demonstrated that associations between gestational diabetes and hypertension and a preexisting diagnosed sleep disorder would bias study results [28]. For the purpose of establishing feasibility and acceptability of recruitment and implementing the intervention, the sample size goal is 10 women per group. To account for attrition, we will oversample by 20%, recruiting 12 women per group.

Recruitment will be conducted through advertisements posted (1) at local commercial and community centers frequented by pregnant women in western MA, (2) at a women's health clinic in western MA, (3) at community centers with diverse memberships such as churches, and (4) on Facebook and Craigslist. The study nurse will also recruit participants in person in the waiting room of the women's health clinics at the University of Massachusetts Medical Center. During screening/enrollment, we will document the source of recruitment for each participant.

Intervention

Conceptual Framework

The conceptual framework for the intervention is based on a meta-regression of lifestyle (ie, healthy eating and physical activity) interventions among adults by Michie et al [29]. The authors found that lifestyle interventions incorporating self-monitoring and one other self-regulatory technique derived from control theory (ie, intention formation, setting goals, feedback on performance, review of behavioral goals) were more effective at promoting behavior change (pooled effect size = 0.38; 95% CI 0.27 to 0.49) than studies not including these techniques (pooled effect size = 0.27; 95% CI 0.21 to 0.34). Likewise, in their systematic review and meta-analysis on the efficacy of postpartum physical activity interventions, Gilinsky et al [30] found that efficacious interventions were twice as likely to include self-monitoring than nonefficacious interventions. Guided by this conceptual framework, the sleep intervention will include (1) behavioral self-monitoring (using the Shine 2), (2) goal setting (30-minute sleep education session), (3) feedback on performance (Shine 2), and (4) review of behavioral goals (Shine 2).

Instrumentation

The intervention group will be given a Shine 2 to monitor sleep throughout the 12-week intervention period. The Shine 2 is a triaxial accelerometer that can be worn on the wrist, waist, neck, pocket, or shoe. The Shine 2 measures steps, intensity, energy expenditure, distance traveled, and sleep duration using a proprietary algorithm. The Shine 2 was selected because it can be worn continuously (even during water-based activities), does not require charging (approximate 6-month battery life), and does not require participants to set the device to sleep mode in order to capture sleep patterns, features that may improve adherence to the intervention. The Shine 2 has been shown to provide equivalent estimates of total sleep time as in-lab polysomnography (r=.87), the gold standard for sleep assessment [31].

Participants will be instructed to wear the device on the wrist, which is better for capturing sleep than the other wear locations. Participants will be instructed on how to self-monitor total sleep

time and select a goal for that behavior. Figures 2 and 3 show examples of the feedback participants receive on their sleep relative to their self-selected sleep goals. Further, participants will view feedback on their sleep time daily on the Shine 2 mobile phone app, which they can use to monitor their progress toward achieving their behavioral goals. The Shine 2 syncs with iPhone, Android, and Windows phones. Instructions for syncing the Shine 2s to phones will be included in an instructional brochure designed by the study team. Study-specific Gmail accounts and passwords will be created for each study ID so we can access their sleep and physical activity data. Participants will download the Shine 2 app to their mobile phone and create a Shine 2 account using the assigned study Gmail address. Participants will also download the IFTTT app, a free platform that facilitates downloading and sharing data from a PHM [32]. Participants will activate two IFTTT applets that connect Shine 2 with a Google spreadsheet in Google Drive: (1) "Save your Misfit Shine 2 sleep logs to a Google spreadsheet" and (2) "Document your daily activity summaries." The applets will automatically download the Shine 2 data for sleep daily to the participant's study Gmail account, making it viewable by the research team. This will allow the research team to view participant progress toward meeting their sleep goals. To enhance data security, the study Gmail account will be used only to collect the Shine 2 data (ie, will not be used for email or other purposes), and the Google Drive accounts will be accessed only by study team members using computers that are password-protected and encrypted. Further, no identifying personal information will be contained within the Gmail account. At the end of the study, participants will be instructed to change their email addresses and passwords in their Shine 2 account to a personal email and password so investigators will no longer have access to their Shine 2 accounts. Once all Shine 2 data have been extracted from the study Google Drive accounts, the accounts will be closed. A team member will conduct follow-up calls to intervention participants one week after enrollment and then monthly to address any issues with using the study device and to answer questions.

Figure 2. Misfit Shine 2 dashboard displaying the previous night's sleep duration and self-selected sleep goal.
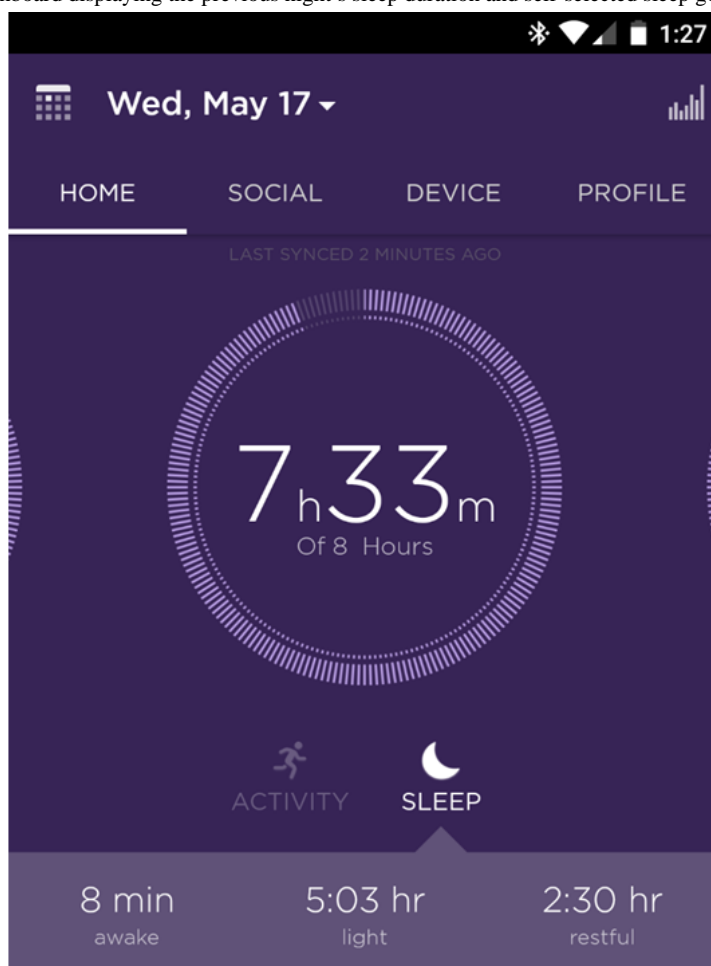
Figure 3. Misfit Shine 2 dashboard displaying weekly trend of sleep duration and self-selected sleep goal.
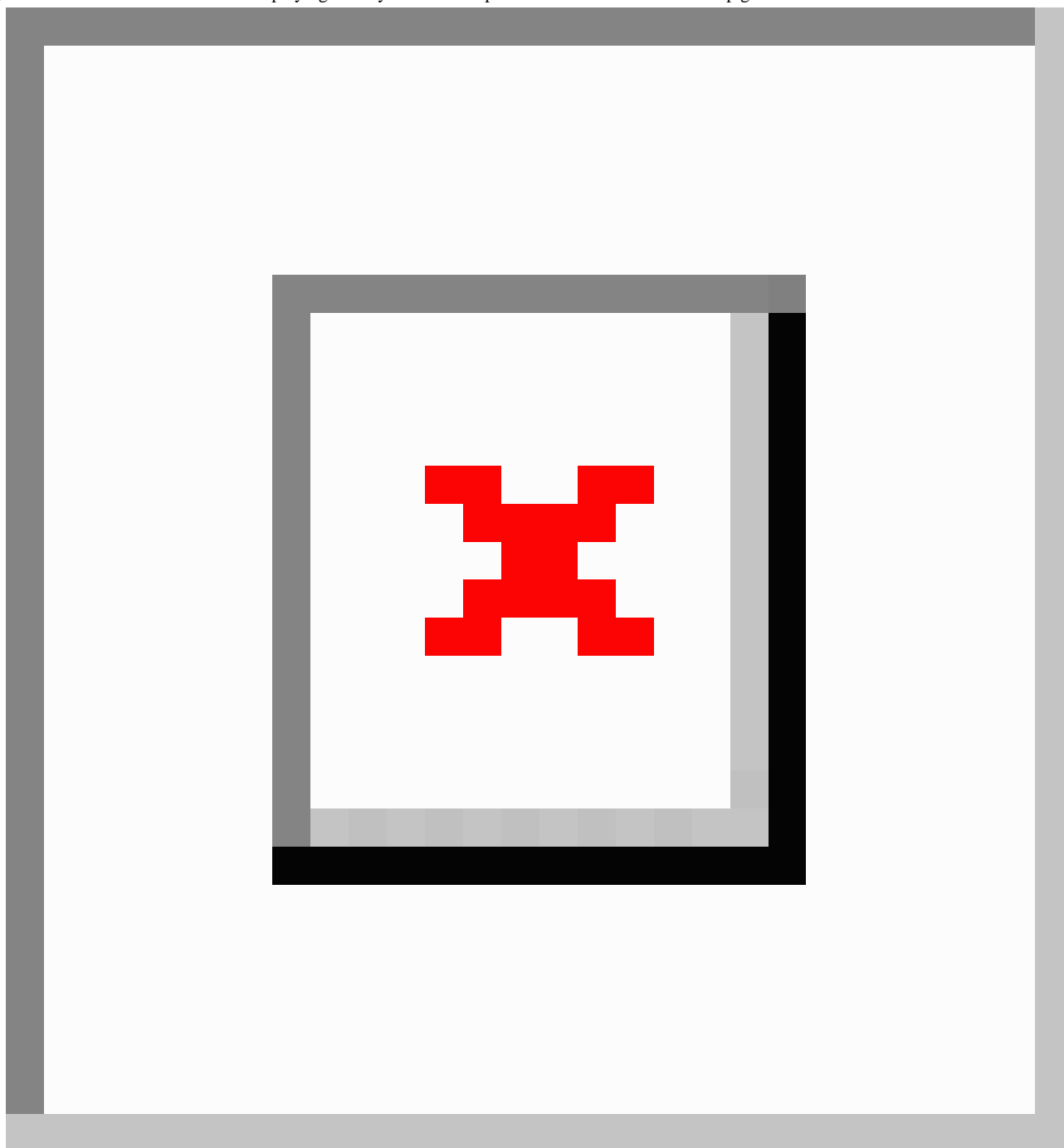
Intervention Fidelity

The sleep hygiene education session will be provided to both groups by the study nurse. To increase fidelity, a specific protocol for delivering the education to each group will be developed by the principal investigator, and the study nurse will be trained by the principal investigator on the protocol. The delivery of education to study participants will be audio recorded with a digital recorder. The principal investigator will listen to each recording to ensure that participants are receiving the same information per the protocol. In the event that deviation from the protocol is noted, the principal investigator and study nurse will meet to review and retrain on the protocol.

Primary Outcomes: Intervention, Feasibility, and Acceptability

Table 1 describes the schedule of study assessments for primary and secondary outcomes measures.

Recruitment

At t_0, participants will be asked how they learned about the study (eg, Facebook, community center), with follow-up questions to gather specific details as necessary. The frequency and percentage of participants recruited by each recruitment method will be calculated. We will present descriptive data on the number screened, eligible, and randomized from each source.

Retention

Retention may be a challenge due to the longitudinal nature of the study. Each participant will be asked to provide multiple contact methods. Increasing incentives will be built into the study at each data collection time point ($20, $30, and $40 gift cards) to encourage retention for both the intervention and control groups. Participants in the intervention arm will keep the Shine 2 device at the study end. Participants in the control arm will be offered a Shine 2 at the study end. The ability to successfully contact participants, deliver the intervention, and collect data at each time point will be recorded to calculate the percentage of participants retained.

To evaluate participant retention with the intervention protocol, the study team will take two actions. First, we will track sleep data weekly for each participant by viewing their sleep log in the study Gmail account. Then, if a lack of data in the sleep log suggests that the participants are not wearing the monitor daily, we will call the participants to find out why they are not using the device, help them problem solve any issues, and encourage them to continue with study adherence.

Table 1. Schedule of study assessments for primary and secondary outcomes.

Characteristics	Assessment			
	Screening/enrollment	Baseline	Active intervention	
	t_0	t_1	t_2	t_3
Primary outcomes				
Recruitment/enrollment				
Eligibility screen	X			
Informed consent	X			
Allocation			X	
Retention				X
Participant satisfaction				X
Secondary outcomes				
Sleep quality and duration		X		X
Sleep disturbances		X		X
Excessive daytime sleepiness		X		X
Fatigue		X		X
Depressive symptoms		X		X

Participant Satisfaction

Participants in both groups will be asked to complete a semistructured interview at the 12-week follow-up data collection meeting (t_3). The interview questions will ask participants about their sleep, barriers and facilitators of sleep, and whether the brief sleep counseling was helpful. Women in the intervention group will additionally be asked about their experience wearing the Shine 2, their likes, dislikes, and whether they believe it would be helpful to other pregnant women.

Secondary Outcomes

Sleep Quality and Duration

The Pittsburgh Sleep Quality Index (PSQI) will be used to measure sleep quality and duration [33]. The PSQI is a 19-item scale measuring perceived sleep quality and disturbance over the past month. For this study, the PSQI will be adapted to measure perceived sleep quality over the past week. The PSQI has been used in numerous studies with pregnant women and has been validated for pregnant women using confirmatory factor analysis and has a reliability of .74 [34]. Scores ≥5 will be considered indicative of poor sleep.

Sleep Disturbances

The PROMIS Short Form v1.0–Sleep Disturbance 6a will be used to measure sleep disturbances [35]. This validated 6-item scale measures individual perceptions of sleep quality and disturbance in the past seven days using a 5-point Likert scale. It is scored by summing responses for all items (two are reverse scored); higher scores indicate higher sleep disturbance.

Excessive Daytime Sleepiness

The Epworth Sleepiness Scale is an 8-item measure of daytime sleepiness [36]. It has been validated for measuring symptoms of daytime sleepiness in pregnant women using principal components and confirmatory factor analysis and has a reliability coefficient of .75 [37]. Higher cumulative scores indicate higher sleepiness.

Fatigue

The PROMIS Fatigue Short Form 4a will be used to measure fatigue. This 4-item scale measures fatigue in the past seven days with a 5-point Likert scale [38]. Responses for all items are summed; higher scores indicate higher fatigue symptoms.

Depressive Symptoms

Two scales will be used to measure depressive symptoms in the past seven days. The PROMIS Depression Short Form 6a is a 6-item scale scored by summing responses for all items, with higher scores indicating higher depressive symptoms. It has been validated for use with several chronic illnesses [39]. Although this scale has not been validated in pregnant women, it is a common data element measurement tool that is supported by the National Institutes of Heath. This will allow us to upload deidentified data gathered with this tool to the National Institute for Nursing Research Common Data Elements Repository. We will additionally measure sleep using the Edinburgh Postnatal Depression Scale, a 10-item scale designed to detect depressive symptoms in postpartum women [40], with support for detection of depressive symptoms in pregnancy [41]. Scores range from 0 to 30; a higher score indicates higher depressive symptoms. We will use scores >12 to indicate depressive symptoms.

Covariates

We will collect information on demographic characteristics and eating habits. Demographic characteristics will include age, ethnic background, racial background, relationship status, level of education, employment status, hours worked per week, and income level. Eating habits will be measured using the Dietary Targets Monitor, a 9-item self-report questionnaire assessing consumption of fruits, vegetables, starchy foods, and other types of foods such as meat and fish [42]. Recommended standard portion for each group will be used to calculate daily, weekly, and monthly consumptions of different foods [42].

Data Analysis

We will use descriptive statistics to determine the feasibility and acceptability of the pilot intervention. Specifically, we describe recruitment and dropout rates in each group and noncompliance with the intervention among the intervention group only. To assess noncompliance we will use data from the IFTTT applet to determine the proportion of days the monitor was worn during the study period. High compliance will be defined as wearing the monitor at least 80% of days during the study period.

Qualitative methods will be used to determine the acceptability of the intervention, using a qualitative descriptive design [43,44]. The audio recorded interviews will be organized using NVivo 11 (QSR International Inc) software. Recordings will be professionally transcribed verbatim, and transcripts will be checked for accuracy. Descriptive coding will be used to identify and link comparable content and categorize data at a basic level [43]. For further analysis of descriptive codes, content analysis will be used to describe patterns and summarize findings in the data [44]. Memos will be written throughout the analytic process to capture the analysis process and the researchers' responses to the data. Trustworthiness will be determined through peer debriefing between the research assistant and the principal investigator and the audit trial of documentation of the analysis, including coding, memos, and findings.

To examine the preliminary efficacy of the intervention, all secondary outcome variables will be modeled as continuous outcomes. We will compare baseline characteristics of the intervention groups to ensure randomization was successful. We will present the mean and standard deviation or median and interquartile range for each secondary outcome at baseline and postintervention. A paired t test will be used to compare baseline and postintervention means separately in each group. A Student t test will then compare the change in each measure between the two intervention groups. Finally, we will calculate within and between-group effect sizes using Cohen d ($M_{\Delta I} - M_{\Delta C}$ / $SD_{pooled\ I\ and\ C}$; $M_{t2} - M_{t3}$ / $SD_{pooled\ t2\ and\ t3}$).

Results

Recruitment for this study began in September 2017 and ended in March 2018. Data collection for the primary and secondary aims was completed in August 2018. We anticipate that the data analysis for primary aims, evaluating the feasibility and acceptability of the intervention, will be completed by July 2019. We anticipate that the data analysis on the secondary aims, examining the preliminary efficacy of the trial on sleep quality and duration, sleep disturbance, excessive daytime sleepiness, fatigue, and depressive symptoms, will be completed by December 2019. The results from this trial will inform the development of a larger National Institutes of Health grant application to test the efficacy of an enhanced version of the sleep intervention that we plan to submit in the year 2020.

Discussion

Principal Findings

This 12-week, parallel arm, pilot randomized controlled trial proposes to establish the feasibility and acceptability of conducting a 12-week intervention for sleep self-management among pregnant women using a PHM device and determine the feasibility of collecting data on sleep and physical activity using a PHM device. Secondarily, the trial will determine the preliminary efficacy of the trial on improving self-reported sleep quality and nocturnal sleep duration and decreasing sleep disturbances, excessive daytime sleepiness, fatigue, and depressive symptoms. This study is novel because it will be the first to apply a PHM device as a tool for promoting self-management of sleep among pregnant women. We anticipate that at least 80% of participants will wear the PHM device daily throughout the study period and report high satisfaction with the intervention. Further, we anticipate that participant feedback will inform the design of a larger randomized controlled trial aiming to improve sleep quality and duration during pregnancy using a PHM device.

Specifically, findings from the pilot study will inform the recruitment and intervention implementation, including strategies to increase adherence. As far as refining our recruitment plan, we will learn important information about the demographic characteristics (eg, race and ethnicity, socioeconomic status, baseline sleep characteristics) of the sample population. Our recruitment goal is for at least 20% of the sample to come from racial and ethnic minority groups. If we fail to recruit a racial and ethnically diverse sample, we will identify specific recruitment barriers and refine our recruitment strategy. Second, we will identify which recruitment approach

(eg, Facebook, flyering at local clinics) is the most cost effective. Third, identifying weekly and monthly recruitment yields will allow us to determine if our recruitment strategy could feasibly enroll participants for an adequately powered study. We will need to consider the length of the recruitment period, adding clinical sites, and budgetary considerations for the larger trial.

In addition to recruitment, participant adherence is another critical challenge this feasibility study will help inform. Since previous studies did not use a PHM device specifically to monitor sleep, we would like to first address if pregnant women will wear them for this purpose. Through the IFTTT app, we will be able to identify when the monitor is not worn and elicit feedback from participants on any barriers to wearing the device. If specific benchmarks for retention are not met (>80%), we will refine our current approach (eg, incentive, engagement strategies) to address any challenges that exist.

Last, this feasibility trial will provide important information about the implementation of the intervention. The intervention was meant to be a low-contact intervention to be consistent with the limited time resources in a clinic setting. Through the participant satisfaction interview, we will determine whether the initial counseling session and sleep brochure address challenges specific to pregnant women. Concerning the instruments for the intervention (Shine 2), we will learn if there are specific features participants liked or did not like about the device and the extent to which participants used the device to self-monitor their sleep. Identifying specific features will allow us to be responsive to rapid changes in technology, steering us to select a device that contains features participants find particularly helpful. More generally, we will learn if wearing the PMH device serves as a motivator to participants to monitor their sleep more closely and identify behaviors affecting their sleep.

Given the prevalence and morbidity of sleep disturbances during pregnancy, it is essential to identify cost-effective approaches to promote better sleep in this group. Wearable devices have the potential to facilitate behavioral interventions as they include theory-driven self-regulatory techniques such as behavioral self-monitoring. There has been increasing interest in incorporating PMH devices into clinical settings. For example, PHM devices have been used to monitor inpatient recovery [45]. In an outpatient care setting, PHM devices can be used to track health behaviors remotely and monitor progress toward meeting patient-centered goals around sleep [46,47]. Data from PHM devices can be integrated into electronic medical records. For the patient, tracking behavior more closely can help the patient identify patterns and make choices to change behavior to improve sleep [48]. No prior studies have used a PHM to promote sleep during pregnancy; therefore, novel intervention strategies need to be developed and refined.

Limitations

This study will, however, have important limitations that warrant discussion. First, the study requires that participants have access to mobile phones compatible with the study monitor; therefore, individuals of low socioeconomic status may potentially be excluded from the study [49]. However, lower income Americans have made gains in technology adoption including the use of mobile phones, which is likely to continue in the future. In addition, since there are multiple components of the intervention (ie, PHM and follow-up calls), we will not be able to discern which component explains differences between the groups, if any are observed. We will keep the follow-up calls briefs and use them primarily as a method of collecting information on the study's feasibility throughout. However, we won't be able to remove the potential effects repeated contacts can have on a participant's ability to change behavior. Third, we will not exclude participants based on their use of any PHM. Therefore, control group participants may have access to a PHM during the study. However, none of the participants will be using our specific study monitor at baseline. Further, participants in the control group will receive study monitors at the end of the study to reduce attrition and limit the use of PHMs during the study. Last, due to the small sample size, we will not have sufficient power to detect any group differences in the secondary outcomes. Therefore, any significant or nonsignificant findings must be interpreted with caution. We will, however, be able to observed trends and generate hypotheses to be tested in a fully powered study.

Conclusion

This pilot feasibility study will provide the foundation for a larger trial that aims to improve sleep during pregnancy using a PHM device. The findings from this feasibility and acceptability pilot will provide valuable feedback on the design and implementation of the intervention.

Acknowledgments

Authors' Contributions

All the authors contributed to the various stages of this study. MP conceived of the study design, researched the literature, and provided feedback on all drafts of the manuscript. MH participated in the design of the study, researched the literature, and drafted the manuscript. FI contributed to study design and brochure development. All the authors read and commented on the drafts and approved of the final version of the manuscript for submission.

Conflicts of Interest

None declared.

XSL•FO

RenderX

References

1. Facco FL, Kramer J, Ho KH, Zee PC, Grobman WA. Sleep disturbances in pregnancy. Obstet Gynecol 2010 Jan;115(1):77-83. [doi: 10.1097/AOG.0b013e3181c4f8ec] [Medline: 20027038]

2. Herring SJ, Nelson DB, Pien GW, Homko C, Goetzl LM, Davey A, et al. Objectively measured sleep duration and hyperglycemia in pregnancy. Sleep Med 2014 Jan;15(1):51-55 [FREE Full text] [doi: 10.1016/j.sleep.2013.07.018] [Medline: 24239498]

3. Reutrakul S, Anothaisintawee T, Herring SJ, Balserak BI, Marc I, Thakkinstian A. Short sleep duration and hyperglycemia in pregnancy: aggregate and individual patient data meta-analysis. Sleep Med Rev 2018 Dec;40:31-42. [doi: 10.1016/j.smrv.2017.09.003] [Medline: 29103944]

4. Cai S, Tan S, Gluckman PD, Godfrey KM, Saw S, Teoh OH, GUSTO Study Group. Sleep quality and nocturnal sleep duration in pregnancy and risk of gestational diabetes mellitus. Sleep 2017 Feb 01;40(2). [doi: 10.1093/sleep/zsw058] [Medline: 28364489]

5. Williams MA, Miller RS, Qiu C, Cripe SM, Gelaye B, Enquobahrie D. Associations of early pregnancy sleep duration with trimester-specific blood pressures and hypertensive disorders in pregnancy. Sleep 2010 Oct;33(10):1363-1371 [FREE Full text] [Medline: 21061859]

6. Li R, Zhang J, Zhou R, Liu J, Dai Z, Liu D, et al. Sleep disturbances during pregnancy are associated with cesarean delivery and preterm birth. J Matern Fetal Neonatal Med 2017 Mar;30(6):733-738. [doi: 10.1080/14767058.2016.1183637] [Medline: 27125889]

7. Lee KA, Gay CL. Sleep in late pregnancy predicts length of labor and type of delivery. Am J Obstet Gynecol 2004 Dec;191(6):2041-2046. [doi: 10.1016/j.ajog.2004.05.086] [Medline: 15592289]

8. Abeysena C, Jayawardana P, Seneviratne RDA. Effect of psychosocial stress and physical activity on low birthweight: a cohort study. J Obstet Gynaecol Res 2010 Apr;36(2):296-303. [doi: 10.1111/j.1447-0756.2009.01155.x] [Medline: 20492380]

9. Ahmed AH, Hui S, Crodian J, Plaut K, Haas D, Zhang L, et al. Relationship between sleep quality, depression symptoms, and blood glucose in pregnant women. West J Nurs Res 2018 Nov 08:193945918809714. [doi: 10.1177/0193945918809714] [Medline: 30406728]

10. Plancoulaine S, Flori S, Bat-Pitault F, Patural H, Lin J, Franco P. Sleep trajectories among pregnant women and the impact on outcomes: a population-based cohort study. Matern Child Health J 2017 Dec;21(5):1139-1146. [doi: 10.1007/s10995-016-2212-9] [Medline: 28083730]

11. Rasmussen MH, Strøm M, Wohlfahrt J, Videbech P, Melbye M. Risk, treatment duration, and recurrence risk of postpartum affective disorder in women with no prior psychiatric history: a population-based cohort study. PLoS Med 2017 Sep;14(9):e1002392 [FREE Full text] [doi: 10.1371/journal.pmed.1002392] [Medline: 28949960]

12. Moline M, Broch L, Zak R. Sleep problems across the life cycle in women. Curr Treat Options Neurol 2004 Jul;6(4):319-330. [Medline: 15157409]

13. Tsai S, Lee C, Wu W, Landis CA. Sleep hygiene and sleep quality of third-trimester pregnant women. Res Nurs Health 2016 Feb;39(1):57-65. [doi: 10.1002/nur.21705] [Medline: 26650922]

14. Bazargan M. Self-reported sleep disturbance among African-American elderly: the effects of depression, health status, exercise, and social support. Int J Aging Hum Dev 1996;42(2):143-160. [doi: 10.2190/GM89-NRTY-DERQ-LC7D] [Medline: 8859496]

15. Baker JH, Rothenberger SD, Kline CE, Okun ML. Exercise during early pregnancy is associated with greater sleep continuity. Behav Sleep Med 2018;16(5):482-493. [doi: 10.1080/15402002.2016.1228649] [Medline: 27739877]

16. Kolu P, Raitanen J, Luoto R. Physical activity and health-related quality of life during pregnancy: a secondary analysis of a cluster-randomised trial. Matern Child Health J 2014 Nov;18(9):2098-2105. [doi: 10.1007/s10995-014-1457-4] [Medline: 24585400]

17. van Lee L, Chia A, Loy SL, Colega M, Tham EKH, Cai S, et al. Sleep and dietary patterns in pregnancy: findings from the GUSTO cohort. Int J Environ Res Public Health 2017 Dec 17;14(11) [FREE Full text] [doi: 10.3390/ijerph14111409] [Medline: 29149071]

18. Hollenbach D, Broker R, Herlehy S, Stuber K. Non-pharmacological interventions for sleep quality and insomnia during pregnancy: a systematic review. J Can Chiropr Assoc 2013 Sep;57(3):260-270 [FREE Full text] [Medline: 23997252]

19. Omura JD, Carlson SA, Paul P, Watson KB, Fulton JE. National physical activity surveillance: Users of wearable activity monitors as a potential data source. Prev Med Rep 2017 Mar;5:124-126 [FREE Full text] [doi: 10.1016/j.pmedr.2016.10.014] [Medline: 28101443]

20. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]

21. Polzien KM, Jakicic JM, Tate DF, Otto AD. The efficacy of a technology-based system in a short-term behavioral weight loss intervention. Obesity (Silver Spring) 2007 Apr;15(4):825-830. [doi: 10.1038/oby.2007.584] [Medline: 17426316]

22. Pellegrini CA, Verba SD, Otto AD, Helsel DL, Davis KK, Jakicic JM. The comparison of a technology-based system and an in-person behavioral weight loss intervention. Obesity (Silver Spring) 2012 Feb;20(2):356-363 [FREE Full text] [doi: 10.1038/oby.2011.13] [Medline: 21311506]

XSL·FO

RenderX

23. Unick JL, O'Leary KC, Bond DS, Wing RR. Physical activity enhancement to a behavioral weight loss program for severely obese individuals: a preliminary investigation. ISRN Obes 2012 Sep 05;2012 [FREE Full text] [doi: 10.5402/2012/465158] [Medline: 24379985]

24. Wang JB, Cadmus-Bertram LA, Natarajan L, White MM, Madanat H, Nichols JF, et al. Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed J E Health 2015 Oct;21(10):782-792. [doi: 10.1089/tmj.2014.0176] [Medline: 26431257]

25. Cadmus-Bertram LA, Marcus BH, Patterson RE, Parker BA, Morey BL. Randomized trial of a Fitbit-Based physical activity intervention for women. Am J Prev Med 2015 Sep;49(3):414-418. [doi: 10.1016/j.amepre.2015.01.020] [Medline: 26071863]

26. Thompson WG, Kuhle CL, Koepp GA, McCrady-Spitzer SK, Levine JA. "Go4Life" exercise counseling, accelerometer feedback, and activity levels in older people. Arch Gerontol Geriatr 2014;58(3):314-319. [doi: 10.1016/j.archger.2014.01.004] [Medline: 24485546]

27. Random integer generator. URL: https://www.random.org/ [accessed 2019-04-01] [WebCite Cache ID 77JXgIHvR]

28. August EM, Salihu HM, Biroscak BJ, Rahman S, Bruder K, Whiteman VE. Systematic review on sleep disorders and obstetric outcomes: scope of current knowledge. Am J Perinatol 2013 Apr;30(4):323-334. [doi: 10.1055/s-0032-1324703] [Medline: 22893551]

29. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

30. Gilinsky AS, Dale H, Robinson C, Hughes AR, McInnes R, Lavallee D. Efficacy of physical activity interventions in post-natal populations: systematic review, meta-analysis and content coding of behaviour change techniques. Health Psychol Rev 2015;9(2):244-263. [doi: 10.1080/17437199.2014.899059] [Medline: 26209211]

31. Mantua J, Gravel N, Spencer RMC. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 Dec 05;16(5) [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]

32. IFTTT. URL: https://ifttt.com/ [accessed 2019-04-01] [WebCite Cache ID 77JXrCtgj]

33. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [Medline: 2748771]

34. Qiu C, Gelaye B, Zhong Q, Enquobahrie DA, Frederick IO, Williams MA. Construct validity and factor structure of the Pittsburgh Sleep Quality Index among pregnant women in a Pacific-Northwest cohort. Sleep Breath 2016 Mar;20(1):293-301 [FREE Full text] [doi: 10.1007/s11325-016-1313-4] [Medline: 26810497]

35. Buysse DJ, Yu L, Moul DE, Germain A, Stover A, Dodds NE, et al. Development and validation of patient-reported outcome measures for sleep disturbance and sleep-related impairments. Sleep 2010 Jun;33(6):781-792 [FREE Full text] [Medline: 20550019]

36. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991 Dec;14(6):540-545. [Medline: 1798888]

37. Baumgartel KL, Terhorst L, Conley YP, Roberts JM. Psychometric evaluation of the Epworth sleepiness scale in an obstetric population. Sleep Med 2013 Jan;14(1):116-121 [FREE Full text] [doi: 10.1016/j.sleep.2012.10.007] [Medline: 23218535]

38. Fatigue: a brief guide to the PROMISE Fatigue instruments. URL: http://www.healthmeasures.net/administrator/components/com_instruments/uploads/PROMIS%20Fatigue%20Scoring%20Manual%2011042016.pdf [accessed 2019-04-01] [WebCite Cache ID 77JY9Vvt9]

39. Schalet BD, Pilkonis PA, Yu L, Dodds N, Johnston KL, Yount S, et al. Clinical validity of PROMIS Depression, Anxiety, and Anger across diverse clinical samples. J Clin Epidemiol 2016 Dec;73:119-127 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.08.036] [Medline: 26931289]

40. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry 1987 Jun;150:782-786. [Medline: 3651732]

41. Kozinszky Z, Dudas RB. Validation studies of the Edinburgh Postnatal Depression Scale for the antenatal period. J Affect Disord 2015 May 01;176:95-105. [doi: 10.1016/j.jad.2015.01.044] [Medline: 25704562]

42. Lean MEJ, Anderson AS, Morrison C, Currall J. Evaluation of a dietary targets monitor. Eur J Clin Nutr 2003 May;57(5):667-673. [doi: 10.1038/sj.ejcn.1601596] [Medline: 12771967]

43. Saldana J. The Coding Manual for Qualitative Researchers. 3rd Edition. Thousand Oaks: Sage Publications; 2016.

44. Thomas E, Magilvy JK. Qualitative rigor or research validity in qualitative research. J Spec Pediatr Nurs 2011 Apr;16(2):151-155. [doi: 10.1111/j.1744-6155.2011.00283.x] [Medline: 21439005]

45. Low CA, Bovbjerg DH, Ahrendt S, Choudry MH, Holtzman M, Jones HL, et al. Fitbit step counts during inpatient recovery from cancer surgery as a predictor of readmission. Ann Behav Med 2018 Jan 05;52(1):88-92 [FREE Full text] [doi: 10.1093/abm/kax022] [Medline: 29538623]

46. Izmailova ES, Wagner JA, Perakslis ED. Wearable devices in clinical trials: hype and hypothesis. Clin Pharmacol Ther 2018 Jul;104(1):42-52 [FREE Full text] [doi: 10.1002/cpt.966] [Medline: 29205294]

47. Shelgikar AV, Anderson PF, Stephens MR. Sleep tracking, wearable technology, and opportunities for research and clinical care. Chest 2016 Dec;150(3):732-743. [doi: 10.1016/j.chest.2016.04.016] [Medline: 27132701]

48.   Bianchi MT. Sleep devices: wearables and nearables, informational and interventional, consumer and clinical. Metabolism 2018 Dec;84:99-108. [doi: 10.1016/j.metabol.2017.10.008] [Medline: 29080814]

49.   Anderson M. Digital divide persists even as lower-income Americans make gains in tech adoption. Washington: Pew Internet and American Life Project; 2017 Mar 22. URL: https://www.pewresearch.org/fact-tank/2017/03/22/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/ [accessed 2019-04-01] [WebCite Cache ID 77JYJ0vAZ]

Abbreviations

ID: identification
PHM: personal health monitor
PSQI: Pittsburgh Sleep Quality Index

XSL•FO
RenderX

Protocol

Gait Characteristics in Patients With Ankylosing Spondylitis: Protocol for a Systematic Review

Julie Soulard[1,2], MSc; Nicolas Vuillerme[1,3], PhD; Jacques Vaillant[1,2], PhD

[1]AGEIS, Université Grenoble Alpes, Grenoble, France

[2]Grenoble Alpes University Hospital, Grenoble, France

[3]Institut Universitaire de France, Paris, France

Corresponding Author:
Julie Soulard, MSc
AGEIS
Université Grenoble Alpes
Faculty of Medicine
Grenoble,
France
Phone: 33 476637104
Email: juliesoulard.physio@gmail.com

Abstract

Background: Ankylosing spondylitis is a subtype of inflammatory rheumatic disease, affecting predominantly the axial skeleton and sacroiliac joints. The main clinical manifestations are spinal stiffness and inflammatory back pain, which can potentially affect gait ability of patients with ankylosing spondylitis. However, published studies show discrepancies regarding gait characteristics in ankylosing spondylitis and heterogeneity in terms of task requirement, types of equipment, data collection, and analysis techniques used to assess gait ability of patients with ankylosing spondylitis.

Objective: This review aimed to determine (1) the consequences of ankylosing spondylitis on gait and (2) how gait is assessed in patients with ankylosing spondylitis.

Methods: Three electronic databases—PubMed, Physiotherapy Evidence Database (PEDro), and Cochrane—were searched systematically with no limit on the publication date in order to identify studies satisfying the search criteria. The research focused on original research, using Boolean operators "AND" and "OR" in the combination of the Medical Subject Headings descriptors found in titles or abstracts: (Gait OR Walk OR Walking OR locomotor OR locomotion) AND (ankylosing spondylitis OR spondyloarthritis). Only English-language original articles were included.

Results: As of September 2018, the search was completed, and 168 records were obtained. After screening titles and abstracts, 19 full texts were reviewed. Of those, 17 were included in the review. We are currently in the process of data extraction and synthesis.

Conclusions: The systematic review will provide a synthesis and comprehensive evaluation of published studies on gait characteristics in patients with ankylosing spondylitis. This work is also intended to help identify the likely relevant directions for future research.

Trial Registration: PROSPERO CRD42018102540; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=102540

International Registered Report Identifier (IRRID): DERR1-10.2196/12470

Introduction

Ankylosing spondylitis is a subtype of inflammatory rheumatic disease, predominantly affecting the axial skeleton and sacroiliac joints. Ankylosing spondylitis is associated with inflammation or new bone formation, with syndesmophytes and ankyloses visible on radiographs [1]. The main clinical manifestations are spinal stiffness and inflammatory back pain, which yield adverse effects on work ability, work productivity, quality of life, and psychological well-being [2,3]. Gait ability is known to

XSL·FO
RenderX

contribute to functional independence and quality of life in patients with stroke [4,5] or following hip arthroplasty [6] and is impaired in patients with ankylosing spondylitis [7]. Indeed, considering the decreased range of movement, pain, and altered posture [8,9] associated with ankylosing spondylitis, previous works have reported that ankylosing spondylitis leads to more cautious gait pattern, shorter stride length, and decreased range of motion at the hip and knee joints [10-12]. However, it is still unclear whether and how gait is modified in patients with ankylosing spondylitis. Previous studies reported that patients with ankylosing spondylitis covered significantly lesser distance than controls during the Six-Minute Walk test [13] and adopted a shorter stride length [12], whereas in other published studies, no statistically significant group differences were reported for the same gait-related parameters [10,11,14]. Furthermore, published studies showed heterogeneity in terms of task requirement, types of equipment, data collection, and analysis techniques used to assess gait ability of patients with ankylosing spondylitis. For instance, gait-related studies encompass an increasingly large variety of tasks, types of equipment, and analysis techniques including, for example, both clinical (eg, Timed-Up-and-Go test and Six-Minute Walk test) and laboratory measurements (kinetic, kinematic, or electromyographic gait analysis), which should be taken into consideration for the assessment of gait ability in patients with ankylosing spondylitis. Thus, this systematic review aims to document the effect of ankylosing spondylitis in gait, specifically focusing on published studies that have reported clinical or laboratory gait measurements in patients with ankylosing spondylitis. More specifically, this review aimed to determine the consequences of ankylosing spondylitis on gait and how gait is assessed in patients with ankylosing spondylitis.

Methods

This protocol has been registered in PROSPERO (CRD42018102540). We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement guidelines provided by Moher et al when conducting our systematic review and meta-analysis [15].

Inclusion Criteria

Original quantitative and qualitative research studies that assessed gait in patients with ankylosing spondylitis were included. To be eligible for inclusion, studies had to be published in English in peer-reviewed scientific journals.

Type of Participants

Studies were included if participants were older than 18 years, with a diagnosis of ankylosing spondylitis.

Type of Outcome Measurements

Studies were included if they reported clinical or laboratory gait measurements.

Type of Studies

Observational and experimental study designs were included.

Exclusion Criteria

The following types of studies were ineligible: case reports, abstracts, editorials, conference abstracts, letters to the editor, reviews, and meta-analysis.

We also excluded studies that reported gait outcomes inadequately (without mean and SD, or median associated with interquartile range or first and third quartiles) or those from which it was not possible to extract data from the results section.

Data Sources and Search Strategy

A computer-aided literature search was conducted in the following electronic databases on June 5, 2018, with no date restrictions: PubMed, Physiotherapy Evidence Database (PEDro), Cochrane library.

Consistent with a similar review, search terms included those related to population, ankylosing spondylitis [16], and the outcome—gait [17]. The search strategy included a combination of the following keywords and Medical Subject Headings terms found in the abstract or title: ("gait" OR "walk" OR "walking" OR "locomotor" OR "locomotion") AND ("ankylosing spondylitis" OR "spondyloarthritis").

Study Selection

Two reviewers independently screened the titles, abstracts, and keywords identified by the search strategy in order to select potentially relevant studies.

After this initial search, full-length texts of the identified potentially relevant studies were obtained. Based on the above mentioned inclusion and exclusion criteria, the two reviewers further screened these full texts to elucidate their eligibility and decide on their inclusion. In case of any disagreement, consensus was reached through discussions between the two reviewers. If no consensus was achieved between the two reviewers, a third reviewer was contacted.

Risk of Bias in Individual Studies

As our aim is not to evaluate the effect of an intervention, we did not use a risk-of-bias assessment. As mentioned above, our aim was to document the effect of ankylosing spondylitis on gait, specifically focusing on published studies that have reported clinical or laboratory gait measurements in patients with ankylosing spondylitis.

Data Extraction

Following the PRISMA guidelines [15], a flow chart of the selection process was created, with the number of citations reviewed at each stage of the review (Figure 1). Additionally, the following four sets of data will be extracted from the retrieved articles [18]:

- Study characteristics: first author(s), title, year of publication, journal name, and country
- Sample description: sample size, age, gender, weight, height, body mass index, health status, disease duration, functional status measurements, level of pain, description of radiographic damage, biologic medications, Bath ankylosing spondylitis functional index, Bath ankylosing

spondylitis disease activity index, and Bath ankylosing spondylitis metrology index

- Methods: task requirement, data acquisition methodology and instrumentation, and parameters assessed
- Main results obtained from gait assessment: clinical measurements of gait (Six-Minute Walk distance and time to complete the Timed-Up-and-Go test) and laboratory measurements of gait such as spatiotemporal parameters (gait speed, stride length, stride time, and cadence) and

kinematic parameters (continuous estimate of relative phase, joint range of motion, and joint moments)

Means and SDs or medians associated with interquartile range or the first and third quartiles will be extracted. Two reviewers will independently extract these data from each enrolled study and compare the data for consistency. Any discrepancies between the two reviewers will be resolved at a consensus meeting. If disagreement persists, a third reviewer will be consulted to achieve a final judgment.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow chart of the selection process.
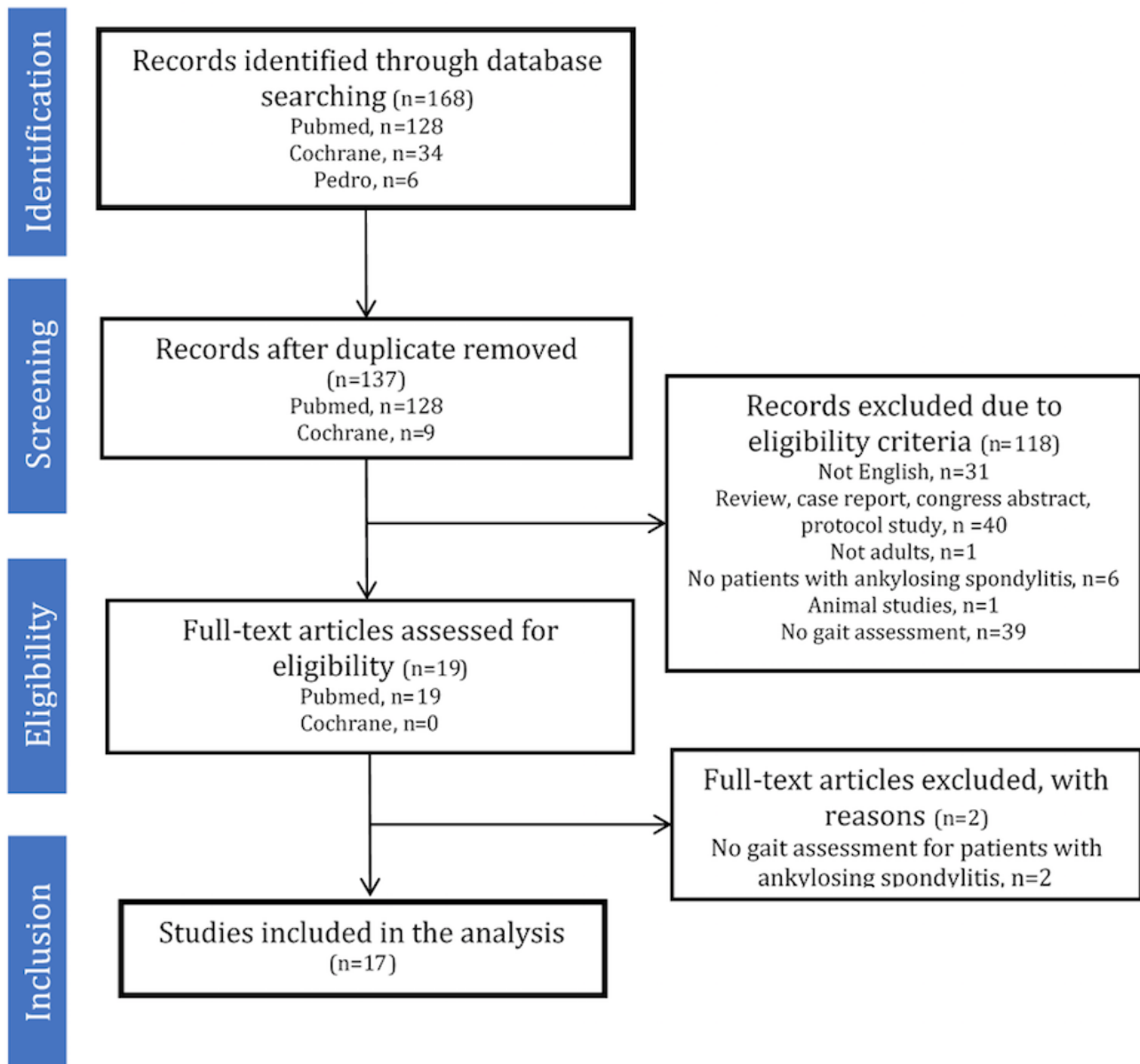
Results

As of September 2018, we have completed the search strategy and obtained 168 records. After screening titles and abstracts, 19 full texts were reviewed. Of those, 17 were included in the review (Figure 1). We are currently in the process of data extraction and synthesis. We expect the final results to be submitted for publication in March 2019.

Discussion

Considering the importance of gait in functional independence and quality of life, there are a growing number of reviews and reports examining gait characteristics in a wide range of populations [19] with neurological disorders [20-22], hip osteoarthritis [23], diabetes [24], frailty [25], or dementia [26] and in older adults [17]. However, as of September 2018, only one review published in 2015 focused on gait characteristics in

rheumatologic patients [7], with only 3 studies reporting results of patients with ankylosing spondylitis [10-12]. Interestingly, the abovementioned review [7] focused on case-control studies only (ie, "studies were included...if they were articles that included a healthy group as means of comparison" [7]). This review did not include studies with clinical measurements of gait, but only studies reporting laboratory measurements (ie, "studies were included...if they reported spatiotemporal, kinematic, kinetic, peak plantar pressure or muscle activity data during gait" [7]). Thus, an update of the published literature is needed.

A strength of this review protocol is that it includes both clinical and laboratory measurements of gait studies on patients with ankylosing spondylitis, reporting precisely the methodology used in each selected study, as recommended by the PRISMA statement (e.g. "how the data was collected and analysed" [15]). Indeed, early identification of gait deficits in patients with ankylosing spondylitis could help us better understand, follow, and predict disease evolution and allow for timely implementation of targeted interventions or treatment to improve gait. Keywords have been chosen based on latest reviews on

ankylosing spondylitis [16] and gait [17] separately and were searched in principal databases, assuring the conduct of a systematic review.

However, there are some limitations related to this review that need to be addressed. We assume that the selection and qualitative synthesis of the eligible studies are a subjective process. However, we will seek to minimize this limitation by duplicating our search and having two reviewers conduct the screening process independently [15]. We plan to present the results of this systematic review at international scientific and clinical conferences and publish them in a peer-reviewed scientific journal. The systematic review will provide a synthesis and comprehensive evaluation of published research on gait characteristics in patients with ankylosing spondylitis. Largely, this work is further intended to help identify the likely relevant directions for future research. For instance, from a clinical perspective, we support the idea that an objective and standardized assessment of gait characteristics should be an integral part of every comprehensive assessment of patients with ankylosing spondylitis.

Authors' Contributions

JS, NV, and JV designed the systematic review protocol. JS prepared the first draft. NV and JV reviewed and revised the first draft. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Braun J, Sieper J. Ankylosing spondylitis. Lancet 2007 Apr 21;369(9570):1379-1390. [doi: 10.1016/S0140-6736(07)60635-7] [Medline: 17448825]
2. Sağ S, Nas K, Sağ MS, Tekeoğlu I, Kamanlı A. Relationship of work disability between the disease activity, depression and quality of life in patients with ankylosing spondylitis. J Back Musculoskelet Rehabil 2018;31(3):499-505. [doi: 10.3233/BMR-169657] [Medline: 29504521]
3. Sieper J, Braun J, Rudwaleit M, Boonen A, Zink A. Ankylosing spondylitis: an overview. Ann Rheum Dis 2002 Dec;61 Suppl 3:iii8-ii18 [FREE Full text] [Medline: 12381506]
4. Price R, Choy NL. Investigating the relationship of the functional gait assessment to spatiotemporal parameters of gait and quality of life of individuals with stroke. J Geriatr Phys Ther 2018 Jan 10. [doi: 10.1519/JPT.0000000000000173] [Medline: 29324509]
5. Khanittanuphong P, Tipchatyotin S. Correlation of the gait speed with the quality of life and the quality of life classified according to speed-based community ambulation in Thai stroke survivors. NeuroRehabilitation 2017;41(1):135-141. [doi: 10.3233/NRE-171465] [Medline: 28527227]
6. Sliwinski M, Sisto S. Gait, quality of life, and their association following total hip arthroplasty. J Geriatr Phys Ther 2006;29(1):10-17. [Medline: 16630370]
7. Carroll M, Parmar P, Dalbeth N, Boocock M, Rome K. Gait characteristics associated with the foot and ankle in inflammatory arthritis: a systematic review and meta-analysis. BMC Musculoskelet Disord 2015 Jun 05;16:134 [FREE Full text] [doi: 10.1186/s12891-015-0596-0] [Medline: 26044780]

8.   Aydog E, Depedibi R, Bal A, Eksioglu E, Unlü E, Cakci A. Dynamic postural balance in ankylosing spondylitis patients. Rheumatology (Oxford) 2006 Apr;45(4):445-448. [doi: 10.1093/rheumatology/kei192] [Medline: 16278280]

9.   Bot SD, Caspers M, Van RBJ, Toussaint HM, Kingma I. Biomechanical analysis of posture in patients with spinal kyphosis due to ankylosing spondylitis: a pilot study. Rheumatology (Oxford) 1999 May;38(5):441-443. [Medline: 10371282]

10. Del DS, Carraro E, Sawacha Z, Guiotto A, Bonaldo L, Masiero S, et al. Impaired gait in ankylosing spondylitis. Med Biol Eng Comput 2011 Jul;49(7):801-809. [doi: 10.1007/s11517-010-0731-x] [Medline: 21229328]

11. Mangone M, Scettri P, Paoloni M, Procaccianti R, Spadaro A, Santilli V. Pelvis-shoulder coordination during level walking in patients with ankylosing spondylitis. Gait Posture 2011 May;34(1):1-5. [doi: 10.1016/j.gaitpost.2011.02.002] [Medline: 21377366]

12. Zebouni L, Helliwell PS, Howe A, Wright V. Gait analysis in ankylosing spondylitis. Ann Rheum Dis 1992 Jul;51(7):898-899 [FREE Full text] [Medline: 1632666]

13. Brambila-Tapia AJL, Rocha-Muñoz AD, Gonzalez-Lopez L, Vázquez-Del-Mercado M, Salazar-Páramo M, Dávalos-Rodríguez IP, et al. Pulmonary function in ankylosing spondylitis: association with clinical variables. Rheumatol Int 2013 Sep;33(9):2351-2358. [doi: 10.1007/s00296-013-2723-2] [Medline: 23543327]

14. Çınar E, Akkoç Y, Karapolat H, Durusoy R, Keser G. Postural deformities: potential morbidities to cause balance problems in patients with ankylosing spondylitis? Eur J Rheumatol 2016 Mar;3(1):5-9 [FREE Full text] [doi: 10.5152/eurjrheum.2015.15104] [Medline: 27708961]

15. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

16. Verhoeven F, Guillot X, Prati C, Mougin F, Tordi N, Demougeot C, et al. Aerobic exercise for axial spondyloarthritis - its effects on disease activity and function as compared to standard physiotherapy: A systematic review and meta-analysis. Int J Rheum Dis 2018 Sep 05. [doi: 10.1111/1756-185X.13385] [Medline: 30187695]

17. Boyer KA, Johnson RT, Banks JJ, Jewell C, Hafer JF. Systematic review and meta-analysis of gait mechanics in young and older adults. Exp Gerontol 2017 Dec;95:63-70. [doi: 10.1016/j.exger.2017.05.005] [Medline: 28499954]

18. Stebbins J, Trinler UK, Baker R, Brunner R, Wren T, Theologis T. Recommendations for reporting gait studies. Gait Posture 2015 Feb;41(2):339-340. [doi: 10.1016/j.gaitpost.2014.11.014] [Medline: 25497141]

19. Shull PB, Jirattigalachote W, Hunt MA, Cutkosky MR, Delp SL. Quantified self and human movement: a review on the clinical impact of wearable sensing and feedback for gait analysis and intervention. Gait Posture 2014;40(1):11-19 [FREE Full text] [doi: 10.1016/j.gaitpost.2014.03.189] [Medline: 24768525]

20. Geroin C, Mazzoleni S, Smania N, Gandolfi M, Bonaiuti D, Gasperini G, Italian Robotic Neurorehabilitation Research Group. Systematic review of outcome measures of walking training using electromechanical and robotic devices in patients with stroke. J Rehabil Med 2013 Nov;45(10):987-996 [FREE Full text] [doi: 10.2340/16501977-1234] [Medline: 24150661]

21. Gor-García-Fogeda MD, Cano DLCR, Carratalá TM, Alguacil-Diego IM, Molina-Rueda F. Observational gait assessments in people with neurological disorders: A systematic review. Arch Phys Med Rehabil 2016 Jan;97(1):131-140. [doi: 10.1016/j.apmr.2015.07.018] [Medline: 26254954]

22. Buckley E, Mazzà C, McNeill A. A systematic review of the gait characteristics associated with Cerebellar Ataxia. Gait Posture 2018 Dec;60:154-163 [FREE Full text] [doi: 10.1016/j.gaitpost.2017.11.024] [Medline: 29220753]

23. Constantinou M, Barrett R, Brown M, Mills P. Spatial-temporal gait characteristics in individuals with hip osteoarthritis: a systematic literature review and meta-analysis. J Orthop Sports Phys Ther 2014 Apr;44(4):291-2B7. [doi: 10.2519/jospt.2014.4634] [Medline: 24450373]

24. Allet L, Armand S, Golay A, Monnin D, de Bie RA, de Bruin ED. Gait characteristics of diabetic patients: a systematic review. Diabetes Metab Res Rev 2008;24(3):173-191. [doi: 10.1002/dmrr.809] [Medline: 18232063]

25. Schwenk M, Howe C, Saleh A, Mohler J, Grewal G, Armstrong D, et al. Frailty and technology: a systematic review of gait analysis in those with frailty. Gerontology 2014;60(1):79-89 [FREE Full text] [doi: 10.1159/000354211] [Medline: 23949441]

26. Mc Ardle R, Morris R, Wilson J, Galna B, Thomas AJ, Rochester L. What can quantitative gait analysis tell us about dementia and its subtypes? A structured review. J Alzheimers Dis 2017;60(4):1295-1312. [doi: 10.3233/JAD-170541] [Medline: 29036826]

Abbreviations

PEDro: Physiotherapy Evidence Database
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis

Protocol

Medication Management Models for Polymedicated Home-Dwelling Older Adults With Multiple Chronic Conditions: Protocol of a Systematic Review

Filipa Pereira[1,2], BSc, MSc; Pauline Roux[3], PhD; Joëlle Rosselet Amoussou[4], BSc; Maria Manuela Martins[2,5], PhD; Armin von Gunten[6], MD; Henk Verloo[1,6], PhD

[1]School of Health Sciences, HES-SO Valais-Wallis, Sion, Switzerland

[2]Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal

[3]Research Center for Psychology of Health, Aging and Sport Examination, University of Lausanne, Lausanne, Switzerland

[4]Psychiatry Library, Education and Research Department, Hospital and University of Lausanne, Lausanne, Switzerland

[5]Higher School of Nursing of Porto, Porto, Portugal

[6]Service of Old Age Psychiatry, Lausanne University Hospital, Lausanne, Switzerland

Corresponding Author:
Filipa Pereira, BSc, MSc
School of Health Sciences
HES-SO Valais-Wallis
5 Chemin de l'Agasse
Sion, 1950
Switzerland
Phone: 41 786661700
Email: filipa.pereira@hevs.ch

Abstract

Background: Older adults with multiple chronic diseases commonly require complex medication regimes. When combined with frailty, cognitive impairment, and changing pharmacological prescriptions, older adults' polymedication regimes increase the risk of medication-related problems (MRPs) and hospitalization. Effective, well-organized medication management could avoid MRPs and their clinical outcomes.

Objective: Identify medication management models and analyze their impact on managing and preventing MRPs for polymedicated, home-dwelling older adults.

Methods: We will conduct a systematic review of published articles in relevant professional scientific journals from inception until March 31, 2019, in the following electronic databases,: Embase; Medline OvidSP; PubMed (NOT Medline[sb]); Cumulative Index to Nursing and Allied Health Literature (CINAHL) EBSCO; PsycINFO OvidSP; Cochrane Library, Wiley; and Web of Science. We will also hand search the bibliographies of all the relevant articles found and search for unpublished studies. We will consider publications in English, French, German, Spanish, Italian, and Portuguese. Retrieved articles will be screened for eligibility. Statistical analyses will be conducted following the recommendations of the Cochrane Handbook for Systematic Reviews of Interventions, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), and Meta-analysis Of Observational Studies in Epidemiology (MOOSE) statements. Data will be analyzed using SPSS Statistics for Windows, version 25.0 (IBM Corp), and Review Manager, version 5.5 (The Nordic Cochrane Centre, The Cochrane Collaboration).

Results: A preliminary search in Embase delivered 3272 references. This preliminary search allows us to complete our research strategy with equation development and to search the other databases. Relevant articles identified will allow for searching the reference lists for unpublished studies. The inclusion and exclusion criteria will be rigorously respected in the study selection. The entire study is expected to be completed by January 2020.

Conclusions: This review will provide an exhaustive view of medication management models that could be effective for polymedicated, home-dwelling older adults and will allow us to analyze their impact on managing and preventing MRPs.

Trial Registration: PROSPERO CRD42018117287; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=117287 (Archived by WebCite at http://www.webcitation.org/77fCfbCjT)

International Registered Report Identifier (IRRID): DERR1-10.2196/13582

XSL•FO
RenderX

Introduction

Health care systems across Europe are being challenged by an ageing population [1]. International statistics demonstrate that home-dwelling older adults consume huge amounts of prescribed and over-the-counter medication [2,3]. The risks of medication prescription and deprescription have been debated since the dissemination of systematic and umbrella reviews pointing out the inconsistency of some pharmacological treatments for older adults [4,5]. However, a substantial proportion of older adults have multiple chronic diseases requiring numerous treatment components and complex medication regimes [6]. Increasingly complex medication regimes, combined with frailty, reduced cognitive function, and changing pharmacokinetics and pharmacodynamics, increase the risk of adverse drug events and other medication-related problems (MRPs) in this population [7]. An MRP is "an event or circumstance involving medication therapy that actually or potentially interferes with desired health outcomes" [8]. MRPs include inappropriate prescribing (ie, wrong drug, dose, dosage frequency, or dosage form), drug interactions, adverse drug reactions, incorrect administration, the need for monitoring, and nonadherence to medication therapy [9]. MRPs occur frequently among polymedicated, home-dwelling older adults and are associated with increased risks of hospitalization, morbidity, and mortality [10-12]. Avoidable adverse drug events are the serious consequences of inappropriate drug prescribing [13]. For instance, adverse drug events alone contribute to 30%-40% of acute hospital admissions among older adults, although many are preventable [14]. The World Health Organization has estimated that 50% of patients suffering from chronic diseases either do not take their medication or fail to follow instructions for their medical prescription [15]. Medication management among polymedicated, home-dwelling older adults is a serious problem because of the increased burden of symptoms and disease, leading to the use of more medicines and a greater chance of suboptimal management. Estimates of medication nonadherence vary from 40% to 75% [16].

The misuse of categories of drugs such as sleeping pills, analgesics, tranquilizers, appetite suppressants, and stimulants is common [17]. Problems associated with low or zero therapeutic adherence are even more evident during the sensitive period following discharge home from hospital [18]. Suboptimal medication management may lead to a deterioration in the patient's clinical condition, avoidable short-term hospitalizations or readmissions, physical and cognitive decline, exacerbated chronic medical conditions, and, consequently, increased health care use and costs [18-20]. The indirect impacts of the adverse effects of drug nonadherence, such as falls, dehydration, or delirium, may also lead to hospitalizations [21]. Older adults often undergo changes to dosage and prescribed medication during hospitalization [22] and during the first few months after hospital discharge due to comorbidities and the need for disease stabilization [19]. Such changes tend to decrease optimal medication management [23]. Older adults may also go back to taking medications that were discontinued during hospitalization, fail to begin a new treatment initiated during hospitalization, or take incorrect dosages [18]; they are particularly at risk of nonadherence in the first days or weeks after hospital discharge [18]. Home-dwelling older adults taking five or more drugs are considered to be more susceptible to the consequences of polypharmacy, such as adverse drug reactions, drug-drug interactions, nonadherence, or drug-food interactions. Monitoring, assessing, and reacting accordingly are requisite skills for optimal medication management in cases involving inappropriate polymedication or excessive polymedication [24]. Optimal medication management should thus be an integral part of older adults' daily lives and is an essential condition for successfully maintaining them at home [25,26]. Moreover, taking into account the high prevalence of multiple chronic conditions in this population, optimal medication management often becomes a determinant of an older adult's state of health and quality of life at home [27]. Medication management in polymedicated, home-dwelling older people has been described as the "single most important health care intervention in the industrialized world" [28]. This brings us to our search for the best practice models that optimize safety, the continuity of medication intake, and overall medication management among home-dwelling older adults [6]. With the overall aim of developing effective strategies to prevent MRPs and avoiding medication-related hospitalizations and rehospitalizations, many researchers and clinicians involved in primary care have tried to generate either general or specific structured, systematic, medication management models for polymedicated home-dwelling older adults.

One recent example of a structured action to prevent MRPs is the Medication Management Model. This is based on risk management and provides home-dwelling patients with medication review services comparable to those that benefit hospital and nursing home patients [29]. A second example concerns the Coordinated Medication Management Model, which involves home care nurses, nurses, physicians, and community pharmacists in medication processes for home-dwelling older adults, supported by home health care services [30]. Medication management models are interprofessional collaboration processes between patients, their informal caregivers, pharmacists, and nurses, based on evidence-based guidelines and bringing, if appropriate, identified problems to a physician's attention [26]. To the best of our knowledge, there have been no systematic reviews summarizing the medication management models developed in community health care and their impact on MRPs among polymedicated, home-dwelling older adults.

Methods

Overview

This protocol has been registered in the International Prospective Register of Systematic Reviews (PROSPERO) (protocol number CRD42018117287). The systematic review will be conducted following the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols (PRISMA-P) and its checklist for reporting harms [31,32], the reporting proposals of the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) [33], and the methods outlined in the Cochrane Handbook for Systematic Reviews of Interventions [34].

Purpose and Research Question

This systematic review's purpose is to identify, examine, and summarize the models developed in community health care systems to optimize medication management for polymedicated, home-dwelling older adults. The following research question will guide this review: Which community health care models for optimizing medication management for polymedicated, home-dwelling older adults have been reported in interventional and observational studies?

Inclusion and Exclusion Criteria

Types of Studies

This review will include randomized controlled trials, cluster randomized controlled trials, and nonrandomized studies. Nonrandomized studies will include quantitative studies examining the effects of medication management models that do not use randomization to allocate patients to comparison groups [34]. We will include retrospective and prospective epidemiological studies, cohort studies, case-control studies, controlled before-and-after studies, interrupted-time-series studies, and controlled trials with inappropriate randomization (ie, quasi-experimental studies) [35,36]. We will search for papers in French, German, English, Spanish, Italian, and Portuguese.

Types of Participants

This review will consider studies involving polymedicated, home-dwelling older adults with multiple chronic conditions and a minimum mean age of 65 years, as well as studies with participants aged 55 years or older. In order to properly include heterogeneity and complexity, we will consider multiple chronic conditions: the co-occurrence of at least two diseases in the same individual, cumulative indices considering both the number and severity of concurrent diseases, and the simultaneous presence of not only diseases but also symptoms and physical and cognitive dysfunctions.

Types of Models

We will examine all types of medication management models, including strategies, interventions, and clinical pathways aimed at optimizing the effects of medication management for polymedicated, home-dwelling older adults with multiple chronic conditions. Given their impact in reducing medication errors and enhancing interprofessional collaboration and patient

safety, electronic medication management systems (EMMS) will be included in this study [37,38].

Where possible, these types of medication management models will be compared with usual care and will include strategies documented in models, interventions, and clinical pathways delivered by a primary health care provider alone or in collaboration with other allied health care professionals at home.

Based on the Effective Practice and Organization of Care (EPOC) taxonomy of health system interventions [39], we will consider medication management models, interventions, and clinical pathways targeting the health care professional level and the patient level, as discussed below, but we will exclude those targeting health care organizations:

1. Optimized medication management at the health care professional level:
   a. Educational programs aimed at optimizing medication management.
   b. Distribution of materials aimed at optimizing medication management.
   c. Feedback to peers and other involved health care professionals on the effects and impact of medication management (ie, medication review from medical records).
   d. Monitoring medication management models, including interventions and clinical pathways (ie, assessment, adjustment or change of medication, and medication deprescription).
   e. Verbal recommendations to polymedicated, home-dwelling older adults by the health care providers involved to optimize medication management (eg, pharmacists and physicians).
   f. The organized activities of teams for medication conciliation, prescription, and deprescription.
   g. EMMS covering prescription, administration, pharmacy review, barcode medication administration, and anything that encompasses medication management processes for polymedicated, home-dwelling older adults with multiple chronic conditions [40].
   h. Evaluations of the involvement of different health care professionals in the optimization of medication management.

2. Optimized medication management at the level of polymedicated, home-dwelling older adults:
   a. Organized interventions aimed at optimizing medication management for polymedicated, home-dwelling older adults (ie, single- or multi-professional interactions conducted by nurses, pharmacists, or physicians, such as counselling on medication and medication compliance or patient education sessions).
   b. Patient reminder systems aimed at optimizing medication management (ie, single- or multi-professional interventions conducted by nurses, pharmacists, or physicians, such as telephone contact and discharge planning; medication adherence aids, such as electronic monitors or pill dispensers; and meetings with the multi-professional health care team in the patient's home).

Types of Outcome Measures

This review's primary outcome measures will be:

1. The identification of models including interventions, clinical pathways, and EMMS aimed at optimizing medication management for polymedicated, home-dwelling older adults in primary health care.
2. The description of the components of the models, interventions, and clinical pathways and the identification of the stakeholders involved (ie, professional and nonprofessional caregivers).
3. The description of the impact of medication management models, interventions, and clinical pathways versus usual care on:
   a. Rates of hospitalization for MRPs.
   b. Rates of emergency department visits for MRPs.

Primary outcomes will be measured by different methods based on dichotomous (ie, yes or no), ordinal, or continuous rates and scores (eg, hospitalization or rehospitalization, frailty severity or progress, emergency department visits for MRP [10], and misuse of medication [41]).

This review's secondary outcome measures will be descriptions of the associations between sociodemographic characteristics, health data, and MRPs (ie, nominal, ordinal, or interval level).

Information Sources and Search Strategy

We will search the following databases, without restriction on the publication date: Embase (from 1947); Medline OvidSP (from 1946); a subset (sb) of PubMed (NOT Medline[sb]) (from 1996); Cumulative Index to Nursing and Allied Health Literature (CINAHL) EBSCO (from 1937); PsycINFO OvidSP (from 1887); Cochrane Library, Wiley (from 1992); and Web of Science (from 1900). Furthermore, we will search in the reference lists of relevant articles identified and for unpublished studies.

The search syntax will use Medical Subject Headings (MeSH) and text terms with Boolean operators. The syntax will consist of the search themes intersected by the Boolean terms "AND" and "OR." MeSH terms and free keywords will include the following:

1. Terms for "Medication management," "Drug therapy management," "Therapeutic medication management," and "Optimizing medication treatment."
2. Terms for "Community-dwelling older adults," "Home-dwelling older adults," "Elderly," "Aged," "Home-care patients," "Older adults," and "Very old adults."
3. Terms for "Home-dwelling," "Living in place," "Homebound," "Primary care," "Community health services," "Community hospital," "Ambulatory care," "Outpatient clinics," "Hospital," "Ambulatory care facilities," "Day care," "Primary health care," "Community health centers," "Health services for the aged," "Community," "Domicile," "Home or home-care or home-based," "Outpatient," "Day patient," "Community care," "Home-care services," "General practice," and "Urban population."
4. Terms for "Multiple chronic conditions," "Frailty," "Vulnerability," "Multimorbidity," "Multimorbid," "Muscle weakness," "Fatigue," "Slow motor performance," "Low physical activity," and "Unintentional weight loss."
5. Terms related to "Drug-related problems," "Medication-related problems," "Misuse of medication," "Medication abuse," "Inappropriate polypharmacy," "Polypharmacy," "Excessive polypharmacy," and "Polymedication."
6. Terms for models, interventions, and clinical pathways: "Care model," "Care map," "Multidisciplinary care," "Evidence-based care," "Guideline," "Patient care plans," "Clinical paths," "Clinical pathways," "Critical paths," and "Interventions."
7. Terms related to "Medication management model," "Medication optimization," and "Reconciliation."
8. Terms for "Computerized provider order entry," "Electronic prescribing," "Computer-assisted diagnosis," "Computer-assisted therapy," and "Medical device."

Study Selection

Two reviewers (FP and PR) will independently screen the titles and abstracts identified in the searches to assess which studies meet the inclusion criteria. Disagreements will be resolved through discussion or, if needed, a consensus will be reached after discussion with coauthors (MMM and HV). The reviewers will then independently assess the full-text articles to ensure that they meet the inclusion criteria. Disagreements will be discussed and resolved with coauthors (MMM and HV). A flowchart of the trial selection process has been drawn in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [42] (see Figure 1).

Figure 1. Blank flow diagram based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [42].
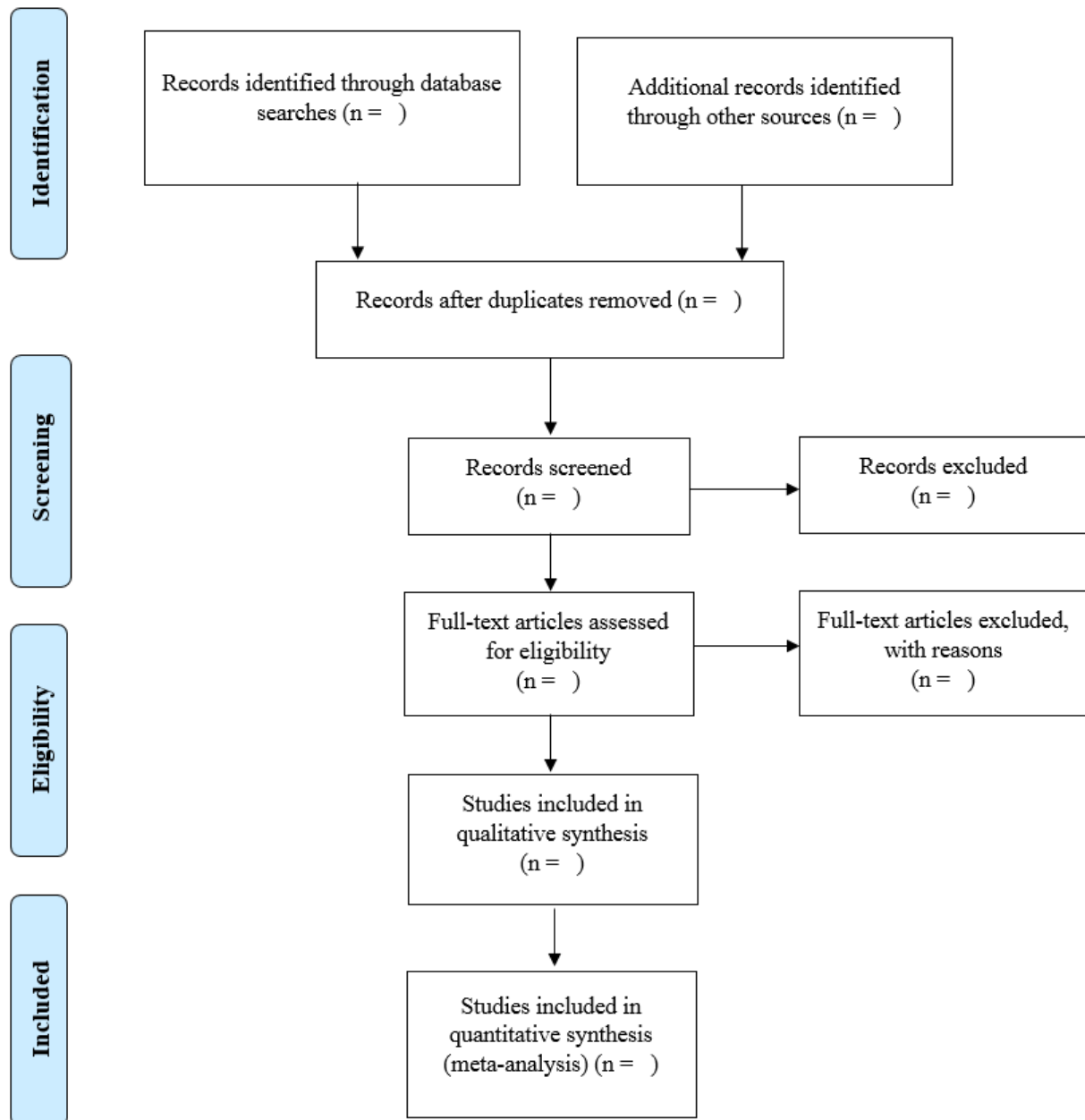
Data Extraction

Data will be extracted independently by two authors (FP and PR) using a specially designed, standardized data extraction form. Discrepancies will be resolved through discussion and consultation with coauthors (MMM and HV).

The following information will be extracted from each study included in the review: (1) study authors, year of publication, and country where the study was conducted; (2) study characteristics, including setting, design, duration of follow-up, and sample size; (3) participants' characteristics, including age, sex, social status, marital status, educational status, level of autonomy, history of hospitalization and rehospitalization for MRPs, and emergency department visits for MRPs; medication management model; interventions; and clinical pathways; (4) multiple chronic conditions measured using indices (ie, the

Charlson Comorbidity Index and the Cumulative Illness Rating Scale-Geriatric); and (5) types of outcome measures.

Assessment of the Risks of Bias in Included Studies

Two reviewers (FP and PR) will independently assess the risks of bias in all the randomized and nonrandomized studies for interventions (NRSI) included, using the validated Cochrane Risk of Bias Tool, version 2.0 (The Cochrane Collaboration) [43]. This tool is based on five domains: (1) bias arising from the randomization process, (2) bias due to deviations from intended interventions, (3) bias due to missing outcome data, (4) bias in outcome measurement, and (5) bias in the selection of the results reported. Each of these five domains will be rated as (1) low risk of bias, (2) some concerns about bias, or (3) high risk of bias. Declaring that a study has a particular level of risk of bias in any individual domain will mean that the study as a whole has a risk of bias. Disagreements will be resolved through discussion and consultation with coauthors (MMM and HV).

We will also use the validated Risk of Bias In Nonrandomized Studies of Interventions (ROBINS-I) tool for assessing the risk of bias in NRSI [44]. This tool covers two dimensions and seven domains through which bias might be introduced into NRSI: (1) preintervention and during the intervention (ie, bias due to confounding, selection of study participants, or classification of the intervention) and (2) postintervention (ie, bias due to deviations from intended interventions, missing data, measurement of outcomes, and selection of the reported result) [44]. Any disagreements in quality assessments will be resolved through discussion.

Statistical Analyses

Statistical analyses will be conducted by FP and PR following the recommendations of the Cochrane Handbook for Systematic Reviews of Interventions [34] and the PRISMA-P and MOOSE statements [31,33]. For dichotomous outcomes, average intervention effects (ie, pooled effect and meta-analysis) will be calculated as relative risks with 95% CIs using a random-effects model [45]. For continuous data, a random-effects model will be used to calculate weighted-mean differences with 95% CIs. If required, we will calculate standard deviations from the standard errors or 95% CIs presented in the articles. Heterogeneity will be quantified using the I^2 and chi-square tests. Funnel plots will be drawn and Egger's test will be computed to explore the possibility of publication bias [46].

Reasons for heterogeneity in effect estimates will be sought in meta-analyses [47,48]. To explore the possible determinants of heterogeneity, we will conduct subgroup analyses according to selected study characteristics (eg, participants' ages; country where the study was conducted; types of professions; and types of models, interventions, and clinical pathways). Furthermore, sensitivity analyses will be conducted by (1) excluding relatively small studies (ie, fewer than 20 participants per randomization group) and (2) restricting analyses to good-quality studies. Data will be analyzed using SPSS Statistics for Windows, version 25.0 (IBM Corp), and Review Manager, version 5.5 (The Nordic Cochrane Centre, The Cochrane Collaboration).

Results

To date, searches in Embase have been performed, delivering 3272 references. As of publication of this protocol, we are developing the search equations in the remaining databases to then initiate the process of study selection, rigorously applying the inclusion and exclusion criteria. The final results are expected in January 2020.

Discussion

Principal Considerations

To the best of our knowledge, this systematic review will be the first to synthesize evidence about medication management models, including their interventions, clinical pathways, and EMMS, as well as their impacts on MRPs among polymedicated, home-dwelling older adults.

Since MRPs are associated with an increased risk of hospital readmissions, morbidity, and mortality and are significant issues for the health care system, it is very important to develop intervention strategies to resolve or prevent them. The suboptimal medication management causing MRPs involving the hospitalization of polymedicated, home-dwelling older adults is underrecognized by community health care providers, especially by frontline community health care nurses. Therefore, an important task for community health care providers is to identify, resolve, and prevent the occurrence of MRPs in this rapidly growing population of polymedicated, home-dwelling older adults.

The results of this systematic review may guide future research in this avenue. The results may also contribute to the development of comprehensive and implementable recommendations for primary health care practitioners and policy makers concerning medication management among polymedicated, home-dwelling older adults.

Strengths and Limitations

This systematic review protocol's strengths are as follows: (1) clear definitions of the major concepts of medication management models, multiple chronic conditions, clinical pathways, and interventions; (2) the use of an appropriate search strategy designed in collaboration with a health librarian experienced in conducting such reviews; and (3) the inclusion criteria, which impose few restrictions on the study's language, age, or geographic location. Nevertheless, there are several limitations that should be noted. The authors' personal judgements may introduce bias into the assessment. Nonetheless, this risk will be reduced by using two reviewers to select and assess the eligibility of studies independently. Furthermore, it is possible that some eligible studies may not be covered by our research strategy. We will seek to minimize this limitation through the contribution of an experienced health librarian. Finally, the expected heterogeneity of studies about medication management models for polymedicated, home-dwelling older adults with multiple chronic conditions may influence our ability to state comprehensive and implementable recommendations from the literature.

Conclusions

This systematic review will synthesize the available evidence about medication management models among polymedicated, home-dwelling older adults and their impact on MRPs. We expect our findings to provide meaningful evidence toward the optimization of health care models, programs, and services for polymedicated, home-dwelling older adults with multiple chronic conditions.

Conflicts of Interest

None declared.

Authors' Contributions

FP is the guarantor and all the authors contributed to drafting the protocol. All authors will contribute to the development of the selection criteria, data extraction and analysis, and the search strategy. All the authors approved the final protocol manuscript.

References

1. OECD/European Union. Health at a Glance: Europe 2018. State of Health in the EU Cycle. Paris, France: OECD Publishing; Nov 02, 2018.

2. Cooper RJ. Over-the-counter medicine abuse: A review of the literature. J Subst Use 2013 Apr;18(2):82-107 [FREE Full text] [doi: 10.3109/14659891.2011.615002] [Medline: 23525509]

3. Calderón-Larrañaga A, Gimeno-Feliu LA, González-Rubio F, Poblador-Plou B, Lairla-San José M, Abad-Díez JM, et al. Polypharmacy patterns: Unravelling systematic associations between prescribed medications. PLoS One 2013;8(12):e84967 [FREE Full text] [doi: 10.1371/journal.pone.0084967] [Medline: 24376858]

4. Salahudeen MS. Deprescribing medications in older people: A narrative review. Drugs Today (Barc) 2018 Aug;54(8):489-498. [doi: 10.1358/dot.2018.54.8.2856495] [Medline: 30209443]

5. Page A, Potter K, Clifford R, Etherton-Beer C. Deprescribing in older people. Maturitas 2016 Sep;91:115-134. [doi: 10.1016/j.maturitas.2016.06.006] [Medline: 27451330]

6. Bachmann N, Burla L, Kohler D. La Santé en Suisse: Le Point sur les Maladies Chroniques. Bern, Switzerland: Hogrefe Verlag; 2015. URL: https://www.obsan.admin.ch/sites/default/files/publications/2015/rapportsante_2015_f_0.pdf [accessed 2019-05-03] [WebCite Cache ID 789Y9Uvu1]

7. Johansen J, Havnes K, Halvorsen K, Haustreis S, Skaue L, Kamycheva E. Interdisciplinary collaboration across secondary and primary care to improve medication safety in the elderly (IMMENSE study): Study protocol for a randomised controlled trial. BMJ Open 2018;8(1). [doi: 10.1136/bmjopen-2017-020106]

8. PCNE Classification for Drug Related Problems V5.01. Zuidlaren, The Netherlands: Pharmaceutical Care Network Europe Foundation; 2006. URL: https://www.pcne.org/upload/files/16_PCNE_classification_V5.01.pdf [accessed 2019-01-23] [WebCite Cache ID 75fI3yhD7]

9. Nolan L, O'Malley K. Prescribing for the elderly. Part I: Sensitivity of the elderly to adverse drug reactions. J Am Geriatr Soc 1988 Feb;36(2):142-149. [Medline: 3276767]

10. Ahmad A, Mast MR, Nijpels G, Elders PJ, Dekker JM, Hugtenburg JG. Identification of drug-related problems of elderly patients discharged from hospital. Patient Prefer Adherence 2014;8:155-165 [FREE Full text] [doi: 10.2147/PPA.S48357] [Medline: 24523581]

11. Benson H, Lucas C, Kmet W, Benrimoj SI, Williams K. Pharmacists in general practice: A focus on drug-related problems. Int J Clin Pharm 2018 Jun;40(3):566-572. [doi: 10.1007/s11096-018-0617-9] [Medline: 29542035]

12. Ensing HT, Koster ES, Dubero DJ, van Dooren AA, Bouvy ML. Collaboration between hospital and community pharmacists to address drug-related problems: The HomeCoMe-program. Res Social Adm Pharm 2019 Jan;15(3):267-278. [doi: 10.1016/j.sapharm.2018.05.001] [Medline: 29773308]

13. Chan M, Nicklason F, Vial JH. Adverse drug events as a cause of hospital admission in the elderly. Intern Med J 2001;31(4):199-205. [Medline: 11456032]

14. Basnet S, Zhang M, Lesser M, Wolf-Klein G, Qiu G, Williams M, et al. Thirty-day hospital readmission rate amongst older adults correlates with an increased number of medications, but not with Beers medications. Geriatr Gerontol Int 2018 Oct;18(10):1513-1518. [doi: 10.1111/ggi.13518] [Medline: 30225904]

15. Sabaté E(. Adherence to Long-Term Therapies: Evidence for Action. Geneva, Switzerland: World Health Organization; 2003. URL: https://apps.who.int/iris/bitstream/handle/10665/42682/9241545992.pdf;jsessionid=17D1922E658FE9AB2DF3923E39FC602B?sequence=1 [accessed 2019-01-22] [WebCite Cache ID 75fJ9UsSh]

16. Doggrell SA. Adherence to medicines in the older-aged with chronic conditions: Does intervention by an allied health professional help? Drugs Aging 2010 Mar 01;27(3):239-254. [doi: 10.2165/11532870-000000000-00000] [Medline: 20210369]

17. Menecier P, Fernandez L. Addictive behavior among the elderly [Article in French]. Presse Med 2012 Dec;41(12 Pt 1):1226-1232. [doi: 10.1016/j.lpm.2012.07.020] [Medline: 23040954]

18. Fallis BA, Dhalla IA, Klemensberg J, Bell CM. Primary medication non-adherence after discharge from a general internal medicine service. PLoS One 2013;8(5):e61735 [FREE Full text] [doi: 10.1371/journal.pone.0061735] [Medline: 23658698]

19. Mansur N, Weiss A, Hoffman A, Gruenewald T, Beloosesky Y. Continuity and adherence to long-term drug treatment by geriatric patients after hospital discharge: A prospective cohort study. Drugs Aging 2008;25(10):861-870. [doi: 10.2165/00002512-200825100-00005] [Medline: 18808210]

XSL•FO

RenderX

20. Nickel CH, Ruedinger JM, Messmer AS, Maile S, Peng A, Bodmer M, et al. Drug-related emergency department visits by elderly patients presenting with non-specific complaints. Scand J Trauma Resusc Emerg Med 2013 Mar 05;21:15 [FREE Full text] [doi: 10.1186/1757-7241-21-15] [Medline: 23497667]

21. Bianco A, Molè A, Nobile CGA, Di Giuseppe G, Pileggi C, Angelillo IF. Hospital readmission prevalence and analysis of those potentially avoidable in southern Italy. PLoS One 2012;7(11):e48263 [FREE Full text] [doi: 10.1371/journal.pone.0048263] [Medline: 23133624]

22. Shelton PS, Mozingo DB, Avissar PS, Karg M, Charboneau AL, Rich W. Measuring adherence in a community-based elderly population. Consult Pharm 2012 Nov;27(11):771-781. [doi: 10.4140/TCP.n.2012.771] [Medline: 23168927]

23. Chan FW, Wong FY, So WY, Kung K, Wong CK. How much do elders with chronic conditions know about their medications? BMC Geriatr 2013 Jun 13;13:59 [FREE Full text] [doi: 10.1186/1471-2318-13-59] [Medline: 23758824]

24. Garnett A, Ploeg J, Markle-Reid M, Strachan PH. Self-management of multiple chronic conditions by community-dwelling older adults: A concept analysis. SAGE Open Nurs 2018 Jan 17;4:237796081775247. [doi: 10.1177/2377960817752471]

25. Doggrell SA. Inadequate management of medicines by the older-aged living in a retirement village. Int J Clin Pharm 2013 Aug;35(4):546-549. [doi: 10.1007/s11096-013-9786-8] [Medline: 23657755]

26. Frey D. The medication management model: An effective tool for home health providers. Home Health Care Manag Pract 2016 Jun 30;15(3):222-230. [doi: 10.1177/1084822302250690]

27. Williams GC, Patrick H, Niemiec CP, Williams LK, Divine G, Lafata JE, et al. Reducing the health risks of diabetes: How self-determination theory may help improve medication adherence and quality of life. Diabetes Educ 2009;35(3):484-492 [FREE Full text] [doi: 10.1177/0145721709333856] [Medline: 19325022]

28. Avorn J. Medication use in older patients: Better policy could encourage better practice. JAMA 2010 Oct 13;304(14):1606-1607. [doi: 10.1001/jama.2010.1495] [Medline: 20940388]

29. Brown NJ, Griffin MR, Ray WA, Meredith S, Beers MH, Marren J, et al. A model for improving medication use in home health care patients. J Am Pharm Assoc (Wash) 1998;38(6):696-702. [Medline: 9861787]

30. Toivo T, Dimitrow M, Puustinen J, Savela E, Pelkonen K, Kiuru V, et al. Coordinating resources for prospective medication risk management of older home care clients in primary care: Procedure development and RCT study design for demonstrating its effectiveness. BMC Geriatr 2018 Jan 16;18(1):74 [FREE Full text] [doi: 10.1186/s12877-018-0737-z] [Medline: 29548304]

31. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]

32. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, PRISMAHarms Group. PRISMA harms checklist: Improving harms reporting in systematic reviews. BMJ 2016 Feb 01;352:i157. [doi: 10.1136/bmj.i157] [Medline: 26830668]

33. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000 Apr 19;283(15):2008-2012. [Medline: 10789670]

34. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons; 2008.

35. Sedgwick P. What is a non-randomised controlled trial? BMJ 2014 Jun 20;348:g4115. [doi: 10.1136/bmj.g4115] [Medline: 24951505]

36. Ferriter M, Huband N. Does the non-randomized controlled study have a place in the systematic review? A pilot study. Crim Behav Ment Health 2005;15(2):111-120. [doi: 10.1002/cbm.43] [Medline: 16470505]

37. Kallio S, Kumpusalo-Vauhkonen A, Järvensivu T, Mäntylä A, Pohjanoksa-Mäntylä M, Airaksinen M. Towards interprofessional networking in medication management of the aged: Current challenges and potential solutions in Finland. Scand J Prim Health Care 2016 Dec;34(4):368-376 [FREE Full text] [doi: 10.1080/02813432.2016.1249055] [Medline: 27822968]

38. Shah K, Lo C, Babich M, Tsao NW, Bansback NJ. Bar code medication administration technology: A systematic review of impact on patient safety when used with computerized prescriber order entry and automated dispensing devices. Can J Hosp Pharm 2016;69(5):394-402 [FREE Full text] [Medline: 27826157]

39. Mowatt G, Grimshaw JM, Davis DA, Mazmanian PE. Getting evidence into practice: The work of the Cochrane Effective Practice and Organization of Care Group (EPOC). J Contin Educ Health Prof 2001;21(1):55-60. [doi: 10.1002/chp.1340210109] [Medline: 11291587]

40. Pearce R, Whyte I. Electronic medication management: Is it a silver bullet? Aust Prescr 2018 Apr;41(2):32-33 [FREE Full text] [doi: 10.18773/austprescr.2018.012] [Medline: 29670307]

41. Badrakalimuthu VR, Rumball D, Wagle A. Drug misuse in older people: Old problems and new challenges. Adv Psychiatr Treat 2018 Jan 02;16(6):421-429. [doi: 10.1192/apt.bp.108.006221]

42. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]

43. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: A conversation between trialists and epidemiologists. Epidemiology 2017 Jan;28(1):54-59 [FREE Full text] [doi: 10.1097/EDE.0000000000000564] [Medline: 27748683]

44. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]

45. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011 Feb 10;342:d549. [doi: 10.1136/bmj.d549] [Medline: 21310794]

46. Lau J, Ioannidis JPA, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ 2006 Sep 16;333(7568):597-600 [FREE Full text] [doi: 10.1136/bmj.333.7568.597] [Medline: 16974018]

47. Chiolero A, Santschi V, Burnand B, Platt RW, Paradis G. Meta-analyses: With confidence or prediction intervals? Eur J Epidemiol 2012 Oct;27(10):823-825. [doi: 10.1007/s10654-012-9738-y] [Medline: 23070657]

48. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester, UK: John Wiley & Sons; 2009.

Abbreviations

CINAHL: Cumulative Index to Nursing and Allied Health Literature
EMMS: electronic medication management systems
EPOC: Effective Practice and Organization of Care
MeSH: Medical Subject Heading
MOOSE: Meta-analysis Of Observational Studies in Epidemiology
MRP: medication-related problem
NRSI: nonrandomized studies for interventions
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Protocols
PROSPERO: International Prospective Register of Systematic Reviews
ROBINS-I: Risk of Bias In Nonrandomized Studies of Interventions
sb: subset

Protocol

A Patient-Centered Mobile Phone App (iHeartU) With a Virtual Human Assistant for Self-Management of Heart Failure: Protocol for a Usability Assessment Study

Lingling Zhang[1], ScD; Sabarish V Babu[2], PhD; Meenu Jindal[3], MD; Joel E Williams[4], PhD; Ronald W Gimbel[4], PhD

[1]College of Nursing and Health Sciences, University of Massachusetts Boston, Boston, MA, United States
[2]School of Computing, Clemson University, Clemson, SC, United States
[3]Department of Medicine, Greenville Health System, Greenville, SC, United States
[4]Department of Public Health Sciences, Clemson University, Clemson, SC, United States

Corresponding Author:
Lingling Zhang, ScD
College of Nursing and Health Sciences
University of Massachusetts Boston
100 Morrissey Blvd
Boston, MA, 02125
United States
Phone: 1 6172877845
Email: lingling.zhang@umb.edu

Abstract

Background: Heart failure (HF) causes significant economic and humanistic burden for patients and their families, especially those with a low income, partly due to high hospital readmission rates. Optimal self-care is considered an important nonpharmacological aspect of HF management that can improve health outcomes. Emerging evidence suggests that self-management assisted by smartphone apps may reduce rehospitalization rates and improve the quality of life of patients. We developed a virtual human–assisted, patient-centered mobile health app (iHeartU) for patients with HF to enhance their engagement in self-management and improve their communication with health care providers and family caregivers. iHeartU may help patients with HF in self-management to reduce the technical knowledge and usability barrier while maintaining a low cost and natural, effective social interaction with the user.

Objective: With a standardized systematic usability assessment, this study had two objectives: (1) to determine the obstacles to effective and efficient use of iHeartU in patients with HF and (2) to evaluate of HF patients' adoption, satisfaction, and engagement with regard to the of iHeartU app.

Methods: The basic methodology to develop iHeartU systems consists of a user-centric design, development, and mixed methods formative evaluation. The iterative design and evaluation are based on the guidelines of the American College of Cardiology Foundation and American Heart Association for the management of heart failure and the validated "Information, Motivation, and Behavioral skills" behavior change model. Our hypothesis is that this method of a user-centric design will generate a more usable, useful, and easy-to-use mobile health system for patients, caregivers, and practitioners.

Results: The prototype of iHeartU has been developed. It is currently undergoing usability testing. As of September 2018, the first round of usability testing data have been collected. The final data collection and analysis are expected to be completed by the end of 2019.

Conclusions: The main contribution of this project is the development of a patient-centered self-management system, which may support HF patients' self-care at home and aid in the communication between patients and their health care providers in a more effective and efficient way. Widely available mobile phones serve as care coordination and "no-cost" continuum of care. For low-income patients with HF, a mobile self-management tool will expand their accessibility to care and reduce the cost incurred due to emergency visits or readmissions. The user-centered design will improve the level of engagement of patients and ultimately lead to better health outcomes. Developing and testing a novel mobile system for patients with HF that incorporates chronic disease management is critical for advancing research and clinical practice of care for them. This research fills in the gap in user-centric design and lays the groundwork for a large-scale population study in the next phase.

International Registered Report Identifier (IRRID): DERR1-10.2196/13502

XSL•FO
RenderX

KEYWORDS

heart failure; mobile health; self-management; patient engagement; virtual human

Introduction

Background

Heart failure (HF) is a chronic condition associated with significant morbidity, early mortality, and impaired quality of life (QoL), which poses tremendous economic and humanistic burden on patients, families, the health care system, and society [1]. Hospital readmission rates for HF are among the highest of any chronic disease [1,2]. Half of the patients with HF experience rehospitalization within 6 months of hospital discharge [3]. Those previously hospitalized for HF had the greatest rates of HF rehospitalization or cardiovascular death [4,5]. Limited access to care may be an additional problem among patients with limited socioeconomic means. Patients lacking access to primary care physicians after hospital discharge often experience clinical deterioration, necessitating hospital readmission [6]. This problem is especially magnified with uninsured patients, who may be forced into a cycle of relying on using costly emergency room visits [7].

According to the American College of Cardiology Foundation and American Heart Association guidelines for the management of HF, lack of improvement in health-related quality of life (HRQoL) after discharge from the hospital is a powerful predictor of rehospitalization and mortality [8-10], and no pharmacological therapy is a consistent determinant of HRQoL [8]. Previous studies have shown that self-management, which is expected to be integral to both the maintenance of wellness and the management of illness [11], can improve patients' HRQoL [12-14]. As most chronic conditions are related to lifestyle, self-management was designed to meet the needs of managing daily treatment and life activities to improve health and health behaviors [11].

Mobile Health and Heart Failure

To foster the ability of patients to practice more effective self-management and offset the difficulties due to lack of access to care, telemonitoring systems may be useful tools in reducing rehospitalization and improving QoL for patients with HF, as suggested by emerging evidence [1]. However, special digital devices required for telemonitoring are costly, limiting its potential for wide use, especially among low-income patients.

Optimal self-care is considered an important nonpharmacological aspect of HF management that can improve health outcomes [15]. Mobile health (mHealth) technologies have emerged as a way to actively engage patients in self-management and health care decision-making processes [16]. The access to mobile phones makes it possible for mHealth apps to transform treatment adherence through improved self-management [17]. About 85% of adults aged 65 years or above in the United States owned a mobile phone, and this proportion is increasing in all age groups [18]. The use of mobile phones to manage daily self-management of chronic diseases has been reported even among older adults with no experience in technology [14]. However, the dedicated, persistent use of self-monitoring systems and user engagement rates are still low [19,20]. Standardized, systematic mHealth usability assessments need to be performed, but even these have been insufficient in the past [21]. The Technology Acceptance Model [22] hypothesized that perceived usefulness, perceived ease of use, and attitude toward the technology were three factors determining whether the user would ultimately use or reject the technology. Patient involvement in the development of new information and communication technology (ICT) is crucial, given that the typical patient with chronic disease is an older adult with difficulties in understanding and using standard ICT equipment [23]. In order to support acceptance and later adoption of the new system, the innovation process should include the perspective of patients and the needs of other stakeholders in the care of chronic diseases [23].

Virtual Human and Patient Engagement

To date, mobile phone–based apps for HF monitoring and management have not been widely researched. To address this gap, iHeartU—a mobile phone app with a novel virtual human assistant ("iHeartHelper")—was developed to help patients with HF in their self-management by providing a holistic engagement experience (Figure 1). Its prototype has been undergoing usability testing since April 2018.

Patient engagement is a key mediator for behavior changes in patients with HF since patients who are engaged as decision-makers in their care tend to be healthier and have better outcomes (Figure 1) [17]. The development of iHeartU is based on the guidelines of the American College of Cardiology Foundation and American Heart Association for the management of HF and the validated Information, Motivation, and Behavioral skills behavior change model. This model emphasizes on information and knowledge about patient behavior, motivation to perform the behavior, and the patient's behavioral skills [24].

A unique feature of this app—iHeartHelper—is an interactive virtual human or embodied conversational agent that resembles a human assistant and provides natural social interaction with users of the system. The personalized conversational interactions facilitated by the iHeartHelper can increase the engagement of patients with HF, health care providers, and family caregivers, which may lead to an improvement in patients' health outcomes and reduction in hospital readmission rates.

Figure 1. Conceptual framework for iHeartU: IMB model with patient engagement as mediator. HF: heart failure; VR: virtual reality.
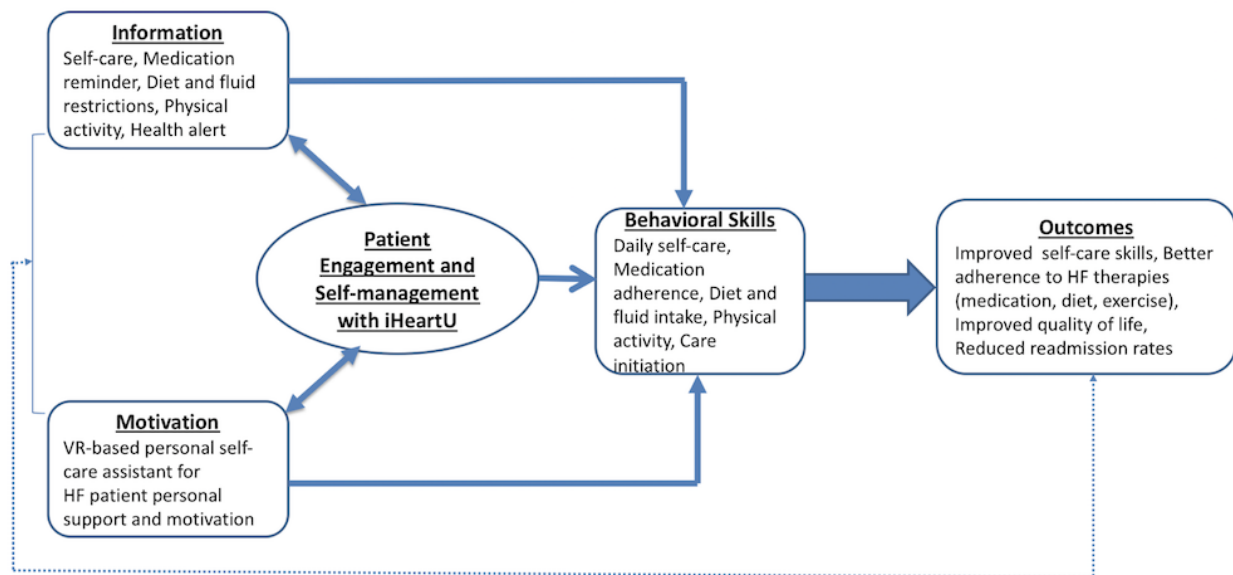
There are existing studies that use virtual environment simulation to treat patients with posttraumatic stress disorder, fear of heights [25,26], and public speaking anxiety [27] and to conduct medical training [28]. However, whether the use of virtual human interface will improve patient engagement has not been systemically studied. The research assumption of this study is that a patient-centered mobile phone app with a virtual human assistant can improve engagement of patients with HF.

Objectives

This study focuses on the user-centered design and assessment of the iHeartU mobile phone app. Unlike mHealth apps in previous studies, iHeartU is featured by the use of virtual human interface, which may involve more assessments in its usability testing but with the expectation of better user experience. Usability testing is the term used to describe the assessment of how easy a user interface is to operate [29]. The objectives of this study are two-fold: (1) to determine the obstacles to effective and efficient use of iHeartU among patients with HF and (2) to evaluate HF patients' adoption, satisfaction, and engagement with regard to iHeartU.

Methods

Study Design

A single-arm prospective observation study is currently underway for the usability assessment of the iHeartU app. A mixed method design is used for this study, including several established usability evaluations through surveys, direct observations, interviews, think-aloud protocol, and focus groups. Additionally, the researchers have been working closely with the Patient Engagement Studio (an HF panel of patients and all stakeholders, including patients, family caregivers, physicians, nurses, and other health care providers) at Greenville Health

System (GHS) to seek input and feedback on the design, development, and evaluation process.

Participants

With the approval of the Institutional Review Board of GHS (Pro00066413), we aim to recruit 10 patients with HF to conduct the usability testing. The sample size is in accordance with the existing literature [30,31], where Kushniruk [30] and Vizri [31] have shown that 70% of severe usability problems can be discovered by the first five users and up to 85%, by the eighth user, following which less problems tend be identified, and these problems are also less significant. Patients are eligible to participate in this study if they have a clinical diagnosis of HF, are English-speaking, are able to operate study devices, have a New York Heart Association Functional Classification I-III, and are willing to provide informed consent for participating in the study. Patients are recruited by their direct care team during scheduled clinic visits. They are informed about the purpose of the study upon providing written consent for participation. There are no expected risks or discomforts in the study other than possible confusion when using iHeartU. Participation in this study is completely voluntary; therefore, patients can refuse to participate or withdraw from the study at any time. Patients who are unable to operate the devices and have a limited cognitive ability, as determined by the patient's clinician, will be excluded from consideration.

Health care providers in the direct care team for patients with HF, including physicians and nurses, are also involved in the testing. One representative of each group is asked to complete the survey questionnaires after providing consent. The research team collects their feedback and input to provide a helpful clinical perspective for improving the design of the mobile phone app.

Measures

An iterative approach is used to refine iHeartU for patients with HF based on their experience of interacting with this app while aiming for continual patient satisfaction. Patients will be the primary users of the app. Health care providers will help identify patient needs and health problems. The following instruments and methods will be used to collect data:

- A demographic and background questionnaire to describe personal and health information as well as experience with mobile phone use.
- Usability evaluation metrics, guided by the International Organization for Standardization 9241-11 Usability framework and mobile health usability research [21], to evaluate effectiveness, efficiency, and satisfaction. Engagement will be added as the fourth dimension (Table 1).
- Think-aloud protocol to capture patients' cognitive processes while performing representative tasks on the app [30].
- Individual interviews with open-ended questions to evaluate task-specific user satisfaction regarding what study participants think about the interface and functions of the app as well as any specific issues that they find confusing.
- Networked minds survey to evaluate social presence in terms of communication [32] when the user interacts with the virtual human assistant.
- System usage logs to quantify patient engagement with the system and guide improved app design and development [33] and nonscheduled patient-initiated feedback to indicate patient satisfaction and engagement.
- Focus groups to define the essential expectations and needs that are personally and clinically relevant and to achieve a consensus on the key features and functions of iHeartU and the overall mobile self-management system.

Procedure

The usability assessment will follow the phases below and last for about 17-24 months.

Enrollment Phase

A panel of patients, family caregivers, and health care providers was hosted in the GHS Patient Engagement Studio to provide initial thoughts and suggestions on the preliminary product of iHeartU. A meeting with a group of health care providers for patients with HF was also held during the prototype development for advice on the design. The eligible patients with HF were introduced for enrollment by their direct care team at GHS.

Intervention Phase

HF patient representatives and health care providers who care for patients with HF complete the demographic and background questionnaire, navigate the features of iHeartU, and test its functions along with the use of supplemental devices. Users perform predefined tasks by the research team and conduct interactive conversations with the iHeartHelper to input clinical variables: weight; systolic and diastolic blood pressure; heart rate; and subjective reports for medication adherence, dietary sodium and fluid intake, and physical activity. In case of any technical difficulty, patients can always choose to enter the data manually. This assessment is conducted by using usability evaluation metrics, the think-aloud protocol, individual interviews, and networked minds survey. The smartphone and supplemental devices are provided by the research team.

Follow-Up Phase

Iterative usability testing with the improved iHeartU app is conducted and evaluated among patients after their use at home. The usability regarding patient engagement is examined by using system usage logs and the User Engagement Scale. A focus group discussion is organized with the patient participants to further identify the essential needs and desirable features and functions of iHeartU to improve patient satisfaction and engagement.

Data Analysis

Quantitative analysis will be performed on the survey data. The descriptive statistics will be summarized on patients' background and experience. The significance of overall System Usability Scale [26] will be assessed by using the Student t test [37,38]. The User Engagement Scale [36] and Networked Minds survey [32] will be used for analysis. The significance of iHeartU use will also be assessed by comparing the ratio of observed use to expected use to the hypothetical mean of 1.0 by using the Student t test [38]. The interrater agreement of raw system usability scores will be assessed by the intraclass correlation coefficient with 95% CIs [39]. The Cronbach alpha test will be performed on the data to evaluate the internal consistency of the responses.

Table 1. Usability evaluation metrics.

Aim	Measure	Method
Effectiveness	• Task completion: with ease, with minor mistakes, or failure • Error coding	• Direct observation
Efficiency	• Time used to complete a task	• Count
Satisfaction	• System Usability Scale [34] • Poststudy Usability [35]	• Questionnaire
Engagement	• User Engagement Scale [36] • System usage log	• Questionnaire • Built-in tracker

XSL•FO
RenderX

Qualitative analysis will be performed on the data collected through think-aloud protocol, individual interviews, and focus groups by using the Grounded Theory to identify emerging themes directly from patients' own words and thoughts. Each think-aloud protocol will be reviewed line-by-line by two researchers in order to reach a consensus on the coding. The ease of use will be assessed and coded by scrutinizing the recorded details (such as facial expression and finger movement) of patients' performance on using iHeartU. Individual interviews and focus group discussions will be analyzed using the constant comparative method. The transcript will be reviewed and coded for recurrent themes independently by two researchers. Themes will be combined by agreement of the two researchers who will create theme tags based on the recurrent themes, and the transcripts will be rereviewed collaboratively by the two researchers in order to tag all instances of the themes. Major themes and subthemes will be developed via an iterative review process.

Results

Development of iHeartU

The development of iHeartU started in May 2017 with developmental approval from the Office of Compliance at Clemson University. Institutional review board approval was obtained (#Pro00066413) from the GHS in October 2017. The prototype of the mobile phone app has been produced. The screenshot of its interface with the iHeartHelper is indicated in Figure 2. A variety of outfits for the iHeartHelper and background options are available for random picking every time when the user opens the app. Key self-care behaviors, including adherence to prescribed HF therapies such as medications, dietary sodium and fluid restrictions, and exercise, are built into the daily checkup page [40]. Besides these HF therapies, patients are taught to recognize and manage changes in symptoms and seek health advice when such changes occur.

The iHeartHelper proactively engages the patient at times prescribed by the clinician; requests the patient to provide a report of their general progress, medications, activity, and other behavioral aspects via natural dialogue; and records the patient's responses as audio files. Real-world interactions between physicians and patients were shadowed at the GHS Heart Failure Clinic by the research team and used to enrich the conversation scenarios with the iHeartHelper. The sample scripts are demonstrated in Textbox 1. A closed caption is installed to meet the needs of patients who may have hearing impairment. The iHeartHelper's speech speed can be customized as per the patients' preference.

The iHeartHelper serves as a health assistant who checks on the health of the patient three times daily and records vital signs, diet, fluid intake, weight, and any symptom that the patient would like to report. The objective data are manually input by the patient in the current development phase, with error checking for a numeric format. Patients receive a push notification to remind them to open the app when it passes the preset time. Patients have the facility at any time to convey messages and alerts to their health care provider and caregivers by reporting issues to the iHeartHelper through audio recording. Two representative screenshots are illustrated in Figures 3 and 4.

Audio files as well as other data can be periodically uploaded by a reporting module to a central data repository, which allows the health care provider to monitor the patient's progress as needed via a Web-based interface (Figure 5). The Web portal for health care providers has also been developed and linked with the iHeartHelper (Figure 6). The audio files and any information (eg, vital signs and weight) recorded by the patient can be accessed by the care team (eg, physician, physician assistant, nurse, and disease manager). Once the patient's data exceed the predefined thresholds with respect to the clinical guideline and physician's suggestions, the health care provider will be alerted. According to the severity of the situation, the health care provider can send a message to the patient through the portal to give him/her suggestions or schedule a phone call or visit. The message is delivered to the patient by the iHeartHelper. There are two-stage confirmations to ensure that the message is delivered to the patient.

Figure 2. Example of the interface of iHeartHelper.

XSL•FO

RenderX

Textbox 1. Sample scripts of the conversation scenarios between the virtual human assistant (iHeartHelper) and the patient.

Have you had anything to eat since we last talked?

- Yes - What did you have to eat?

- No - Please be sure to eat three times daily!

What fluids have you had to drink today and how much of each did you drink? Have you taken your medication for this meal yet?

- Yes - Great!

- No - Please remember to take your medications at every meal.

Have you done any exercise since our last session?

- Yes - what kinds of activities did you do?

- No - Alright, try to do a little bit of exercise every day.

Have you felt any shortness of breath since we last spoke?

- Yes - Can you please describe the shortness of breath and when it happened?

- No - Okay, good! Remember to always alert your doctor immediately if you feel any shortness of breath.

Do you have any other symptoms you would like your doctor to be aware of?

- Yes - What are the symptoms that are worrying you?

- No - Okay, please keep your doctor informed if you experience any unexpected symptoms.

Figure 3. Example of daily check-up and monitoring by the iHeartHelper.

Figure 4. Example of follow-up and notifications by the iHeartHelper.

Figure 5. Data transfer structure for iHeartU patient-centered mobile self-management system. CG: caregiver.
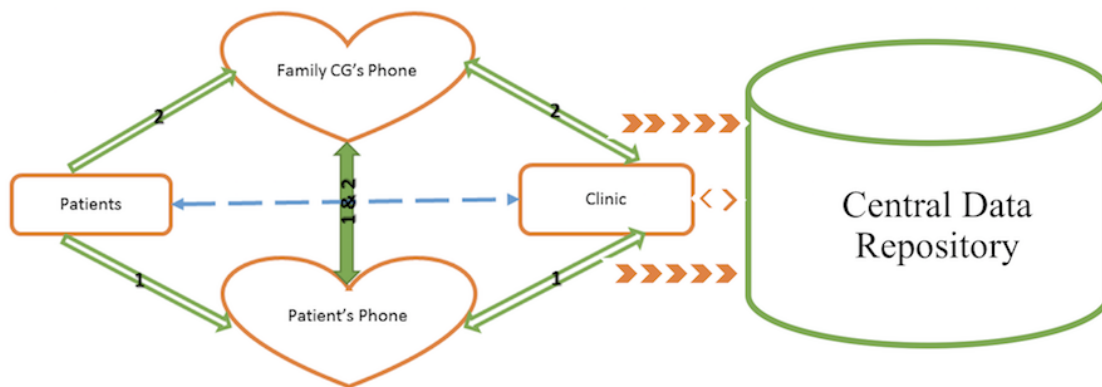
Usability Assessment

Two panel meetings with all key stakeholder representatives, including patients, family caregivers, and health care providers, have already been conducted to collect inputs and suggestions at the GHS Patient Engagement Studio before and during the development of iHeartU. The feedback was also solicited from a meeting with health care providers in the Department of Internal Medicine and the Department of Cardiology at GHS. The suggestions from the stakeholders and health care providers were incorporated into the iHeartU development for usability testing. Enrollment of study participants began in January 2018.

As of September 2018, the preliminary data collection in the intervention phase has been completed. The follow-up phase is expected to be completed and the entire usability testing is expected to be concluded by the end of 2019. The data will be analyzed accordingly, and the results will be submitted for publication upon completion of the study. Two representatives of health care providers will be invited to conduct a usability assessment following the same process as patients and provide inputs from their perspectives. Their feedback will serve as important supplemental information to guide further improvement on iHeartU. Thus far, testing has been completed by a physician in the direct care team.

Figure 6. Screenshot of the Web portal for health care providers.

Discussion

Uniqueness of iHeartU

The main contribution of this work is the development of an HF patient–centered self-management system with a virtual human interface, which may support patients' self-care at home and strengthen communications among patients, health care providers, and family caregivers for better care transition and coordination. To our knowledge, this is the first mobile phone app with a virtual human interface for HF self-management.

There are several foreseeable advantages to this design of iHeartU. First, the virtual human assistants can carry out conversations proactively to record a daily log for patients' self-management behaviors. Since most patients with HF are older adults with relatively lower rates of technology adaption, the intuitive voice-based data collection and easy-to-use features are better than the features of traditional mobile apps that require patients to manually input data. The interactive process aims to resemble the clinical workflow in reality but be more proactive with closer monitoring on patients. Second, the virtual human interface may provide more psychological comfort and companion to patients with HF by sharing news, weather, and jokes with patients to make daily conversions more engaging and socially motivating. Third, the virtual human interface provides personalized health coaching through virtual "face-to-face" communications using education materials tailored for patients' individual situation. Fourth, the virtual

human interface serves as a digital personal assistant to remind patients of their daily medication use, which may improve patients' compliance to HF therapies. This virtual assistant can also remind patients of their physician office visits and facilitate their follow-up visits. This feature will be expected to expand their accessibility to care, improve care coordination, and reduce the cost incurred due to unnecessary emergency visits or readmissions, especially for low-income patients with HF. Fifth, data collection via this mobile self-management system in community settings may lay the groundwork for future clinical studies and big data analysis.

Further Development of iHeartU

Besides the embedded functions described above, the Bluetooth-enabled devices (eg, weight scale and blood pressure monitor) will be integrated into iHeartU. For patients with these Bluetooth sensors, data will be automatically transmitted to iHeartU. They can also choose to record their data with validation and error checking. The manual input option will be still available as a backup.

A personalized customization on the iHeartHelper such as gender, race, and age will be developed. The entertainment features will also be embedded in the app. The iHeartHelper can talk about jokes and report the weather and news as per the user's commands. Such interaction may resemble the interaction of the user with Alexa and Siri. The iHeartHelper will put on an outfit according to the weather change as well, for instance, wearing a rain coat when it is raining.

During the usability assessment, besides testing the close-to-finish product, a list of desirable features will be obtained from the end users to know their thoughts and solicit feedbacks. Thereafter, further development and improvement will be undertaken and tailored to the needs of patients based on the outcome of the usability testing.

The Web portal for health care providers will be further enhanced. In addition, the mobile phone app for family caregivers will be developed. The caregiver app will show the real-time patient's data (with the patient's consent) and receive messages from the patient's health care providers, so that he/she can help monitor the patient's health. An overall assessment will be conducted on the entire mobile self-management system after its full development.

Conclusions

Findings from the standardized usability testing on iHeartU will be used to refine its preliminary design and contents, which may enhance its implementation and dissemination in the later phases. Such a usability testing model can be translational and adapted to other health management programs through disease-specific customization with the involvement of patients and key stakeholders. Further research is needed to establish a standardized, systematic mHealth usability assessment for chronic HF patient care, which can improve methodological consistency and make it possible to compare findings across different mHealth app evaluations [41].

Acknowledgments

Conflicts of Interest

None declared.

Multimedia Appendix 1

Peer-review report from Greenville Health System.

[PDF File (Adobe PDF File), 76KB - resprot_v8i5e13502_app1.pdf]

References

1. Zan S, Agboola S, Moore S, Parks KA, Kvedar JC, Jethwani K. Patient engagement with a mobile web-based telemonitoring system for heart failure self-management: a pilot study. JMIR Mhealth Uhealth 2015 Apr 01;3(2):e33 [FREE Full text] [doi: 10.2196/mhealth.3789] [Medline: 25842282]

2. Fingar K, Barrett M, Jiang H. Healthcare Cost and Utilization Project. 2019. Statistical Brief #230: A Comparison of All-Cause 7-Day and 30-Day Readmissions, 2014 URL: https://www.hcup-us.ahrq.gov/reports/statbriefs/sb230-7-Day-Versus-30-Day-Readmissions.pdf [accessed 2019-05-12] [WebCite Cache ID 78JrDClXE]

3. Roger VL, Go AS, Lloyd-Jones DM, Benjamin EJ, Berry JD, Borden WB, American Heart Association Statistics CommitteeStroke Statistics Subcommittee. Heart disease and stroke statistics--2012 update: a report from the American Heart Association. Circulation 2012 Jan 03;125(1):e2-e220 [FREE Full text] [doi: 10.1161/CIR.0b013e31823ac046] [Medline: 22179539]

4. Bello N, Claggett B, Desai A, McMurray J, Granger C, Yusuf S, et al. Influence of previous heart failure hospitalization on cardiovascular events in patients with reduced and preserved ejection fraction. Circ Heart Fail 2014 Jul;7(4):590-595 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.113.001281] [Medline: 24874200]

5. Benjamin E, Muntner P, Alonso A, Bittencourt MS, Callaway CW, Carson AP, American Heart Association Council on EpidemiologyPrevention Statistics CommitteeStroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. Circulation 2019 Dec 05;139(10):e56-e528. [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]

6. Cibulskis C, Giardino A, Moyer V. Care transitions from inpatient to outpatient settings: ongoing challenges and emerging best practices. Hosp Pract (1995) 2011 Aug;39(3):128-139. [doi: 10.3810/hp.2011.08.588] [Medline: 21881400]

7. Misky G, Wald H, Coleman E. Post-hospitalization transitions: Examining the effects of timing of primary care provider follow-up. J Hosp Med 2010 Sep;5(7):392-397. [doi: 10.1002/jhm.666] [Medline: 20578046]

8. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association

XSL•FO

RenderX

Task Force on practice guidelines. Circulation 2013 Oct 15;128(16):1810-1852. [doi: 10.1161/CIR.0b013e31829e8807] [Medline: 23741057]

9. Moser D, Yamokoski L, Sun J, Conway G, Hartman K, Graziano J, Escape Investigators. Improvement in health-related quality of life after hospitalization predicts event-free survival in patients with advanced heart failure. J Card Fail 2009 Nov;15(9):763-769 [FREE Full text] [doi: 10.1016/j.cardfail.2009.05.003] [Medline: 19879462]

10. Rodríguez-Artalejo F, Guallar-Castillón P, Pascual CR, Otero C, Montes A, García AN, et al. Health-related quality of life as a predictor of hospital readmission and death among patients with heart failure. Arch Intern Med 2005 Jun 13;165(11):1274-1279. [doi: 10.1001/archinte.165.11.1274] [Medline: 15956007]

11. Grady P, Gough L. Self-Management: A Comprehensive Approach to Management of Chronic Conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041]

12. Lorig K, Mazonson P, Holman H. Evidence suggesting that health education for self-management in patients with chronic arthritis has sustained health benefits while reducing health care costs. Arthritis & Rheumatism 1993 Apr;36(4):439-446. [doi: 10.1002/art.1780360403]

13. Lorig K, Ritter P, Villa F, Armas J. Community-Based Peer-Led Diabetes Self-management. Diabetes Educ 2009 Apr 30;35(4):641-651. [doi: 10.1177/0145721709335006]

14. Seto E, Leonard K, Cafazzo J, Barnsley J, Masino C, Ross HJ. Mobile phone-based telemonitoring for heart failure management: a randomized controlled trial. J Med Internet Res 2012 Feb 16;14(1):e31 [FREE Full text] [doi: 10.2196/jmir.1909] [Medline: 22356799]

15. Jaarsma T. Non-Pharmacological Management and Patient Educaiton in Heart Failure Patients. European Cardiovascular Disease 2006:108-110. [doi: 10.15420/ecr.2006.1.108]

16. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do Mobile Phone Applications Improve Glycemic Control (HbA1c) in the Self-management of Diabetes? A Systematic Review, Meta-analysis, and GRADE of 14 Randomized Trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]

17. Athilingam P, Labrador M, Remo E, Mack L, San Juan AB, Elliott A. Features and usability assessment of a patient-centered mobile application (HeartMapp) for self-management of heart failure. Appl Nurs Res 2016 Dec;32:156-163. [doi: 10.1016/j.apnr.2016.07.001] [Medline: 27969021]

18. Pew Research Center: Internet & Technology. 2018. Mobile Fact Sheet URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-05-12] [WebCite Cache ID 78K4Wxu2s]

19. Tatara N, Arsand E, Skrøvseth SO, Hartvigsen G. Long-term engagement with a mobile self-management system for people with type 2 diabetes. JMIR Mhealth Uhealth 2013 Mar 27;1(1):e1 [FREE Full text] [doi: 10.2196/mhealth.2432] [Medline: 25100649]

20. Triantafyllidis A, Velardo C, Chantler T, Shah SPC, Paton C, Khorshidi R, SUPPORT-HF Investigators. A personalised mobile-based home monitoring system for heart failure: The SUPPORT-HF Study. Int J Med Inform 2015 Oct;84(10):743-753. [doi: 10.1016/j.ijmedinf.2015.05.003] [Medline: 26037921]

21. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]

22. Davis F, Bagozzi P, Warshaw P. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]

23. Brunner-La Rocca H, Fleischhacker L, Golubnitschaja O, Heemskerk F, Helms T, Hoedemakers T, et al. Challenges in personalised management of chronic diseases—heart failure as prominent example to advance the care process. EPMA Journal 2016 Jan 30;7(1):2. [doi: 10.1186/s13167-016-0051-9]

24. Chang S, Choi S, Kim S, Song M. Intervention Strategies Based on Information-Motivation-Behavioral Skills Model for Health Behavior Change: A Systematic Review. Asian Nursing Research 2014 Sep;8(3):172-181. [doi: 10.1016/j.anr.2014.08.002]

25. Rothbaum B, Rizzo AS, Difede J. Virtual reality exposure therapy for combat-related posttraumatic stress disorder. Ann N Y Acad Sci 2010 Oct;1208:126-132. [doi: 10.1111/j.1749-6632.2010.05691.x] [Medline: 20955334]

26. Hodges L, Kooper R, Meyer T, Rothbaum B, Opdyke D, de Graaff J, et al. Virtual environments for treating the fear of heights. Computer 1995 Jul;28(7):27-34. [doi: 10.1109/2.391038]

27. Chollet M, Wörtwein T, Morency L, Shapiro A, Scherer S. Exploring feedback strategies to improve public speaking: an interactive virtual audience framework. 2015 Sep Presented at: The ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2015; Osaka, Japan p. 7-11. [doi: 10.1145/2750858.2806060]

28. Cordar A, Robb A, Wendling A, Lampotang S, White C, Lok B. Virtual Role-Models: Using Virtual Humans to Train Best Communication Practices for Healthcare Teams. In: Intelligent Virtual Agents: 15th International Conference, Iva 2015, Delft, The Netherlands, August 26-28, 2015, Proceedings. Cham: Springer International Publishing; 2015.

29. Irizarry T, DeVito Dabbs A, Curran C. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

30. Kushniruk A, Patel V. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform 2004 Feb;37(1):56-76 [FREE Full text] [doi: 10.1016/j.jbi.2004.01.003] [Medline: 15016386]

31.   Virzi R. Refining the Test Phase of Usability Evaluation: How Many Subjects Is Enough? Hum Factors 2016 Nov
       23;34(4):457-468. [doi: 10.1177/001872089203400407]

32.   Biocca F, Harms C, Gregg J. The Networked Minds measure of social presence: Pilot test of the factor structure and
       concurrent validity. 2001 May Presented at: The 4th International Workshop on Presence; 2001; Philadelphia, PA.

33.   Chen E, Cimino J. Automated discovery of patient-specific clinician information needs using clinical information system
       log files. AMIA Annu Symp Proc 2003:145-149 [FREE Full text] [Medline: 14728151]

34.   Brooke J. SUS: A Quick and Dirty Usability Scale. In: Usability Evaluation in Industry. London, UK: Taylor & Francis;
       1996.

35.   Lewis J. Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ. In: Proceedings of the
       Human Factors and Ergonomics Society Annual Meeting. 2016 Aug 06 Presented at: The Human Factors and Ergonomics
       Society Annual Meeting; 1992; Boca Raton, FL p. 1259-1260. [doi: 10.1177/154193129203601617]

36.   O'Brien H, Toms E. Examining the generalizability of the User Engagement Scale (UES) in exploratory search. Information
       Processing & Management 2013 Sep;49(5):1092-1107. [doi: 10.1016/j.ipm.2012.08.005]

37.   Kurosu E. The factor structure of the system usability scale. In: Human Centered Design: First International Conference,
       HCD 2009, Held As Part Of HCI International 2009, San Diego, CA, USA, July 19-24, 2009 Proceedings (Lecture Notes
       In Computer Science). Berlin, Heidelberg: Springer; 2019.

38.   Gardner C, Flanagan M, Franklin C, John-Swayers C, Walsh-Pouch S, Bryant F, et al. Electronic physiologic and subjective
       data acquisition in home-dwelling heart failure patients: An assessment of patient use and perception of usability. Int J Med
       Inform 2016 Dec;93:42-48. [doi: 10.1016/j.ijmedinf.2016.06.001] [Medline: 27435946]

39.   Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline:
       18839484]

40.   Evangelista LS, Shinnick MA. What do we know about adherence and self-care? J Cardiovasc Nurs 2008;23(3):250-257
       [FREE Full text] [doi: 10.1097/01.JCN.0000317428.98844.4d] [Medline: 18437067]

41.   Baig M, Gholamhosseini H. Smart health monitoring systems: an overview of design and modeling. J Med Syst 2013
       Apr;37(2):9898. [doi: 10.1007/s10916-012-9898-z] [Medline: 23321968]

Abbreviations

GHS: Greenville Health System
HF: heart failure
HRQoL: health-related quality of life
ICT: information and communication technology
mHealth: mobile health
QoL: quality of life

XSL•FO

RenderX

<u>Proposal</u>

Optimizing the Impact of Public-Academic Partnerships in Fostering Policymakers' Use of Research Evidence: Proposal to Test a Conceptual Framework

Christina D Kang-Yi[1,2], PhD

[1]Center for Mental Health, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States
[2]Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Christina D Kang-Yi, PhD
Center for Mental Health
Perelman School of Medicine
University of Pennsylvania
3535 Market Street, 3rd Floor
Philadelphia, PA, 19104
United States
Phone: 1 215 746 6715
Email: ckangyi@upenn.edu

Abstract

Background: Previous research has reported that public-academic partnerships (PAPs) can effectively promote PAP leaders' use of research evidence in improving youth outcomes. However, the existing literature has not yet clarified whether and how PAP leaders' use of research evidence evolves along the PAP life cycle and whether PAP partners' concordant perceptions of usefulness of their PAP has an impact on PAP leaders' use of research evidence. Developing a conceptual framework that recognizes the PAP life cycle and empirically identifying contexts and mechanisms of PAPs that promote PAP leaders' use of research evidence from the PAP life cycle perspective are imperative to guide researchers and policymakers to successfully lead PAPs and foster policymakers' use of research evidence for improving youth outcomes.

Objective: Utilizing an integrated framework of organizational life cycle perspective, a social partnership perspective, and a realist evaluation, this study examines the extent to which PAP development and PAP leaders' use of research evidence can be characterized into life cycle stages and identifies PAP contexts and mechanisms that explain the progress of PAPs and PAP leaders' use of research evidence through life cycle stages.

Methods: Recruiting PAPs across the United States that aim to improve mental health and promote well-being of youth aged 12-25 years, the study conducts a document analysis and an online survey of PAPs to inform policymakers and academic researchers on the contexts and mechanisms to increase PAP sustainability and promote policymakers' use of research evidence in improving youth outcomes.

Results: Fifty-three PAPs that meet the recruitment criteria have been identified, and document review of PAPs and participant recruitment for the online survey of PAP experience have been conducted.

Conclusions: This paper will help policymakers and researchers gain a deeper knowledge of the contexts and mechanisms for each PAP life cycle stage in order to optimize PAP leaders' use of research evidence in achieving positive youth outcomes.

International Registered Report Identifier (IRRID):   DERR1-10.2196/14382

KEYWORDS

XSL•FO
RenderX

Introduction

Background

This proposed project aims to develop a conceptual framework to understand dynamic and complex public-academic partnerships (PAPs) and reveal contexts and mechanisms for each PAP life cycle stage in order to optimize PAP leaders' use of research evidence in improving youth mental health and well-being. The proposed project defines a PAP as a partnership between the state and county policymakers (administrators and program directors) and researchers at academic institutes, formed to promote evidence-informed policymaking and practice. PAP leaders are policymakers who influence public identification of problems, design and implement programs, and make policy decisions as administrators and program directors.

This project focuses specifically on PAPs that aim to improve mental health and promote well-being of youth aged 12-25 years. Many psychiatric disorders such as mood disorders, substance abuse problems, and schizophrenia develop during adolescence [1]. About 20% of US youths aged 13-16 years will experience a psychiatric disorder during their lifetime [2-4], and the rates are higher among youth served by the public care sector [5-7]. PAPs are vital to improve the health and well-being of vulnerable populations [8-11]. PAPs seek to bridge the historic divide and disconnect that has evolved among researchers and policymakers and improve the degree to which the knowledge generated by researchers is utilized for the benefit of the individuals being served by the public care sector [12,13]. A critical means through which PAPs accomplish this aim is the use of research evidence. Research evidence is defined as relevant conceptual frameworks or reviews and empirical findings from systematic qualitative, quantitative, or mixed research methods projects [14]. Use of research evidence is defined as acquiring, evaluating, and directly applying research evidence [15,16] and conceptually using research [17,18] to understand the nature of and frame social and community problems; to design and implement public services and programs; and to make policy decisions [19,20].

Previous research has reported that PAPs can effectively promote PAP leaders' use of research evidence in improving youths' outcomes [12,20,10]. However, the existing literature has not yet informed the specific mechanisms of PAPs that promote PAP leaders' use of research evidence in improving youths' outcomes; whether and how PAP leaders' use of research evidence evolves along the PAP life cycle; and whether PAP partners' concordant perceptions of usefulness of their PAP has an impact on PAP leaders' use of research evidence. Developing a conceptual framework that recognizes the PAP life cycle, and empirically identifying contexts and mechanisms of PAPs that promote PAP leaders' use of research evidence from the PAP life cycle perspective is imperative to guide researchers and policymakers to successfully lead PAPs and foster policymakers' use of research evidence in improving youth outcomes. This proposed project will examine the extent to which PAP development and PAP leaders' use of research evidence can be characterized into life cycle stages, and identify PAP contexts and mechanisms that explain the progress of PAPs and PAP leaders' use of research evidence through life cycle stages.

Theoretical and Empirical Rationale

Although a number of empirical studies have helped identify specific factors that are associated with PAP success and failure, there is no research evidence that offers a framework of contexts and mechanisms that should occur at each stage of PAP development in order to maximize the chances of success and increase PAP leaders' use of research evidence. The present project proposes an integrated framework of social partnerships, organizational life cycle, and realist evaluation perspectives to address the complex contexts and corresponding mechanisms that result in successful sustainment of PAPs and foster PAP leaders' use of research evidence, for PAPs that serve youth in the public care sector. This project classifies the PAP life cycle stages into initiated/not initiated, formed/failed to be formed, matured/not matured, and sustained/declined (Figure 1) [21-23].

The social partnership perspective [7,22-28] posits that partnerships have common roots in their intended impact on some societal problems such as education, poverty, and health by building on the capabilities, resources, and expertise of each partner [25,27,28]. This perspective aligns well with PAP development, as PAPs are typically initiated to address social issues such as youth mental health and well-being. A social partnership perspective is that three key partnership processes should occur concurrently to evolve through stages such as initiation, formation, and maturity: issue crystallization; coalition building through mutual benefits and trust, top management support, convener's role, and on-the-spot decision-making power; and purpose formulation through determining structure, goals, primary function, and process of setting agenda [20]. Although the perspective offers a compelling explanation for factors that support successful PAP initiation, formation, and maturity, the framework does not explain how PAP partnerships sustain. *The organizational life cycle perspective* [29-34] offers a longitudinal and sequenced approach to explain how PAPs grow and change over time. Organizations go through stages of birth, maturity, and decline and the goals, priorities, and definitions of organizational effectiveness differ across not only organizations but also these stages within organizations [29,34]. By applying this perspective, we can explain how PAP partnerships transition through life cycle stages and the potential presence of a sustained versus declined stage, which the social partnership perspective is missing. An integration of aspects of the organizational life cycle perspective with social partnership perspective contributes to the conceptualization of PAP progress by distinguishing specific life cycle stages relevant to the progress of organizations. *The realist-evaluation approach* [35-37] helps deepen the conceptualization of how the contexts of PAPs are involved in both their sustainability and use of research evidence. Realist evaluation focuses on three concepts—context, mechanism, and outcomes—which link together to form a context-mechanism-outcome (CMO) configuration (Figure 1). This proposed project adapts the definitions of context, mechanism, and outcome from the study of Jagosh et al [38] who used the realist evaluation approach to understand partnerships in community participatory research.

Figure 1. Proposed conceptual framework to understand public-academic partnerships as social partnerships transitioning through life cycle stages.
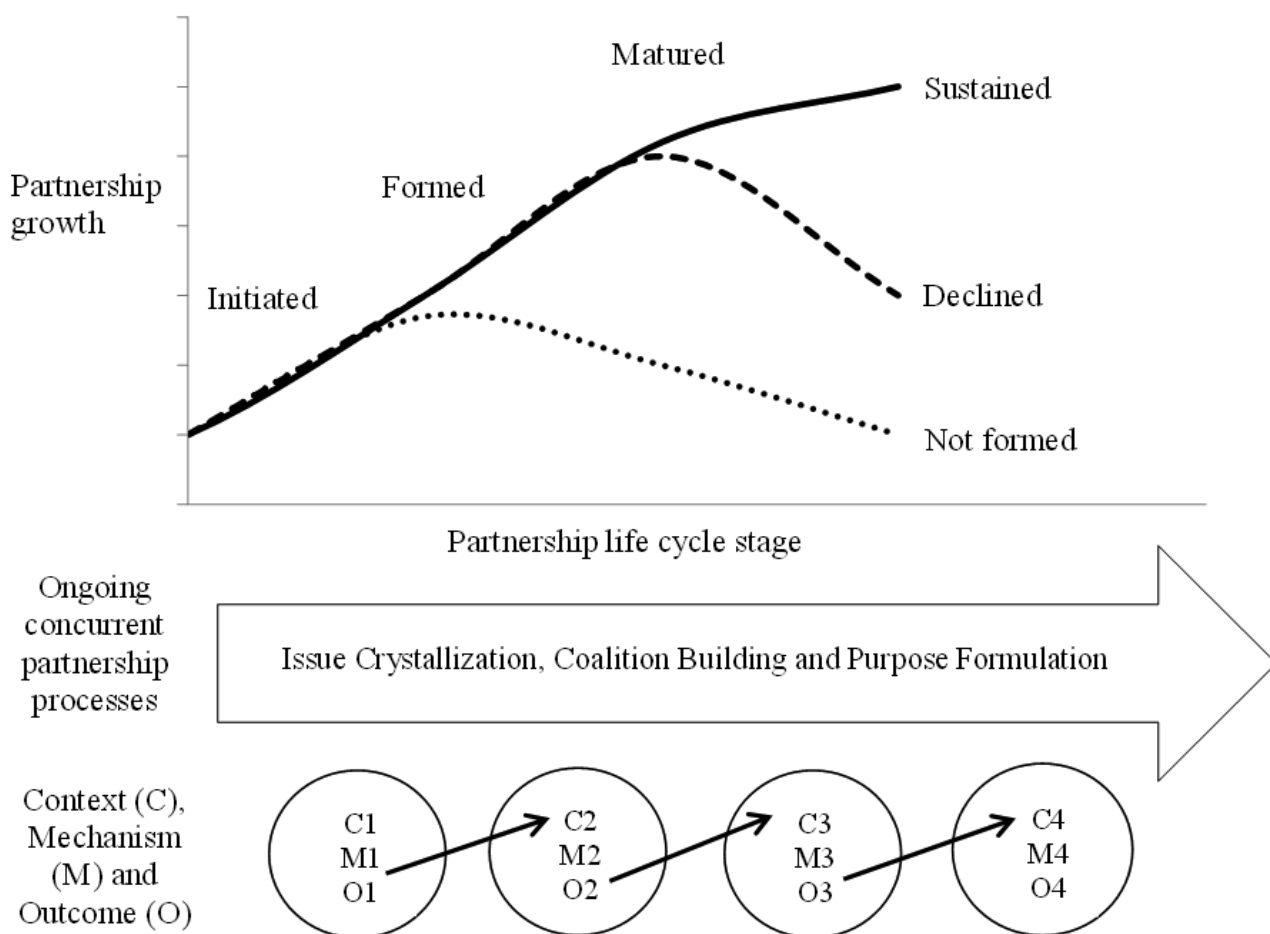
Describing context, mechanism, and outcome relationships based on the realist evaluation framework helps generate and refine an explanatory theory, which, in this case, are the theory of how PAPs evolve and what promotes or inhibits PAP leaders' use of research evidence (Multimedia Appendix 1). The features of a realist evaluation approach can be applied to PAPs to clarify the nature and contributions of contexts and mechanisms to PAP sustainability and PAP leaders' use of research evidence according to the PAP life cycle stage in real-world settings.

Neither social partnerships nor organizational life cycle perspectives recognize the embedding of PAPs in a system and do not offer a means to connect partnership contexts and mechanisms with PAP leaders' use of research evidence. The realist evaluation approach is based on determining not just what approaches work, but more specifically, "what works, for whom, under what circumstances, and why and how" [35-37]. Thus, this perspective helps deepen the conceptualization of how the contexts of PAPs are involved in both their sustainability and use of research evidence.

Considering that PAP leaders' perception of their academic partners and coalition building with their academic partners can have an impact on PAP sustainability [16,25,39], PAP sustainability and PAP leaders' use of research evidence are likely to be associated. However, the lack of empirical findings warrants further studies. The integrated conceptual framework that includes a theory-building process will contribute to further examination of the relationship between PAP sustainability and

PAP leaders' use of research evidence. Extending the previous research [16,25,39], this proposed project seeks to examine the following aspects:

1. Do PAPs go through a life cycle of being initiated, formed, matured, and sustained?
   - Can the partnership processes (issue crystallization, coalition building, and purpose formulation) be traced through the PAP life cycle stages? If yes, how do partnership processes differ by PAP life cycle stage?

2. Which factors promote or interfere in the progression and ultimate sustainability of PAPs?
   - Do PAP partners' perceptions of alignment between PAP purpose formulation and their own organizational purpose formulation differ by PAP life cycle stage?
   - Do PAP partners' perceptions of PAP coalition building differ by PAP life cycle stage?

3. Are there different patterns of use of research evidence (in terms of obtaining, evaluating, and using research evidence) associated with each PAP life cycle?

4. Which PAP factors promote or inhibit PAP leaders' use of research evidence?
   - Does PAP leaders' use of research evidence differ by PAP partners' perceptions of alignment between PAP purpose formulation and their own organizational purpose formulation?

- Does PAP leaders' use of research evidence differ by PAP partners' perceptions of PAP coalition building?

Methods

Sampling and Participant Recruitment

PAPs that are comprised of at least one or more state or local county child welfare or mental health agencies, and one or more academic researchers will be eligible for inclusion. In the proposed project, mental health includes both mental health and substance abuse. The PAPs can be formed on a project/program/intervention basis or as a consortium. PAPs will be included if their aims include improving mental health and well-being outcomes for youth aged 12-25 years. If multiple academic researchers from one academic research institute are working on separate projects/programs/interventions with the same public care agency, these partnerships will be counted as multiple PAPs, as the partnerships may be in different life cycle stages with different contexts and mechanisms that affect PAP leaders' use of research evidence differently. If one academic researcher has multiple projects with one public care agency, multiple projects will be treated as a single PAP. PAPs that were terminated before 2007 will be excluded. PAPs considered to be in an "initiating" or "failed to be formed" stage will be excluded from this recruitment stage, as members of these PAPs are likely to be difficult to identify. PAPs situated outside the United States or focusing on youth outside the United States will be also excluded. A variety of strategies including online search and contacting partners through emails and phone calls will be undertaken to recruit as many eligible PAPs as possible and collect consistent information from all PAPs. If only the PAP leaders or the academic researchers of each PAP agree to participate, we will still include the PAP in the study for supplementary data.

Document Review and Online Survey of Public-Academic Partnership Leaders and Academic Researchers

The project team will invite all PAPs that meet the inclusion criteria to participate and concurrently conduct PAP document review and an online survey of PAP leaders and academic researchers. We will seek PAP documents that include information on PAP structure, goals, primary function, actor roles, process of setting agenda, and funding sources for the PAP processes. The document review will apply the iterative CMO configuration process of the realist evaluation approach [36,38]. An iterative CMO configuration process will be conducted, in which the PAP documents are reviewed by the project team utilizing a review protocol drafted to serve as the guiding tool in the CMO configuration of PAP life cycle based on the potential CMO of each PAP life cycle stage (Multimedia Appendix 1). PAP mechanisms and PAP leaders' use of research evidence by life cycle stage will be detailed through the document review, and the protocol will be refined through multiple rounds of review of PAP documents.

Online survey of PAP leaders and academic researchers will collect data on partners' PAP experience and PAP leaders' use of research evidence. The intent is to obtain over a 75% survey

response rate. PAP Survey 1: The Structured Interview for Evidence Use (SIEU) scale [16] will determine PAP leaders' engagement level of research evidence, which refers to the frequency of using various types of sources for research evidence; PAP leaders' ratings of the importance of evaluating the validity, reliability, and relevance of research evidence; and various factors leading PAP leaders to use/ignore research evidence in deciding to adopt a new program or intervention. The original SIEU scale items will be used. The Input scale (20 items) assesses the source of the research evidence PAP leaders obtain. The Process scale assesses how PAP leaders evaluate the research evidence obtained and includes three subscales of self-assessment for validity and reliability of research evidence (10 items), reliance on others (5 items), and self-assessment for relevance (5 items). The Output scale (20 items) assesses if PAP leaders eventually use research evidence or ignore the evidence. The measurement asks respondents to indicate responses using a Likert-type scale ranging from 1 (not at all) to 5 (all the time) for the items contained in the Input scale, and a 5-point Likert-type scale ranging from 1 (not important) to 5 (very important) for the items contained in the Process and Output scales. The SIEU shows high internal consistency reliability (α=.88) [16]. PAP Survey 2: PAP Experience (Multimedia Appendix 2) will measure PAP leaders' and academic researchers' most recent PAP experience through items that address issue crystallization (clear issue pursued), purpose formulation (structure, goals, primary function, and agenda setting process), coalition building (mutual benefits and trust, top management support, convener's role, and on-the spot decision-making power); PAP partners' perceptions of their PAP life cycle stage; and PAP leaders' use of research evidence. The online surveys will be built in and administered through the Research Electronic Data Capture, a secure Web-based data collection tool that includes data entry forms and Web surveying features.

Analysis

For the document review, the principal investigator and a research assistant will first independently classify the life cycle for each PAP (formed, but not yet matured; matured, but not reached a sustained stage yet; and sustained/declined) by applying the potential CMO of each PAP life cycle stage (Multimedia Appendix 1) and then review the classification of PAPs until a consensus is reached on the classification. Information on PAPs such as partnership structure, goals, and primary function can vary depending on the available documentation. For missing or incomplete data during the document review process, the project team will follow-up with academic researchers and PAP leaders through emails and phone calls to request and obtain the missing information. The online survey data will also complement the missing data from the document review as the domains of the survey questionnaire are consistent with those of the document review protocol.

Concurrently, the online survey data on PAP leaders' use of research evidence and experience with PAPs will be analyzed in relation to the CMO configuration process. For the online survey, reliability of the SIEU will be calculated by using the Cronbach α internal consistency for each of the subscales and the total scale. We will descriptively test for mean differences

in PAP leaders' engagement level of research evidence by (1) PAP partners' rating of the level of alignment between PAP structure, goals, primary function, and process of setting agenda and their organizational structure, goals, primary function, and process of setting agenda; (2) PAP partners' rating of the level of mutual benefit and trust, top management support, convener's role, and on-the-spot decision-making power; and (3) PAP life cycle stage (formed, but not yet matured; matured, but not reached a sustained stage yet; and sustained/declined). Concordance levels of PAP leaders' and academic researchers' perceptions of PAP contexts and mechanisms will be calculated according to the PAP life cycle stage using Cohen κ coefficients and McNemar test [40]. Pearson product-moment correlations of the concordance levels and SIEU subscale and total scores will test the relationship between the concordance of partner's perceptions and PAP leaders' use of research evidence. The potential CMO of each PAP life cycle stage (Multimedia Appendix 1) will be refined through multiple rounds of review of PAP documents and quantitative data for the development of a middle-range theory.

Results

Fifty-three eligible PAPs have been identified, document review of 20 PAPs have been conducted, and 16 PAP researchers have been reached out for additional information. The principal investigator and the research assistant are in the process of classifying the life cycle for each PAP based on the document review by applying the potential CMO of each PAP life cycle stage (Multimedia Appendix 1). The classification of PAPs will be continued until consensus is reached on the classification. Concurrently, the project team is recruiting PAP leaders and academic researchers who will participate in the online survey and will conduct analysis of CMO PAP life cycle stages and its relationship to PAP leaders' use of research evidence.

Discussion

The proposed project is expected to help policymakers and researchers gain a deeper knowledge of the contexts and mechanisms for each PAP life cycle stage in order to optimize PAP leaders' use of research evidence in achieving positive youth outcomes. Although we will focus on youth mental health and well-being, our findings are likely to be relevant to other vulnerable populations. Future studies should include PAPs in an "initiating" or "failed to be formed" stages, as the PAPs are likely to provide valuable learning about attempted partnerships.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Potential Context, Mechanism and Outcome (CMO) of each Public-Academic Partnership (PAP) life cycle stage.

[PDF File (Adobe PDF File), 576KB - resprot_v8i5e14382_app1.pdf]

Multimedia Appendix 2

Public-Academic Partnership (PAP) Survey Questionnaire.

[PDF File (Adobe PDF File), 81KB - resprot_v8i5e14382_app2.pdf]

Multimedia Appendix 3

Peer-reviewer report from the William T Grant Foundation.

[PDF File (Adobe PDF File), 152KB - resprot_v8i5e14382_app3.pdf]

References

1. Locke J, Kang-Yi C, Pellecchia M, Marcus S, Hadley T, Mandell DS. Ethnic Disparities in School-Based Behavioral Health Service Use for Children With Psychiatric Disorders. J Sch Health 2017 Dec;87(1):47-54 [FREE Full text] [doi: 10.1111/josh.12469] [Medline: 27917490]
2. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch Gen Psychiatry 2003 Aug;60(8):837-844. [doi: 10.1001/archpsyc.60.8.837] [Medline: 12912767]

XSL•FO

RenderX

3.   Lewinsohn P. Major depressive disorder in older adolescents: Prevalence, risk factors, and clinical implications. Clinical Psychology Review 1998 Nov;18(7):765-794 [FREE Full text] [doi: 10.1016/S0272-7358(98)00010-5]

4.   Merry SN, Hetrick SE, Cox GR, Brudevold-Iversen T, Bir JJ, McDowell H. Psychological and educational interventions for preventing depression in children and adolescents. Cochrane Database Syst Rev 2011 Dec 07(12):CD003380. [doi: 10.1002/14651858.CD003380.pub3] [Medline: 22161377]

5.   Courtney M, Okpych N, Charles P, Mikell D, Stevenson B, Park K, et al. Findings from the California Youth Transitions to Adulthood Study (CalYOUTH): Conditions of Youth at Age. Chicago, IL: Chapin Hall at the University of Chicago; 2016:19-3125.

6.   Havlicek J, Garcia A, Smith DC. Mental Health and Substance Use Disorders among Foster Youth Transitioning to Adulthood: Past Research and Future Directions. Child Youth Serv Rev 2013 Jan;35(1):194-203 [FREE Full text] [doi: 10.1016/j.childyouth.2012.10.003] [Medline: 23766549]

7.   Gricar B, Brown L. Conflict, Power, and Organization in a Changing Community. Human Relations 2016 Apr 22;34(10):877-893 [FREE Full text] [doi: 10.1177/001872678103401004]

8.   Finnerty M, Neese-Todd S, Bilder S, Olfson M, Crystal S. Best Practices: MEDNET: a multistate policy maker-researcher collaboration to improve prescribing practices. Psychiatr Serv 2014 Nov 01;65(11):1297-1299 [FREE Full text] [doi: 10.1176/appi.ps.201400343] [Medline: 25756882]

9.   Livingood WC, Goldhagen J, Little WL, Gornto J, Hou T. Assessing the status of partnerships between academic institutions and public health agencies. Am J Public Health 2007 Apr;97(4):659-666. [doi: 10.2105/AJPH.2005.083188] [Medline: 17329657]

10. Srihari VH, Breitborde NJK, Pollard J, Tek C, Hyman L, Frisman LK, et al. Public-academic partnerships: early intervention for psychotic disorders in a community mental health center. Psychiatr Serv 2009 Nov;60(11):1426-1428 [FREE Full text] [doi: 10.1176/ps.2009.60.11.1426] [Medline: 19880454]

11. Stirman SW, Buchhofer R, McLaulin JB, Evans AC, Beck AT. Public-academic partnerships: the Beck Initiative: a partnership to implement cognitive therapy in a community behavioral health system. Psychiatr Serv 2009 Oct;60(10):1302-1304 [FREE Full text] [doi: 10.1176/ps.2009.60.10.1302] [Medline: 19797367]

12. Hansen E, Seybolt D, Sundeen S. Building a successful public-academic partnership to support state policy making. Psychiatr Serv 2014 Jun 01;65(6):710-712 [FREE Full text] [doi: 10.1176/appi.ps.201400067] [Medline: 24881682]

13. Jacobs S, Steiner J. The Yale Textbook of Public Psychiatry. New York: Oxford University Press; 2016.

14. Tseng V. The uses of reasrch in policy and practice. Sharing Child and Youth Development Knowledge 2012;26:1-23.

15. Nutley S, Walter I, Davies H. Promoting Evidence-based Practice. Research on Social Work Practice 2009 May 27;19(5):552-559 [FREE Full text] [doi: 10.1177/1049731509335496]

16. Palinkas LA, Garcia AR, Aarons GA, Finno-Velasquez M, Holloway IW, Mackie TI, et al. Measuring Use of Research Evidence: The Structured Interview for Evidence Use. Res Soc Work Pract 2016 Sep;26(5):550-564 [FREE Full text] [doi: 10.1177/1049731514560413] [Medline: 27616869]

17. Farrell C, Coburn C. William T Grant Foundation. 2016. What is the conceptual use of research, and why is it important? URL: http://wtgrantfoundation.org/conceptual-use-research-important [accessed 2019-05-15] [WebCite Cache ID 78P2MWQ11]

18. McBeath B, Jolles MP, Carnochan S, Austin MJ. Organizational and Individual Determinants of Evidence Use by Managers in Public Human Service Organizations. Human Service Organizations: Management, Leadership & Governance 2015 Jun;39(4):267-289. [doi: 10.1080/23303131.2015.1044588]

19. Walter I, Nutley S, Davies H. What works to promote evidence-based practice? A cross-sector review. Evidence & Policy: A Journal of Research, Debate and Practice 2005 Sep 01;1(3):335-364 [FREE Full text] [doi: 10.1332/1744264054851612]

20. Palinkas L, Short C, Wong M. William T Grant Foundation. Research-practice-policy partnerships for implementation of evidence-based practices in child welfarechild mental health URL: http://wtgrantfoundation.org/library/uploads/2015/10/Research-Practice-Policy_Partnerships.pdf [accessed 2019-05-15] [WebCite Cache ID 78P2ddM0d]

21. Sysney L. Organizational Physics. 2011. Organizational development stability chart URL: http://www.organizationalphysics.com/ [accessed 2017-12-07] [WebCite Cache ID 78P35DyLe]

22. Waddock SA. Understanding Social Partnerships. Administration & Society 2016 Jul 26;21(1):78-100 [FREE Full text] [doi: 10.1177/009539978902100105]

23. Jagosh J, Bush PL, Salsberg J, Macaulay AC, Greenhalgh T, Wong G, et al. A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. BMC Public Health 2015 Jul 30;15:725 [FREE Full text] [doi: 10.1186/s12889-015-1949-1] [Medline: 26223523]

24. Austin J, Whitehead J, Hesselbein F, Whitehead J. Collaboration Challenge: How Nonprofits and Businesses Succeed Through Strategic Alliances. New York: John Wiley & Sons, Incorporated; 2000.

25. Selsky JW, Parker B. Cross-Sector Partnerships to Address Social Issues: Challenges to Theory and Practice. Journal of Management 2016 Jul;31(6):849-873. [doi: 10.1177/0149206305279601]

26. Warner M, Sullivan R. Putting partnerships to work: Strategic alliances for development between government and private sector and civil society. New York: Routledge; 2004.

27.   Wymer W, Samu S. Dimensions of Business and Nonprofit Collaborative Relationships. Journal of Nonprofit & Public Sector Marketing 2003 Feb 03;11(1):3-22 [FREE Full text] [doi: 10.1300/J054v11n01_02]

28.   Seitanidi MM, Koufopoulos DN, Palmer P. Partnership Formation for Change: Indicators for Transformative Potential in Cross Sector Social Partnerships. J Bus Ethics 2011 Feb 22;94(S1):139-161. [doi: 10.1007/s10551-011-0784-2]

29.   Quinn R, Cameron K. Organizational Life Cycles and Shifting Criteria of Effectiveness: Some Preliminary Evidence. Management Science 1983 Jan;29(1):33-51 [FREE Full text] [doi: 10.1287/mnsc.29.1.33]

30.   Kimberly J, Miles R. The organizational life cycle: Issues in the creation, transformation, and decline of organizations. San Francisco, CA: Jossey-Bass Publishers; 1980.

31.   Tushman M, Romanelli E. Organizational evaluation: A metamorphosis model of convergencereorientation. In: Research in organizational behavior. Greenwich, CT: JAI Press; 1985:171-222.

32.   Miller D. Evolution and revolution: A quantum view of structural change in organizations. J Management Studies 1982 Apr;19(2):131-151. [doi: 10.1111/j.1467-6486.1982.tb00064.x]

33.   Miller D, Friesen P. Archetypes of Organizational Transition. Administrative Science Quarterly 1980 Jun;25(2):268-299. [doi: 10.2307/2392455]

34.   Whetten DA. Organizational Growth and Decline Processes. Annu Rev Sociol 1987 Aug;13(1):335-358. [doi: 10.1146/annurev.so.13.080187.002003]

35.   Pawson R. Evidence-based policy: A realist perspective. Thousand Oaks, CA: Sage Publications; 2006.

36.   Pawson R. The science of evaluation: A realist manifesto. London, UK: Sage Publications; 2013.

37.   Pawson R, Tilley N. Realist evaluation. London, UK: Sage Publications; 1997.

38.   Jagosh J, Macaulay A, Pluye P, Salsberg J, Bush PL, Henderson J, et al. Uncovering the benefits of participatory research: implications of a realist review for health research and practice. Milbank Q 2012 Jun;90(2):311-346 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00665.x] [Medline: 22709390]

39.   Ellen M, Horowitz E, Vaknin S, Lavis JN. Views of health system policymakers on the role of research in health policymaking in Israel. Isr J Health Policy Res 2016;5:24 [FREE Full text] [doi: 10.1186/s13584-016-0088-1] [Medline: 27330738]

40.   Baggott C, Cooper B, Marina N, Matthay KK, Miaskowski C. Symptom assessment in pediatric oncology: how should concordance between children's and parents' reports be evaluated? Cancer Nurs 2014;37(4):252-262. [doi: 10.1097/NCC.0000000000000111] [Medline: 24936750]

Abbreviations

CMO: context-mechanism-outcome
PAP: public-academic partnership
SIEU: Structured Interview for Evidence Use

Protocol

An Intervention for Changing Sedentary Behavior Among African Americans With Multiple Sclerosis: Protocol

Jessica F Baird[1], PhD; Jeffer Eidi Sasaki[2], PhD; Brian M Sandroff[1], PhD; Gary Cutter[1], PhD; Robert W Motl[1], PhD

[1]University of Alabama at Birmingham, Birmingham, AL, United States
[2]Federal University of Triangulo Mineiro, Uberaba, Brazil

Corresponding Author:
Robert W Motl, PhD
University of Alabama at Birmingham
1716 9th Ave S
Birmingham, AL, 35233
United States
Phone: 1 205 934 5905
Email: robmotl@uab.edu

Abstract

Background: Sedentary behavior is a major concern among patients with multiple sclerosis (MS), as it may accelerate disease progression and exacerbate physical disability. This is especially concerning among African Americans, a segment of the MS population who present with greater neurological disability and higher odds of physical comorbidities than their Caucasian counterparts.

Objective: To date, researchers have not proposed interventions that focus on changing sedentary behavior in African Americans with MS.

Methods: This paper describes a pilot study that examines the feasibility and efficacy of using text messaging along with theory-driven newsletters and behavioral coaching for changing sedentary behavior in African Americans with MS. We herein present the methods, procedures, and outcomes for our ongoing study.

Results: Enrollment began in February 2018 and is expected to conclude in April 2019. Study results will be reported in the fall of 2019.

Conclusions: After completion of this pilot intervention, we will summarize our study results in manuscripts for publication in peer-reviewed journals that will provide critical information on the feasibility and efficacy of our strategy. These results will inform future studies and, potentially, larger interventions for remotely reducing sedentary behavior in African Americans with MS.

Trial Registration: ClinicalTrials.gov NCT03671499; https://clinicaltrials.gov/ct2/show/NCT03671499 (Archived by WebCite at http://www.webcitation.org/77MZnxyNy)

International Registered Report Identifier (IRRID): DERR1-10.2196/12973

Introduction

Sedentary behavior is defined as any waking activity performed in a seated or lying position with energy expenditure ≤1.5 metabolic equivalents of task (MET), with 1 MET being the metabolic rate at rest [1]. Sedentary behavior represents a major public health concern based on associations with morbidity and mortality [2,3], independent of physical activity [1].

Sedentary behavior has recently received attention among persons with the chronic, disabling disease, multiple sclerosis (MS) [4]. Sedentary behavior is two times higher in MS than in the general population, and sitting time increases across levels of worsening MS-related mobility disability [5]. Such an association likely involves sedentary behavior inducing comorbid conditions [6] that can accelerate disease progression [4] and ultimately worsen MS disability over time. This may be particularly pertinent for African Americans with MS. This

group demonstrates worse neurological disability [7] and increased odds of physical comorbidities compared with Caucasian counterparts [8]. There is evidence that African Americans with MS experience a more aggressive disease course and present a blunted response to disease modifying therapies [9]. There are no data on racial differences in sedentary behavior in MS, but evidence in the general population indicates that African Americans spend more time sitting than Caucasians [10].

We recently reported that only 1.7% of participants in studies of exercise and physical activity in persons with MS were African American [11]. Furthermore, there is no research on the management of MS through reduction of sedentary behavior in African Americans with MS [12]. This may be associated with barriers toward implementing behavior change interventions in African Americans, such as diminished health care accessibility based on socioeconomic disparities and greater likelihood of living in low-income neighborhoods [13,14] where rehabilitation centers and facilities are often inaccessible. The use of electronic technology may be part of a solution for this problem and represents a strategy that can be applied for increasing the reach of behavior change interventions by allowing intervention delivery remotely, reducing participant burden, offering access, and reducing implementation costs and research personnel burden [15]. Accordingly, text messaging represents a promising medium for reaching a large number of persons remotely, as 94% of Americans own a mobile phone [16]. This strategy has already been used for promoting changes in health behaviors (eg, eating habits and medication use) and preventing chronic diseases (eg, cardiovascular disease and diabetes) in other populations [17-23]. We therefore believe that change in sedentary behavior might be facilitated by text messaging and newsletters, supplemented with one-on-one behavioral coaching. Studies have further reported that text messaging interventions have resulted in sedentary behavior reductions and increases in physical activity in the general population [17,18,24,25], yet we are unaware of any studies using text messaging to reduce sedentary behavior in African Americans with MS.

Text messaging is a promising strategy for reducing sedentary behavior in MS, and the literature indicates that anchoring intervention content with a behavior change theory likely increases the chance of success [26]. Our group has demonstrated that Social Cognitive Theory (SCT) has been effective in promoting changes in sedentary behavior and physical activity among persons with MS [27-29]. SCT provides a good framework for guiding changes in health behavior, as it includes well-tested principles and assumptions from different fields that investigate human behavior, such as psychology, anthropology, and sociology [30]. SCT further identified targets of behavior change, for example self-efficacy, outcome expectations, and goal setting, for designing the content of an intervention.

In view of the current evidence and literature gaps, we developed a technology-based behavior change intervention for reducing sedentary behavior in African Americans with MS. The intervention combines text messaging, print newsletters, and telephone-based coaching, all informed by SCT for reducing

sedentary behavior in African Americans with MS. The proposed intervention will consist of a 12-week program divided into two phases: Phase 1 targets breaking up and reducing sedentary behavior and Phase 2 aims to reduce sedentary behavior by replacing it with light physical activity. To that end, sedentary behavior is the primary outcome of our intervention, but we will secondarily target increases in physical activity as a means of replacing sedentary behavior. The strategy of primarily focusing on a single behavior is based on the need of having a well-defined target for a successful intervention [26,31]. The project will examine the feasibility (ie, process, resources, management, and science) and preliminary efficacy (ie, changes in volume and patterns of sedentary behavior) of this intervention approach for interrupting and reducing sedentary behavior. The lessons learned from this study will inform the development of future large-scale interventions targeting reductions of sedentary behavior in African Americans with MS. The current protocol paper reports on the development, methodology, and outcome measures of this ongoing project.

Methods

Study Design and Overview

The proposed study takes place at the University of Alabama at Birmingham (UAB), United States, and was designed based on guidelines for feasibility studies [32]. The study will use a single-group, pre-post intervention design that examines multiple domains of feasibility, including scientific efficacy of a 12-week, technology-based behavior change intervention for reducing sedentary behavior among African Americans with MS (ie, the *Sit Less, Move More* program). Feasibility will be assessed under four distinct domains: process, resources, management, and scientific metrics [33-35]. The primary scientific metric will be a change in sedentary behavior, measured both objectively with activity monitors and subjectively with self-report questionnaires. Secondary outcomes will include a change in physical activity and health-related quality of life. We will recruit a sample of 30 African American adults with MS. After being screened for eligibility, participants will be sent the informed consent document and the baseline assessment materials—questionnaires and activity monitors—via postal mail along with a preaddressed, prestamped envelope for returning the materials. After the baseline (T1) assessment, participants will start the 12-week intervention, which is divided into two phases: Phase 1 (weeks 1-6) targets breaking up and reducing sedentary behavior and Phase 2 (weeks 7-12) targets replacing sedentary behavior with physical activity. The intervention itself was developed through stakeholder involvement and consists of daily text messages and biweekly newsletters and telephone calls with a behavioral coach. The content of the text messages, newsletters, and coaching sessions will include SCT-based strategies for behavior change, as these approaches have successfully induced behavior change in home-based programs among persons with MS [30,36,37]. Midpoint (T2) and postintervention (T3) assessments will be conducted during week 6 and immediately after week 12, respectively. Participant flow from recruitment through completion of the program is depicted in Figure 1. Participants will be compensated US $25 for completing each of the three

outcome assessment periods. Ethical approval to undertake the study has been obtained from the UAB Institutional Review Board, and the study has been registered with ClinicalTrials.gov (NCT03671499).

We will identify potential participants from those who have contacted our laboratory regarding research opportunities and who meet our inclusion criteria, outlined below, regarding race and age. An email will be distributed among those potential participants, followed by a phone call among those who do not respond to the initial email. We will also recruit individuals from the community at events sponsored by regional chapters of the National MS Society and other local organizations. If necessary, based on enrollment success, mass emails providing information about the study will be sent through the National MS Society, the North American Research Committee on Multiple Sclerosis registry, and iConquerMS. Those who are interested in participating will be instructed to call the laboratory, and authorized personnel will describe the study objectives and procedures. If a prospective participant continues to express interest in the study, we will conduct a telephone screening to ensure that the prospective participant satisfies the inclusion and exclusion criteria. Based on previous feasibility studies in persons with MS, we aim to recruit 30 individuals to participate in this study [38,39]. As this study is designed to assess the feasibility of conducting the Sit Less, Move More program, rather than confirming the efficacy of the intervention, an a priori power analysis for estimating sample size was not performed.

Inclusion criteria for the proposed study are as follows: (1) African American; (2) participant-reported MS diagnosis; (3) relapse-free in the last 30 days; (4) ambulatory with or without assistance based on a Patient-Determined Disease Steps (PDDS) score of 0-5; (5) self-report daily engagement in sedentary behavior of ≥480 minutes per day; (6) health contribution score of <14 calculated from the Godin Leisure-Time Exercise Questionnaire [40]; (7) absence of major musculoskeletal problems and/or cardiovascular, cardiopulmonary, and/or metabolic diseases that are contraindications for changing physical activity and sedentary behavior levels; (8) living in the United States; and (9) ownership of a mobile phone capable of receiving text messages.

Feasibility Metrics

Feasibility will be assessed based on four domains: process (eg, recruitment and retention rates), resources (eg, communication and monetary requirements), management (eg, researcher preparation and capacity), and science (eg, safety and efficacy of the intervention). To assess process-related feasibility, we will document recruitment, retention, and adherence rates. Recruitment will be evaluated as the number of individuals who initially expressed interest in participating and agreed to be screened for eligibility after receiving more information about the study. Retention will be evaluated as the number of participants retained from enrollment through completion of the study. We will maintain a log of the number of participants who complete the biweekly behavioral coaching sessions to determine adherence. To measure resource-related feasibility, we will maintain logs of time spent communicating with participants and of monetary costs involved in the conduction of the study. Management-related feasibility will be evaluated by maintaining a log indicating the efforts and time spent by research personnel related to collecting and entering data, maintaining equipment and research-related material, and managing the study logistics. Safety and efficacy of the intervention will be included as scientific feasibility metrics. To assess safety, we will maintain records of any adverse or serious adverse events and any medical concerns reported by the participants. We will follow standard UAB protocols in the reporting of any such events. The primary scientific outcome is a change in sedentary behavior. Secondary scientific outcomes include changes in physical activity and quality of life (see Behavioral Outcomes section for more detail). After completing the study, the participants will provide feedback regarding satisfaction and personal experiences with the study using a survey with both Likert scale questions and open-ended questions. This information, along with the results of the feasibility analysis, will be valuable for developing larger-scale studies designed to establish the efficacy and effectiveness of this treatment approach and to identify future strategies of effective behavior change.

Figure 1. Flowchart outlining the program timeline from recruitment through completion of the program. MS: multiple sclerosis; ICD: informed consent document; T1: baseline; T2 midpoint; T3: postintervention.
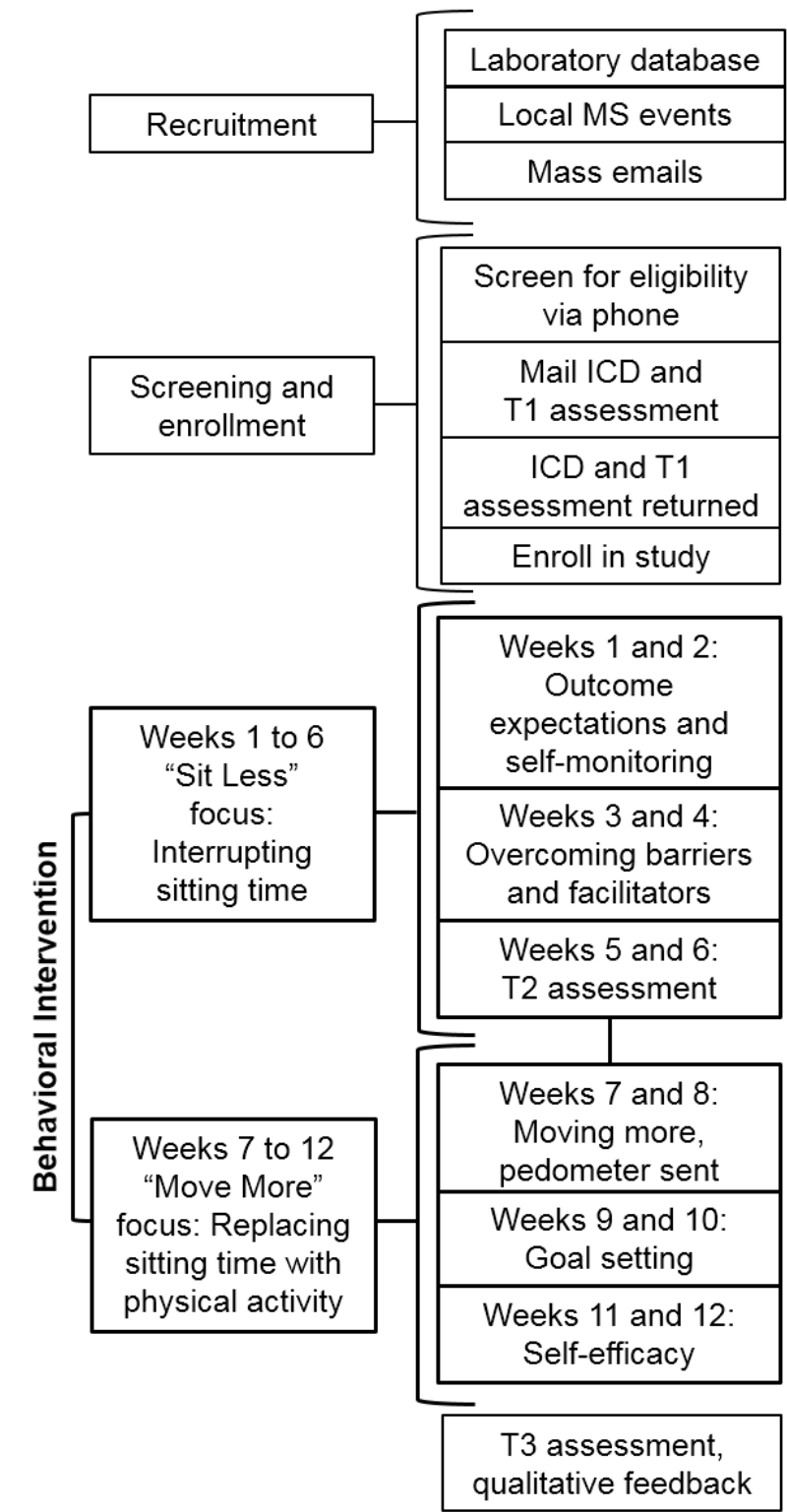
Behavioral Outcomes

Participants will provide information on clinical and demographic characteristics (eg, disease course and duration, education level, and level of income) and disability status (ie, PDDS) during the baseline (T1) assessment period. The primary behavioral outcome will include a change in sedentary behavior, with changes in physical activity and health-related quality of life assessed as secondary outcomes. All outcomes will be measured at three different time points: baseline (T1 assessment), midpoint (T2 assessment), and postintervention (T3 assessment). Sedentary behavior and physical activity behavior will be measured objectively with accelerometry, using the activPAL activity monitor (PAL Technologies) and the ActiGraph model GT3X+ activity monitor (ActiGraph LLC). During each assessment time point, participants will wear both

activity monitors for a 7-day period, only removing them for sleeping or during water-based activities (eg, bathing and swimming). The activPAL monitor will be worn at the middle of the anterior aspect of the thigh and the ActiGraph will be worn on an elastic belt around the waist at the nondominant hip. Standardized instructions describing how to wear the activity monitors, including picture references, will be provided at each assessment time point. Data obtained from the activPAL will be classified as time spent sitting or lying down, standing, or during movement using a proprietary algorithm; these metrics represent the primary sedentary behavior outcome [41]. The activPAL data will further indicate how frequently sitting is interrupted through the number of sit-to-stand transitions. Sedentary behavior measures from the ActiGraph device will supplement measures from the activPAL. The ActiGraph data will be assessed as 60-second epochs and physical activity behavior will be defined by established cut points for persons with MS: sedentary behavior is <100 activity counts per minute, light physical activity is between 100 and 1583 activity counts per minute, and moderate-to-vigorous physical activity is ≥1584 activity counts per minute [42]. The number of minutes spent at each activity level (ie, sedentary vs light physical activity vs moderate-to-vigorous physical activity) will be the primary outcome from the ActiGraph data. In addition, sedentary behavior interruption rate will be assessed from ActiGraph data as the number of transitions from <100 activity counts per minute to ≥100 activity counts per minute. In addition to the objective accelerometry measures, subjective self-report measures of sedentary behavior and physical activity will be completed using the Godin Leisure-Time Exercise Questionnaire [40,43] and the International Physical Activity Questionnaire short version [44,45]. The 36-item Short Form Health Survey (SF-36) [46] will be completed at each measurement time point for self-reported measurement of physical and mental indices of health-related quality of life.

Behavioral Intervention

The intervention will be 12 weeks in duration and divided into two, 6-week phases. Phase 1 (weeks 1-6) will focus on interrupting long periods of sedentary behavior, and Phase 2 (weeks 7-12) will focus on replacing sedentary behavior with light physical activity. Participants will be mailed biweekly (ie, all odd-numbered weeks) SCT-based newsletters that highlight ways of overcoming difficulties in effectively interrupting sedentary behavior and replacing sedentary behavior with physical activity; see Figure 1 for a week-by-week list of SCT-based themes. For example, newsletters focus on anticipating outcomes and setting goals, which are characteristic features of human agency, one of the primary tenets of SCT [30]. We further highlight self-efficacy within the newsletters, as it is a focal determinant of health behavior change within SCT [47,48]. Within 5 days of receiving the newsletter, a trained behavioral coach will call the participant via telephone to discuss the information provided in the newsletter and review strategies for behavior change. Throughout the 12-week intervention period, participants will receive two daily text messages, Monday through Friday, which will be sent in the morning and the afternoon and contain content relevant to the topic being reviewed that week. For example, a week 1 (*Outcome*

Expectations and Self-Monitoring) text message will read, "Expectations are important for sticking with your plan to reduce sitting time. Remember to look up the benefits of sitting less," and a week 7 (*Moving More*) text message will read, "Try to walk around the house during TV commercials." Text messages will be standardized across participants. At the outset of the intervention, participants will be provided with a journal to log behavior and monitor progress. This journal is provided as a self-monitoring tool rather than a method of strictly tracking all sedentary behavior and physical activity. The purpose of the journal is to help the participant (1) recognize how much time is spent sitting, (2) identify opportunities to reduce sitting, and (3) track progress throughout the program. During the first phase of the study (weeks 1-6), participants will be encouraged to record time spent sitting. Immediately following the first phase of the study (ie, beginning of week 7), participants will be provided with a Digi-Walker SW200 pedometer (Yamax) for tracking daily step counts. This is an important study aspect, given that the second phase of the study involves the program transitioning from *sitting less* to *moving more*. To that end, during the second phase of the study (weeks 7-12), participants will be encouraged to record time spent physically active, including the number of steps per day, in the provided journal.

Of note, while developing the study materials, we engaged in the process of stakeholder feedback through an informal focus group with five African American members of a local MS support group. We presented the study materials (ie, text messages, newsletters, and journals) to the focus group members and asked them to provide comments and suggestions for improving the material content. We obtained direct feedback on the content and vocabulary of the text messages, newsletters, and journals. Participants reached a consensus that the materials needed lay vocabulary. In terms of the text messages, each message was inspected for sensitive content and suitability to daily routine. The focus group suggested modifications to wording of some messages and to the feasibility of some activities that were deemed impractical (eg, "Eat while standing at the restaurant"). The suggestions were incorporated into the development of the patient-informed and culturally tailored final study materials.

Data Analysis

Feasibility data for process, resources, and management will be examined via percentage, frequency analysis, and descriptive statistics. Regarding scientific outcomes, data normality will be verified using the Shapiro-Wilk test and descriptive statistics will be computed for all variables per each assessment point. The influence of the intervention on sedentary behavior, physical activity, and physical function—a component of the SF-36—will be evaluated using repeated-measures linear mixed models, respecting the interdependence of measures over time. Point-by-point improvements in quality of life related to reductions in sedentary behavior and increases in physical activity will be assessed using estimates at each measurement time point, adjusted for sociodemographic and clinical covariates. We will further conduct other exploratory analyses as permitted by the data.

Results

Enrollment began in February 2018 and is expected to conclude in April 2019. Intervention delivery will conclude in August 2019. Data analysis with full study results is expected in the fall of 2019.

Discussion

This study will identify the demands and procedures of the proposed intervention strategy (ie, text messaging, newsletters, and behavioral coaching) for changing sedentary behavior in African Americans with MS. The knowledge acquired from this intervention will be valuable in designing future studies for reducing sedentary behavior in a larger number of African Americans with MS. We expect that our intervention will promote reductions and changes in the volume and pattern of sedentary behavior and, by extension, an increase in light physical activity in the study participants. It is important to highlight that, even though the intervention focuses on both sedentary behavior and physical activity, we purposefully place the emphasis on the former, as the literature indicates that intervention success requires a well-defined target [26,31].

The target criterion for considering the intervention a success will be a reduction in sitting time of 60+ minutes per day, as there is evidence that replacement of one hour per day of sitting with physical activity of any intensity can improve quality of life [49]. Studies have further adopted this same volume of sedentary behavior reduction, indicating that it is a reasonable goal for an intervention [24,50]. If the data analyses indicate that this intervention is both feasible (ie, adequate results for process, resources, management, and science) and efficacious (ie, sedentary behavior reductions of 60+ minutes per day), the study results may provide practical and scientific support for applying such an approach in subsequent, large-scale studies that can efficiently target a large number of African Americans with MS who engage in high levels of sedentary behavior (ie, go/no-go decision for a subsequent trial).

The use of technology for reducing sedentary behavior represents an important step for reaching persons with MS remotely [27] and for promoting behavior changes that may improve health [21,51,52]. As of 2018, almost every American (ie, 94%) owned a mobile phone [16], and this makes text messaging an optimal medium for disseminating health information, including information that is directed toward educating people about changing behaviors. The examination of whether text messaging, coupled with the provision of SCT-based newsletters and behavior coaching, may help change sedentary behavior in African Americans with MS is central for scaling up the intervention among a larger number of persons from this population group who might not have access to resources for engaging in regular interventions and/or in-person behavioral intervention. Furthermore, such an approach is advantageous for optimizing the intervention based on the socioeconomic and cultural reality of African Americans with MS. For example, all the intervention materials were designed based on specific characteristics of the target population, as per the recommendations and feedback from a focus group of African American persons with MS (ie, stakeholder engagement). This process contributed toward creating cultural awareness among the researchers and developing strategies appropriate for the target population to maximize the potential to reduce sedentary behavior with the intervention.

There is evidence of the increasing use of technology in theory-based behavior change interventions in people with disabilities, including MS [15]. These strategies have demonstrated promising results in increasing exercise and physical activity levels in such groups of people [15]. The use of technology is ideal for promoting behavior change remotely in large groups of people, given its affordability, reduced personnel burden, and lack of reliance on physical space [20,51,52]. The use of technology can further minimize or even eliminate in-person attendance at specific physical spaces (eg, gyms and recreational clubs) [27-29]. By reducing the burden on both researchers and participants, remote interventions enhance the feasibility and likelihood of successful uptake and adoption of interventions that reduce sedentary behavior. To date, researchers have used computer software [53], DVD-delivered interventions [27,28], and text messaging [24] for promoting reductions in sedentary behavior in the general population and persons with MS. Our study will be a further step in amplifying the possibilities of promoting changes in sedentary behavior remotely in large groups of underserved persons with MS, such as African Americans.

Another advantageous tenet of our intervention involves the particularly high frequency of stimuli for promoting the intended changes in behavior, as text messages can be sent more than once per day over different time periods. This is critical in reminding individuals about continually monitoring sitting behavior and identifying ways to interrupt it. Sedentary behavior is a highly prevalent behavior during a regular day that is difficult to self-monitor. It is usually easier to remember activities that are more intense and less frequent during the day (eg, sports, dance, and recreational activities) [54]. As such, increased frequency of stimuli may be important for effectively reducing sedentary behavior. The results may provide further directions to more efficaciously target this health behavior in future studies.

One strength of this study is the use of both the activPAL and ActiGraph monitors to assess sedentary behavior and physical activity. We will be able to cross-examine the estimates for agreement as well as provide sedentary behavior estimates based on both posture and lack of movement. The inclusion of both devices might further provide data on acceptability and preferences for a device for future trials. Some limitations of this study are noteworthy. The study sample will not be representative of all African Americans with MS since we will only recruit a small number of participants (ie, 30 participants). Nevertheless, the results will provide an indication of whether the proposed strategy has the potential to be delivered and efficacious in a larger sample of African Americans with MS. Another limitation is the short intervention duration; the entire intervention will last for only 3 months. Yet, behavior change is usually considered stable when maintained for longer than 6 months. Funding limitations did not permit collection of such data, but we may be able to obtain data regarding behavior

change maintenance (ie, 6+ months) if funding opportunities arise. The study will not include a control condition. This would be important for ensuring that any behavior changes are due to our intervention itself and not due to other factors; this will be important for future, larger-scale interventions. Another possibility would be to include a comparison group comprised of Caucasians with MS, as this would provide insight regarding differences in responses between African Americans and Caucasians with MS. One final limitation of the study is the lack of a double baseline assessment week. This would be important for assessing reactivity to wearing the activity monitors and, therefore, to better isolate the effects of the intervention on the primary study outcomes. Lastly, we did not examine, in depth, the level of familiarization of participants with portable electronic technology. If more complex electronic strategies are used for prompting behavior change, then researchers need to consider participants' prior history of mobile technology usage for expanding the use of such technology in changing behavior in MS.

In summary, this paper describes the methods and outcomes of interest of a technology-based intervention aimed at reducing sedentary behavior in African Americans with MS. This pilot study will provide data on the feasibility and efficacy of our strategy in promoting sedentary behavior reductions in the aforementioned group and we anticipate the results will be important for researchers investigating strategies for reducing sedentary behavior in MS. After completion of this pilot intervention, we will summarize our study results in manuscripts for publication in peer-reviewed journals that will provide critical information on the feasibility and efficacy of our strategy. These results will inform future studies and, potentially, larger interventions for remotely reducing sedentary behavior in African Americans with MS.

Conflicts of Interest

None declared.

References

1. Owen N, Healy GN, Matthews CE, Dunstan DW. Too much sitting: The population health science of sedentary behavior. Exerc Sport Sci Rev 2010 Jul;38(3):105-113 [FREE Full text] [doi: 10.1097/JES.0b013e3181e373a2] [Medline: 20577058]

2. de Rezende LF, Rodrigues Lopes M, Rey-López JP, Matsudo VK, do Carmo Luiz O. Sedentary behavior and health outcomes: An overview of systematic reviews. PLoS One 2014;9(8):e105620 [FREE Full text] [doi: 10.1371/journal.pone.0105620] [Medline: 25144686]

3. Young DR, Hivert MF, Alhassan S, Camhi SM, Ferguson JF, Katzmarzyk PT, Physical Activity Committee of the Council on Lifestyle and Cardiometabolic Health; Council on Clinical Cardiology; Council on Epidemiology and Prevention; Council on Functional Genomics and Translational Biology; and Stroke Council. Sedentary behavior and cardiovascular morbidity and mortality: A science advisory from the American Heart Association. Circulation 2016 Dec 27;134(13):e262-e279. [doi: 10.1161/CIR.0000000000000440] [Medline: 27528691]

4. Veldhuijzen van Zanten JJ, Pilutti LA, Duda JL, Motl RW. Sedentary behaviour in people with multiple sclerosis: Is it time to stand up against MS? Mult Scler 2016 Dec;22(10):1250-1256. [doi: 10.1177/1352458516644340] [Medline: 27072688]

5. Sasaki JE, Motl RW, Cutter G, Marrie RA, Tyry T, Salter A. National estimates of self-reported sitting time in adults with multiple sclerosis. Mult Scler J Exp Transl Clin 2018;4(1):2055217318754368 [FREE Full text] [doi: 10.1177/2055217318754368] [Medline: 29375889]

6. Hubbard EA, Motl RW, Fernhall B. Sedentary behavior and blood pressure in patients with multiple sclerosis. Int J MS Care 2018;20(1):1-8 [FREE Full text] [doi: 10.7224/1537-2073.2016-021] [Medline: 29507537]

7. Kaufman MD, Johnson SK, Moyer D, Bivens J, Norton HJ. Multiple sclerosis: Severity and progression rate in African Americans compared with whites. Am J Phys Med Rehabil 2003 Aug;82(8):582-590. [doi: 10.1097/01.PHM.0000078199.99484.E2] [Medline: 12872014]

8. Marrie RA, Horwitz R, Cutter G, Tyry T, Campagnolo D, Vollmer T. Comorbidity, socioeconomic status, and multiple sclerosis. Mult Scler 2008 Sep;14(8):1091-1098. [doi: 10.1177/1352458508092263] [Medline: 18728060]

9. Klineova S, Nicholas J, Walker A. Response to disease modifying therapies in African Americans with multiple sclerosis. Ethn Dis 2012;22(2):221-225. [Medline: 22764646]

10. Thomas BE, Charles N, Watson B, Chandrasekaran V, Senthil Kumar R, Dhanalakshmi A, et al. Prevalence of chest symptoms amongst brick kiln migrant workers and care seeking behaviour: A study from South India. J Public Health (Oxf) 2015 Dec;37(4):590-596. [doi: 10.1093/pubmed/fdu104] [Medline: 25538142]

XSL•FO

RenderX

11. Lai B, Cederberg K, Vanderbom KA, Bickel CS, Rimmer JH, Motl RW. Characteristics of adults with neurologic disability recruited for exercise trials: A secondary analysis. Adapt Phys Activ Q 2018 Oct 01;35(4):476-497. [doi: 10.1123/apaq.2017-0109] [Medline: 30382753]

12. Khan O, Williams MJ, Amezcua L, Javed A, Larsen KE, Smrtka JM. Multiple sclerosis in US minority populations: Clinical practice insights. Neurol Clin Pract 2015 Apr;5(2):132-142 [FREE Full text] [doi: 10.1212/CPJ.0000000000000112] [Medline: 26137421]

13. Kershaw KN, Albrecht SS, Carnethon MR. Racial and ethnic residential segregation, the neighborhood socioeconomic environment, and obesity among Blacks and Mexican Americans. Am J Epidemiol 2013 Feb 15;177(4):299-309 [FREE Full text] [doi: 10.1093/aje/kws372] [Medline: 23337312]

14. Lewis CE, Raczynski JM, Heath GW, Levinson R, Hilyer JJ, Cutter GR. Promoting physical activity in low-income African-American communities: The PARR project. Ethn Dis 1993;3(2):106-118. [Medline: 8324487]

15. Lai B, Young HJ, Bickel CS, Motl RW, Rimmer JH. Current trends in exercise intervention research, technology, and behavioral change strategies for people with disabilities: A scoping review. Am J Phys Med Rehabil 2017 Oct;96(10):748-761. [doi: 10.1097/PHM.0000000000000743] [Medline: 28398967]

16. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2018 Feb 05. Mobile fact sheet URL: https://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-04-03] [WebCite Cache ID 77MjS3k4V]

17. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: A meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]

18. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]

19. Park LG, Beatty A, Stafford Z, Whooley MA. Mobile phone interventions for the secondary prevention of cardiovascular disease. Prog Cardiovasc Dis 2016;58(6):639-650 [FREE Full text] [doi: 10.1016/j.pcad.2016.03.002] [Medline: 27001245]

20. Palmer MJ, Barnard S, Perel P, Free C. Mobile phone-based interventions for improving adherence to medication prescribed for the primary prevention of cardiovascular disease in adults. Cochrane Database Syst Rev 2018 Dec 22;6:CD012675. [doi: 10.1002/14651858.CD012675.pub2] [Medline: 29932455]

21. Anglada-Martinez H, Riu-Viladoms G, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona J, Codina-Jane C. Does mHealth increase adherence to medication? Results of a systematic review. Int J Clin Pract 2015 Jan;69(1):9-32. [doi: 10.1111/ijcp.12582] [Medline: 25472682]

22. Ramirez M, Wu S, Beale E. Designing a text messaging intervention to improve physical activity behavior among low-income Latino patients with diabetes: A discrete-choice experiment, Los Angeles, 2014-2015. Prev Chronic Dis 2016 Dec 22;13:E171 [FREE Full text] [doi: 10.5888/pcd13.160035] [Medline: 28005532]

23. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: A systematic review and meta-analysis. Prev Med 2017 Jul;100:194-207 [FREE Full text] [doi: 10.1016/j.ypmed.2017.04.033] [Medline: 28456513]

24. Cotten E, Prapavessis H. Increasing nonsedentary behaviors in university students using text messages: Randomized controlled trial. JMIR Mhealth Uhealth 2016 Aug 19;4(3):e99 [FREE Full text] [doi: 10.2196/mhealth.5411] [Medline: 27543317]

25. Downing KL, Salmon J, Hinkley T, Hnatiuk JA, Hesketh KD. A mobile technology intervention to reduce sedentary behaviour in 2- to 4-year-old children (Mini Movers): Study protocol for a randomised controlled trial. Trials 2017 Dec 03;18(1):97 [FREE Full text] [doi: 10.1186/s13063-017-1841-7] [Medline: 28253904]

26. Baranowski T, Cullen KW, Nicklas T, Thompson D, Baranowski J. Are current health behavioral change models helpful in guiding prevention of weight gain efforts? Obes Res 2003 Oct;11 Suppl:23S-43S [FREE Full text] [doi: 10.1038/oby.2003.222] [Medline: 14569036]

27. Klaren RE, Hubbard EA, Motl RW. Efficacy of a behavioral intervention for reducing sedentary behavior in persons with multiple sclerosis: A pilot examination. Am J Prev Med 2014 Nov;47(5):613-616. [doi: 10.1016/j.amepre.2014.05.036] [Medline: 25070271]

28. McAuley E, Wójcicki TR, Learmonth YC, Roberts SA, Hubbard EA, Kinnett-Hopkins D, et al. Effects of a DVD-delivered exercise intervention on physical function in older adults with multiple sclerosis: A pilot randomized controlled trial. Mult Scler J Exp Transl Clin 2015;1:2055217315584838 [FREE Full text] [doi: 10.1177/2055217315584838] [Medline: 28607692]

29. Adamson BC, Learmonth YC, Kinnett-Hopkins D, Bohri M, Motl RW. Feasibility study design and methods for Project GEMS: Guidelines for Exercise in Multiple Sclerosis. Contemp Clin Trials 2016 Mar;47:32-39. [doi: 10.1016/j.cct.2015.12.002] [Medline: 26655434]

30. Motl RW, Pekmezi D, Wingo BC. Promotion of physical activity and exercise in multiple sclerosis: Importance of behavioral science and theory. Mult Scler J Exp Transl Clin 2018;4(3):2055217318786745 [FREE Full text] [doi: 10.1177/2055217318786745] [Medline: 30090642]

31. Fishbein M, Yzer MC. Using theory to design effective health behavior interventions. Commun Theory 2003;13(2):164-183. [doi: 10.1111/j.1468-2885.2003.tb00287.x]

XSL•FO

RenderX

32. Learmonth YC, Motl RW. Important considerations for feasibility studies in physical activity research involving persons with multiple sclerosis: A scoping systematic review and case study. Pilot Feasibility Stud 2018;4:1 [FREE Full text] [doi: 10.1186/s40814-017-0145-8] [Medline: 28616252]

33. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: The what, why and how. BMC Med Res Methodol 2010 Jan 06;10:1 [FREE Full text] [doi: 10.1186/1471-2288-10-1] [Medline: 20053272]

34. Lancaster GA. Pilot and feasibility studies come of age!. Pilot Feasibility Stud 2015;1(1):1 [FREE Full text] [doi: 10.1186/2055-5784-1-1] [Medline: 29611687]

35. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: Recommendations for good practice. J Eval Clin Pract 2004 May;10(2):307-312. [doi: 10.1111/j.2002.384.doc.x] [Medline: 15189396]

36. Suh Y, Weikert M, Dlugonski D, Sandroff B, Motl RW. Social cognitive correlates of physical activity: Findings from a cross-sectional study of adults with relapsing-remitting multiple sclerosis. J Phys Act Health 2011 Jul;8(5):626-635. [Medline: 21734307]

37. Ellis T, Motl RW. Physical activity behavior change in persons with neurologic disorders: Overview and examples from Parkinson disease and multiple sclerosis. J Neurol Phys Ther 2013 Jun;37(2):85-90. [doi: 10.1097/NPT.0b013e31829157c0] [Medline: 23632452]

38. Learmonth YC, Adamson BC, Kinnett-Hopkins D, Bohri M, Motl RW. Results of a feasibility randomised controlled study of the guidelines for exercise in multiple sclerosis project. Contemp Clin Trials 2017 Dec;54:84-97. [doi: 10.1016/j.cct.2016.11.012] [Medline: 27916699]

39. Sebastião E, McAuley E, Shigematsu R, Adamson BC, Bollaert RE, Motl RW. Home-based, square-stepping exercise program among older adults with multiple sclerosis: Results of a feasibility randomized controlled study. Contemp Clin Trials 2018 Oct;73:136-144. [doi: 10.1016/j.cct.2018.09.008] [Medline: 30243811]

40. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985 Sep;10(3):141-146. [Medline: 4053261]

41. Coote S, O'Dwyer C. Comparative validity of accelerometer-based measures of physical activity for people with multiple sclerosis. Arch Phys Med Rehabil 2012 Nov;93(11):2022-2028. [doi: 10.1016/j.apmr.2012.05.010] [Medline: 22634293]

42. Sandroff BM, Motl RW, Suh Y. Accelerometer output and its association with energy expenditure in persons with multiple sclerosis. J Rehabil Res Dev 2012;49(3):467-475 [FREE Full text] [Medline: 22773205]

43. Sikes EM, Richardson EV, Cederberg KJ, Sasaki JE, Sandroff BM, Motl RW. Use of the Godin leisure-time exercise questionnaire in multiple sclerosis research: A comprehensive narrative review. Disabil Rehabil 2018 Jan 17:1-25. [doi: 10.1080/09638288.2018.1424956] [Medline: 29343122]

44. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]

45. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): A systematic review. Int J Behav Nutr Phys Act 2011 Oct 21;8:115 [FREE Full text] [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]

46. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]

47. Bandura A. Social cognitive theory: An agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]

48. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]

49. Balboa-Castillo T, León-Muñoz LM, Graciani A, Rodríguez-Artalejo F, Guallar-Castillón P. Longitudinal association of physical activity and sedentary behavior during leisure time with health-related quality of life in community-dwelling older adults. Health Qual Life Outcomes 2011 Jun 27;9:47 [FREE Full text] [doi: 10.1186/1477-7525-9-47] [Medline: 21708011]

50. Lewis LK, Rowlands AV, Gardiner PA, Standage M, English C, Olds T. Small Steps: Preliminary effectiveness and feasibility of an incremental goal-setting intervention to reduce sitting time in older adults. Maturitas 2016 Mar;85:64-70. [doi: 10.1016/j.maturitas.2015.12.014] [Medline: 26857881]

51. Bond DS, Thomas JG, Raynor HA, Moon J, Sieling J, Trautvetter J, et al. B-MOBILE--A smartphone-based intervention to reduce sedentary time in overweight/obese individuals: A within-subjects experimental trial. PLoS One 2014;9(6):e100821 [FREE Full text] [doi: 10.1371/journal.pone.0100821] [Medline: 24964010]

52. McCarroll R, Eyles H, Ni Mhurchu C. Effectiveness of mobile health (mHealth) interventions for promoting healthy eating in adults: A systematic review. Prev Med 2017 Dec;105:156-168. [doi: 10.1016/j.ypmed.2017.08.022] [Medline: 28882743]

53. Evans RE, Fawole HO, Sheriff SA, Dall PM, Grant PM, Ryan CG. Point-of-choice prompts to reduce sitting time at work: A randomized trial. Am J Prev Med 2012 Sep;43(3):293-297. [doi: 10.1016/j.amepre.2012.05.010] [Medline: 22898122]

54. Baranowski T. Validity and reliability of self-report measures of physical activity: An information-processing perspective. Res Q Exerc Sport 1988;59(4):314-327. [doi: 10.1080/02701367.1988.10609379]

Abbreviations

ICD: informed consent document
MET: metabolic equivalents of task
MS: multiple sclerosis
PDDS: Patient-Determined Disease Steps
SCT: Social Cognitive Theory
SF-36: 36-item Short Form Health Survey
T1: baseline
T2: midpoint
T3: postintervention
UAB: University of Alabama at Birmingham

XSL•FO
RenderX

Protocol

Acceptability and Feasibility of Self-Collecting Biological Specimens for HIV, Sexually Transmitted Infection, and Adherence Testing Among High-Risk Populations (Project Caboodle!): Protocol for an Exploratory Mixed-Methods Study

Akshay Sharma[1,2], MBBS, MPH, PhD; Rob Stephenson[1,3], MSc, PhD; Gregory Sallabank[1], MA; Leland Merrill[1], BA; Stephen Sullivan[1], MPH; Monica Gandhi[4], MD, MPH

[1]Center for Sexuality and Health Disparities, University of Michigan School of Nursing, Ann Arbor, MI, United States

[2]Department of Health Behavior and Biological Sciences, University of Michigan School of Nursing, Ann Arbor, MI, United States

[3]Department of Systems, Population and Leadership, University of Michigan School of Nursing, Ann Arbor, MI, United States

[4]Division of HIV, Infectious Disease, and Global Medicine, Department of Medicine, University of California, San Francisco, San Francisco, CA, United States

Corresponding Author:
Akshay Sharma, MBBS, MPH, PhD
Center for Sexuality and Health Disparities
University of Michigan School of Nursing
400 North Ingalls Street
Ann Arbor, MI, 48109
United States
Phone: 1 734 647 0151
Fax: 1 734 763 0681
Email: akshaydr@umich.edu

Abstract

Background: Men who have sex with men (MSM) in the United States experience a disproportionate burden of HIV and bacterial sexually transmitted infections (STIs), such as gonorrhea and chlamydia. Screening levels among MSM remain inadequate owing to barriers to testing such as stigma, privacy and confidentiality concerns, transportation issues, insufficient clinic time, and limited access to health care. Self-collection of specimens at home and their return by mail for HIV and bacterial STI testing, as well as pre-exposure prophylaxis (PrEP) adherence monitoring, could be a resource-efficient option that might mitigate some of these barriers.

Objective: Project Caboodle! is a mixed-methods study that explores the acceptability and feasibility of self-collecting and returning a bundle of 5 different specimens for HIV and bacterial STI testing, as well as PrEP adherence monitoring, among sexually active HIV-negative or unknown status MSM in the United States aged 18 to 34 years.

Methods: Participants will be recruited using age, race, and ethnicity varied advertising on social networking websites and mobile gay dating apps. In Phase 1, we will send 100 participants a box containing materials for self-collecting and potentially returning a finger-stick blood sample (for HIV testing), pharyngeal swab, rectal swab, and urine specimen (for gonorrhea and chlamydia testing), and hair sample (to assess adequacy for potential PrEP adherence monitoring). Specimen return will not be incentivized, and participants can choose to mail back all, some, or none of the specimens. Test results will be delivered back to participants by trained counselors over the phone. In Phase 2, we will conduct individual in-depth interviews using a video-based teleconferencing software (VSee) with 32 participants from Phase 1 (half who returned all specimens and half who returned some or no specimens) to examine attitudes toward and barriers to completing various study activities.

Results: Project Caboodle! was funded in May 2018, and participant recruitment began in March 2019. The processes of designing a study logo, creating advertisements, programming Web-based surveys, and finalizing step-by-step written instructions accompanied by color images for specimen self-collection have been completed. The boxes containing 5 self-collection kits affixed with unique identification stickers are being assembled, and shipping procedures (for mailing out boxes to participants and for specimen return by participants using prepaid shipping envelopes) and payment procedures for completing the surveys and in-depth interviews are being finalized.

XSL•FO

RenderX

Conclusions: Self-collection of biological specimens at home and their return by mail for HIV and bacterial STI testing, as well as PrEP adherence monitoring, might offer a practical and convenient solution to improve comprehensive prevention efforts for high-risk MSM. The potentially reduced time, expense, and travel associated with this approach could facilitate a wider implementation of screening algorithms and remote monitoring strategies.

Introduction

Background

Despite representing a small fraction of the US population [1,2], men who have sex with men (MSM) account for 58.95% (648,500/1.1 million) of all people living with HIV [3] and 66.79% (26,570/39,782) of all new HIV diagnoses annually [4]. According to the Centers for Disease Control and Prevention (CDC), HIV diagnoses among MSM increased from 25,155 in 2005 to 26,612 in 2014 [5]. The rate of incident diagnoses in this risk group is more than 44 times that of other men [3], and the rate of prevalent diagnoses is more than 57 times greater [6]. MSM in the United States also face a high burden of bacterial sexually transmitted infections (STIs), such as gonorrhea and chlamydia. Prevalence proportions for pharyngeal, rectal, and urethral gonorrhea among MSM are estimated to be 7.90% (1,144/14,484), 10.24% (1,136/11,092), and 11.14% (2,056/18,460), respectively, and those for pharyngeal, rectal, and urethral chlamydia are estimated to be 2.86% (199/6,961), 14.14% (1,427/10,091), and 8.35% (1,495/17,898), respectively [7]. These data are concerning because bacterial STIs acting through biological mechanisms, in addition to behaviors associated with their acquisition and transmission, are well-established risk factors for HIV [8-12].

The CDC currently recommends at least annual screening for HIV, gonorrhea, and chlamydia for all sexually active MSM [13,14]. More frequent screening for HIV (eg, at 3- to 6-month intervals) could be offered to those at increased risk based on an assessment of their individual risk factors, local HIV epidemiology, and local testing policies [13,15]. Testing is the first step in offering pre-exposure prophylaxis (PrEP) for HIV prevention to those who test negative or initiating treatment for HIV and other STIs among those who test positive. However, screening levels among MSM remain inadequate [16,17] owing to barriers to testing such as stigma, privacy and confidentiality concerns, transportation issues, insufficient clinic time, and limited access to health care [18-30]. In 2014, 71.14% (5,864/8,243) of MSM participating in the National HIV Behavioral Surveillance reported testing for HIV in the past year [17], and 46.98% (8,984/19,124) reported testing for gonorrhea or chlamydia in the past year [31]. Novel strategies are needed to increase current rates of HIV and other STI screening among MSM. Self-collection of specimens at home and their return by mail for laboratory testing could be a resource-efficient option [32] and has the potential to reduce both personal and logistical barriers to regular testing [33].

Oral PrEP with tenofovir (TFV) disoproxil fumarate and emtricitabine (FTC) significantly reduces the risk of HIV among MSM but relies on adequate adherence for effectiveness [34-38]. Measuring biomarkers in biological specimens for assessing PrEP adherence played a key role in the interpretation of landmark placebo-controlled randomized clinical trials [39-41]. The delivery of proven biomedical interventions needs to be accompanied by complementary strategies for measuring and increasing adherence to optimize their effectiveness [42-46]. Although recent demonstration projects have reported high levels of PrEP use and adherence in the context of known efficacy [38,47], adherence biomarkers need to be incorporated into PrEP implementation and roll-out programs to assess its effectiveness in real-world settings and to tailor adherence interventions [48,49]. Hair samples can be used for PrEP adherence measurement [48-58], both in daily and in event-driven PrEP users. As distance along the hair shaft serves as a marker of time, the segmental analysis of hair samples allows for an objective assessment of event-driven PrEP adherence at various time points over previous months [59,60]. Self-collection at home and return by mail of this nonbiohazardous, easy-to-ship specimen that is stable at room temperature might facilitate remote PrEP adherence monitoring and thereby allow for the appropriate identification of MSM facing adherence difficulties for intervention.

Numerous studies on biological specimen self-collection have been conducted in clinical settings, wherein patients immediately return their samples to clinic staff for subsequent testing in a laboratory. For example, self-collected nasal swabs have been used to diagnose respiratory tract infections [61-63], self-collected vaginal swabs have been used for bacterial STIs and cervical cancer screening among women [64-71], and self-collected genital specimens have been used for human papillomavirus screening among men [72-74]. This approach prioritizes patient comfort and facilitates clinic flow, but it cannot reach individuals who do not come to the clinic for testing. Recently, there has been an increase in research focusing on home specimen self-collection for HIV and other STI testing among diverse populations, including sexual minorities [75-87]. Several of these studies involved either cross-sectional surveys or focus group discussions to assess the acceptability of this approach but few studies have also examined feasibility, that is, whether MSM would actually return individual specimens and whether these specimens would be adequate for laboratory testing. To our knowledge, no study has evaluated the acceptability and feasibility of self-collecting at home and returning by mail a bundle of different types of biological

specimens that can be used for HIV and bacterial STI screening (eg, a finger-stick blood sample, pharyngeal swab, rectal swab, and urine specimen) and for potential PrEP adherence monitoring (eg, a hair sample) among young sexually active MSM.

Objectives

Combining the self-collection of biological specimens from different anatomical sites for HIV and bacterial STI testing along with the self-collection of a hair sample, a specimen that has been shown to have utility in PrEP adherence measurement [48-58], could hold promise as a remote monitoring strategy for individuals at risk. Gathering and evaluating data on the specimens that MSM are willing to self-collect at home, the ones they actually mail back, and the adequacy of returned specimens for laboratory testing are critical to developing interventions to help increase HIV and other STI screening rates, as well as adherence to PrEP in this high-risk group. The purpose of this paper is to describe the protocol for an innovative mixed-methods study seeking to evaluate the acceptability and feasibility of biological specimen self-collection and return from a diverse sample of young sexually active HIV-negative or unknown status MSM in the United States. The procedures described below have been reviewed and approved by the Institutional Review Board at the University of Michigan in Ann Arbor (HUM00153673) and have been deemed to pose no more than minimal risk to study participants.

Methods

Study Overview

Project Caboodle! is an exploratory 2-year study to obtain both quantitative and qualitative data regarding the acceptability and feasibility of biological specimen self-collection and return from sexually active HIV-negative or unknown status MSM aged 18 to 34 years residing in the United States or dependent areas. In Phase 1, we will send 100 participants a box containing instructions and materials for self-collecting and potentially returning a finger-stick blood sample (for HIV testing), pharyngeal swab, rectal swab, and urine specimen (for gonorrhea and chlamydia testing), and hair sample (to assess adequacy for potential PrEP adherence monitoring in the future). Specimen return is not incentivized, and participants can choose to mail back all, some, or none of these 5 specimens. In Phase 2, we will conduct individual in-depth interviews using a video-based teleconferencing software (VSee) with 32 participants from Phase 1 to examine attitudes toward and barriers to self-collecting and returning each type of specimen.

Guided by the theoretical constructs of the Information-Motivation-Behavioral skills (IMB) model [88,89], the specific aims of our study are as follows:

1. *Aim 1:* Explore the acceptability and feasibility of self-collecting at home and returning by mail all, some, or none of the following biological specimens: (1) finger-stick blood sample, (2) pharyngeal swab, (3) rectal swab, (4) urine specimen, and (5) hair sample, among 100 sexually active HIV-negative or unknown status US MSM (aged 18 to 34 years, 50 black and 50 white including those of

Hispanic ethnicity) recruited through social media platforms.

2. *Aim 2:* Collect qualitative data via individual in-depth interviews conducted using VSee from 2 subsamples of 16 MSM each, including participants who (1) returned all 5 specimens and (2) returned some or no specimens to examine attitudes toward and barriers to completing various study activities.

Theoretical Approach

Our study is grounded in the IMB model of HIV prevention [88] that describes pivotal constructs pertaining to health self-management [89]. Theoretical constructs of the IMB model will be applied to explore the acceptability and feasibility of self-collecting and potentially returning the 5 different types of specimens by MSM. *Information* will be assessed through a baseline survey in Phase 1, measuring pre-existing awareness of commercially available self-collection kits using questions adapted from our past research on rapid home HIV testing using oral fluid and finger-stick blood samples [90,91]. The survey will also assess participants' knowledge of transmission and prevention of HIV and other STIs using the Sexually Transmitted Disease Knowledge Questionnaire (STD-KQ) [92]. *Motivation* will be examined through questions in the baseline survey on participants' theoretical willingness to self-collect and return each type of specimen, expanding upon our previous study focusing exclusively on dried blood spot specimens [86] and by calculating the proportions of each type of specimen actually returned for laboratory testing. Depending upon their inclination, participants can choose to mail back all, some, or none of the 5 specimens. *Behavioral skills* will be measured by assessing the adequacy of returned specimens for current and future laboratory testing, thereby evaluating the ability of participants to self-collect and return useful specimens without supervision. Theoretical constructs of the IMB model will also be explored through qualitative in-depth interviews in Phase 2, which will focus on familiarity with the concept of biological specimen self-collection (information) and attitudes toward self-collecting and returning each type of specimen (motivation), as well as personal and logistical barriers encountered while performing study activities (behavioral skills).

Participant Recruitment

Project Caboodle! will recruit and enroll 100 young (aged 18 to 34 years), sexually active HIV-negative or unknown status MSM residing in the United States. Participants will be recruited online using age-appropriate, racially and ethnically diverse advertising on social networking websites (eg, Facebook and Instagram) and mobile gay dating apps (eg, Grindr and Scruff). We will aim to ensure that our sample includes equal numbers of black and white MSM including those of Hispanic ethnicity. Our advertisements will include images of men hugging, kissing, or holding hands, the study logo, and call-to-action text. These will appear as posts on social networking websites and will be targeted to profiles reflecting gay *interests*, that is, topics users have accessed (eg, same-sex marriage), and pages they have *liked* (eg, pride events).

Individuals who click on our advertisements will be directed to our study's landing page (programmed in Qualtrics, a secure

Web-based platform approved by the University of Michigan) that will provide a brief overview of the study protocol, in addition to basic information on the burden of HIV and bacterial STIs among MSM in the United States. Those who are not interested can exit by closing the landing page in their browser. Interested individuals can click a button to continue, which will direct them to a comprehensive informed consent document. This document will provide information regarding our study's purpose, contents of the eligibility screener, the baseline survey, and the posttest survey, study procedures in Phases 1 and 2, potential risks of participation, and the right to refuse participation or withdraw at any time. Individuals will be asked to provide consent to (1) be screened for eligibility, (2) be asked for their contact information, (3) participate in the biological specimen self-collection activities (ie, Phase 1) if they meet the eligibility criteria, and (4) be potentially asked to participate in an in-depth interview (ie, Phase 2).

Individuals who consent will be asked to complete an 8-question eligibility screener, programmed in Qualtrics. The eligibility criteria include the following: (1) assigned male sex at birth, (2) currently identify as male, (3) aged 18 to 34 years (owing to the high burden of HIV and bacterial STIs among MSM in this age group [4,5]), (4) currently reside in the United States or dependent areas, (5) are a legal adult in their state of residence, (6) report HIV-negative or unknown status, (7) had ≥2 male sex partners in the past 3 months, and (8) are willing to receive a box containing instructions and materials for specimen self-collection at their preferred mailing address. Eligible MSM will then be asked to register by providing their full name, email address, and mobile phone number. They will also be asked to indicate whether they prefer being contacted by email, phone call, or text to remind them to complete Phase 1 surveys (baseline and posttest) and to be potentially invited to participate in Phase 2 in-depth interviews. Finally, they will be asked to provide a mailing address where they would be willing to receive a box containing instructions and materials

for self-collecting and potentially returning different types of biological specimens.

Potential participants' identities will be verified using 3 steps. First, the Internet Protocol (IP) address of the device that was used to complete the eligibility screener will be recorded within Qualtrics and checked by the study staff to verify that (1) the IP address location is within the United States and (ii) there are no duplicate entries from the same IP address. Second, each potential participant will be asked to reply to an email sent to their preferred email address to confirm its accuracy and functionality. Third, the email address, mobile phone number, and mailing address provided will be validated using Spokeo, a Web-based search platform that aggregates publicly available social media and archival data. Spokeo will be used to further authenticate that the email address and mobile phone number provided by a potential participant correctly links to their provided full name and key demographic eligibility criteria (eg, cisgender male identity). To continue as a participant in the study, one needs to (1) have their IP address located in the United States, (2) reply to the confirmation message sent to their provided email address, (3) have at least 1 aspect of their provided contact information (email address, mobile phone number, and mailing address) be linked to their provided full name or key demographic eligibility criteria. Individuals whose identities cannot be verified will be sent an email informing them that they cannot continue in the study and thanking them for their interest. Such email notifications will take place on an ongoing basis immediately after the verification process is complete.

Individuals who do not consent, do not meet the eligibility criteria, or do not provide valid contact information will be excluded from participating any further and will be directed to the CDC HIV Risk Reduction Tool website containing links to sexual health information, PrEP, and other HIV and STI prevention resources [93]. Those who consent, who meet the eligibility criteria, and whose identities have been verified will be registered as study participants (see Figure 1).

Figure 1. Participant recruitment process for Project Caboodle!. CDC: Centers for Disease Control and Prevention.
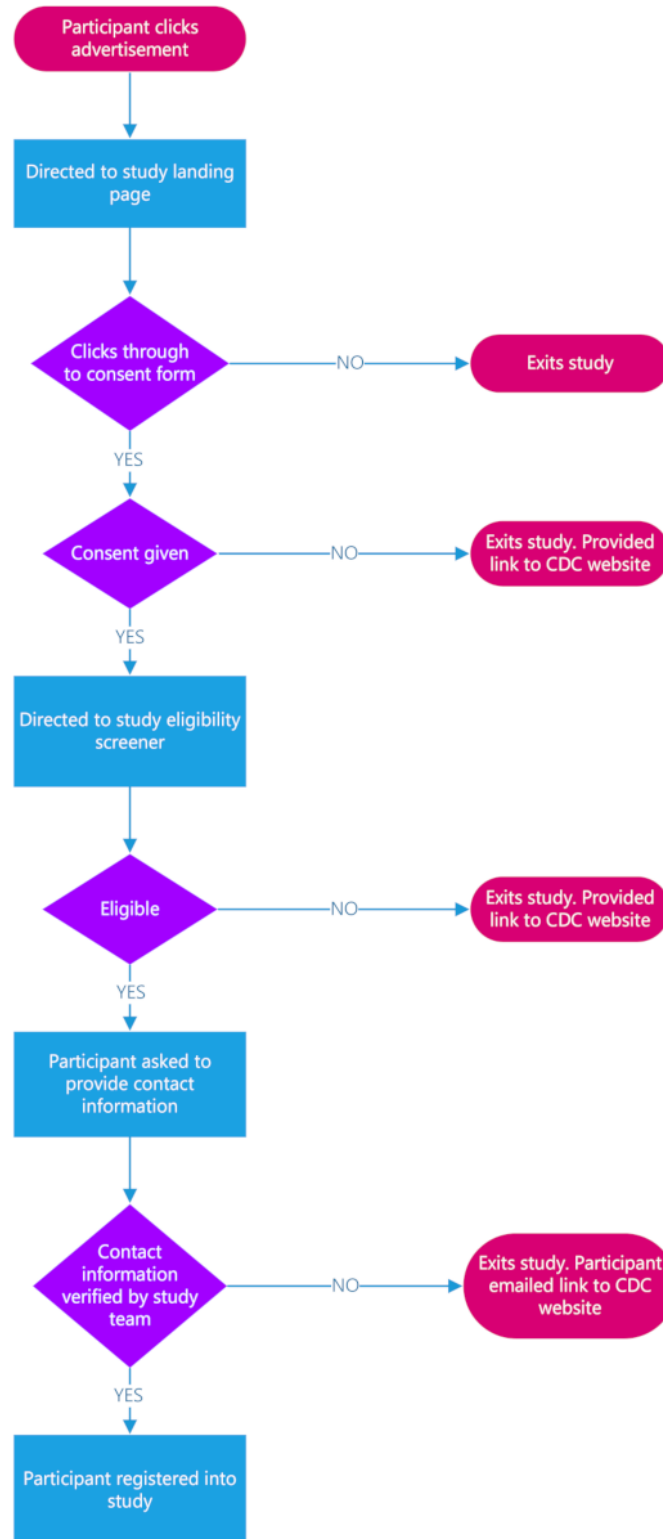
Phase 1 Procedures

Registered participants will be sent an email containing a link to the baseline survey, programmed in Qualtrics. The survey will include questions pertaining to the following domains—*demographic characteristics:* data on age, race and ethnicity, educational level, sexual orientation, employment status, history of incarceration, history of homelessness, health insurance coverage, and access to health care will be collected [94]; *HIV*

testing history: participants will be asked about whether they have ever been tested for HIV, the time frame of their most recent test, the location of their most recent test, and their annual frequency of testing or their reasons for never testing [81,86,94]; *home-based HIV testing:* data will be collected on whether participants have ever used a commercially available self-collection kit (eg, Home Access HIV-1 Test System and OraQuick In Home HIV Test) and the type of test kit they have used or their reasons for never using a self-collection kit; *PrEP*

knowledge, use, and intentions: questions will be adapted from recent studies of PrEP attitudes among MSM [95-97] to collect data on whether participants have heard about PrEP, are currently using PrEP, their previous use of PrEP, reasons for never using or discontinuing PrEP, access to PrEP, and intentions for future use of PrEP; *bacterial STI testing history:* participants will be asked about whether they have ever been tested for bacterial STIs, the time frame of their most recent tests, the location of their most recent tests, different anatomical sites at which they have been tested, whether they have ever been diagnosed with any STIs in the past, and if so, whether they received treatment or their reasons for never testing [98,99]; *sexual behavior:* data will be collected on relationship status, the number of sex partners with whom participants have had condom-less anal or oral sex in the past 3 months, and whether any of those partners were known to be positive for HIV or other STIs [100,101]; *knowledge of transmission and prevention of HIV and other STIs:* the 27-item STD-KQ will be used to assess participants' knowledge about different STIs [92]; *perceived risk of contracting HIV and other STIs:* participants will be asked to indicate how concerned they are about contracting HIV and other STIs on a 10-point scale ranging from not concerned at all to very concerned [102]; *experiences with medical care:* the Group-Based Medical Mistrust Scale will be used to assess participants' previous experiences of racial discrimination and feelings of discomfort and suspicion they might have toward health care personnel and medical treatments [103]; *theoretical willingness and perceived ability to self-collect and return biological specimens:* Likert-item questions will be used to assess participants' intentions to self-collect and return each of the 5 different types of biological specimens similar to previous studies [81,90], as well as their confidence to complete these activities using simple step-by-step written instructions accompanied by color images; *substance use and psychological distress:* participants will be asked about their alcohol, tobacco, and other drug use in the past 6 months and to self-report depressive symptomology on the revised Center for Epidemiologic Studies Depression Scale [104]. The baseline survey will take approximately 30 min to complete, and participants will receive an incentive of US $40 in the form of an Amazon gift card.

After participants have been sent the original email containing a link to the baseline survey, they will receive up to 3 reminders at weekly intervals using their preferred communication method indicated during registration (email, phone call, or text). The final reminder will advise participants that noncompletion of the survey within the next 7 days will suggest that they are no longer interested in continuing, unless they contact the study staff and ask for an extension. They will be also be advised that they will no longer qualify to receive a specimen self-collection box in the mail after that time has lapsed. Participants who do not complete the baseline survey within 7 days of receiving the final reminder (ie, by the end of 4 weeks after the original email has been sent) will be thanked for their interest in the study and emailed information about the CDC HIV Risk Reduction Tool website.

Participants who complete the baseline survey will be shipped a box containing instructions and materials for self-collecting and potentially returning 5 different types of biological specimens (ie, a finger-stick blood sample, pharyngeal swab, rectal swab, urine specimen, and hair sample). Each box and its components will be affixed with a unique identification sticker to enable specimen identification upon return. Boxes will be shipped from the University of Michigan School of Nursing (UMSN) in Ann Arbor using United Parcel Service (UPS) in plain unmarked packaging, and their delivery will be tracked by the study staff on an ongoing basis. Contents of the box will include the following:

General Instructions

Participants will be provided with a brief overview of the study procedures and a description of the 5 specimen self-collection kits contained in the box. They will be informed that they can choose to self-collect and return all, some, or none of the specimens depending on their comfort levels. They will also be informed that the box includes 2 prepaid shipping envelopes affixed with FedEx labels—#1 for biological specimens to be returned to the Emory University Clinical Virology Research Laboratory in Atlanta for HIV, gonorrhea, and chlamydia testing (ie, a finger-stick blood sample, pharyngeal swab, rectal swab, and urine specimen) and #2 for a hair sample to be returned to the Hair Analytical Laboratory (HAL) at the University of California, San Francisco (UCSF).

Specific Instructions

Each specimen self-collection kit will contain simple step-by-step written instructions accompanied by color images provided by the laboratories.

1. Finger-stick blood sample self-collection kit (for HIV testing): Participants will be instructed to wash their hands with soap and warm water, clean their middle or ring finger from their nondominant hand using the alcohol wipe, stimulate blood flow by shaking their hand below the waist for 15 seconds, prick their chosen finger using the safety lancet, wipe away the first drop of blood using the sterile gauze, collect blood using the capillary into the transport tube marked with a fill line, secure the transport tube's cap, apply the bandage to their finger, gently mix the blood with the anticoagulant in the transport tube by flipping it upside down, and finally place the tube in the provided biohazard bag.

2. Pharyngeal swab self-collection kit (for gonorrhea and chlamydia testing): Participants will be instructed to open their mouth wide, wipe the swab around their tonsils on both sides of their throat, place it in the transport tube, screw the transport tube's cap back on securely, and finally place the tube in the provided biohazard bag.

3. Rectal swab self-collection kit (for gonorrhea and chlamydia testing): Participants will be instructed to apply 1 drop of lubricant to the tip of the swab, insert it approximately 1.5 inches into their rectum, gently rotate it for 5 to 10 seconds in a circular motion, withdraw it carefully and place it in the transport tube, screw the transport tube's cap back on securely, and finally place the tube in the provided biohazard bag.

4. Urine specimen self-collection kit (for gonorrhea and chlamydia testing): Participants will be instructed to collect

the first part of their urine stream into a sample collection cup marked with a fill line, use a pipette to transfer urine from the cup into the transport tube, repeating the process until the tube is filled between the minimum and maximum fill lines, screw the transport tube's cap back on securely, and finally place the tube in the provided biohazard bag.

Once all specimens (intended to be returned by a participant) have been placed in the biohazard bag, participants will be informed that they should seal it and place it in the prepaid shipping envelope #1 to be returned to the Emory University Clinical Virology Research Laboratory.

5. Hair sample self-collection kit (to assess adequacy for potential PrEP adherence monitoring in the future): Participants will be instructed to clean scissor blades with the alcohol wipe, cut a segment of hair (about 20 to 30 fibers) from the side of their head as close to their scalp as possible, tape the hair sample to the piece of aluminum foil with an adhesive label placed on the hair end furthest from the scalp, fold the foil shut, and finally place it in the clear plastic bag provided.

Once the hair sample (intended to be returned by a participant) has been placed in the clear plastic bag, participants will be informed they should seal it and place it in the prepaid shipping envelope #2 to be returned to the HAL at UCSF.

Specimen return is completely voluntary, and no incentives will be provided to participants for completing this step. Returned finger-stick blood samples, pharyngeal swabs, rectal swabs, and urine specimens will be tested for HIV, gonorrhea, and chlamydia at the Emory University Clinical Virology Research Laboratory and returned hair samples will be visually inspected for adequacy for PrEP drug level testing at the HAL at UCSF. HAL staff have analyzed tens of thousands of hair samples for TFV and FTC concentrations and can readily determine by visual inspection if the self-collected specimens are adequate for potential PrEP adherence monitoring. No identifiable information will be provided to laboratory personnel at Emory University or UCSF, and the specimens will be connected to the results solely on the basis of the box ID. Results will be returned to the study staff at UMSN through a password-protected file shared over Box, a secure cloud storage and collaboration platform approved by the University of Michigan.

HIV, gonorrhea, and chlamydia test results will be delivered back to participants by trained counselors over the phone. Each of the counselors will have experience in the provision of HIV Counseling, Testing, and Referral. They will therefore have experience in answering participants' questions about the HIV and STI testing processes, addressing concerns around sexual risk behaviors, and initiating linkage to care. For anyone testing positive for HIV, gonorrhea, or chlamydia, the study staff will compile a list of local treatment providers in their area using resources such as United Way 2-1-1 [105] and Emory University's AIDSVu testing and treatment locator [106]. Participants with positive test results will be counseled about the importance of accessing treatment, notifying their sexual partners, and sexual risk reduction measures. Within 24 hours

of delivering the positive test results, the study staff will send them an email including a list of local treatment providers. Participants will be contacted by phone 2 more times: (1) 2 weeks after the initial delivery of positive test results to confirm whether or not an appointment was made and (2) 4 weeks after the initial delivery of positive test results to assess engagement in care and provide any additional resources requested. Participants testing negative for HIV, gonorrhea, and chlamydia will be counseled about the importance of regular screening and engaging in safe sexual behaviors. For anyone testing negative for HIV and not on PrEP, the study staff will provide more information about this prevention option. During the phone call to return HIV and STI test results, participants will also be informed about whether the self-collected hair samples they returned were of adequate quality for potential PrEP adherence monitoring. If a participant's hair sample quality analysis results have not been received before their HIV, gonorrhea, and chlamydia results, the study staff will follow-up with participants as soon as they are available (ie, the return of HIV and STI test results will be prioritized).

Participants will be given 6 weeks from box delivery to collect their biological specimens and return them by mail for laboratory processing. Those who have returned some or all of their specimens will be emailed a link to a short posttest survey, within 24 hours of the results delivery phone call. The survey, programmed in Qualtrics, will assess any change in their previous willingness and perceived ability to self-collect and return different types of specimens since using the specimen self-collection kits. Participants who have not returned any specimens within 6 weeks of box delivery will also be emailed a link to the posttest survey at that time point to assess any change in their previous willingness and perceived ability to self-collect and return specimens after seeing or attempting to use the 5 different types of specimen self-collection kits. The survey will also include questions to elicit reasons for not returning each type of specimen using lists of pre-specified options, as well as open-ended response fields. The posttest survey will take approximately 10 min to complete, and participants will receive an incentive of US $10 in the form of an Amazon gift card.

Similar to the baseline survey, after participants have been sent the original email containing a link to the posttest survey, they will receive up to 3 reminders at weekly intervals using their preferred communication method indicated during registration (email, phone call, or text). The final reminder will advise participants that noncompletion of the survey within the next 7 days will suggest that they are no longer interested in continuing, unless they contact study staff and ask for an extension. Participants who do not complete the posttest survey within 7 days of receiving the final reminder (ie, by the end of 4 weeks after the original email has been sent) will no longer be contacted.

Phase 1 Outcomes

Specific outcomes to be measured during this phase include— *Information:* (1) Pre-existing awareness of commercially available self-collection kits for home-based HIV testing, (2) Knowledge of transmission and prevention of HIV and other

STIs, and (3) Variations in knowledge levels across categories of selected characteristics (eg, age, race and ethnicity, educational level, sexual orientation, HIV and bacterial STI testing history, relationship status, sexual behaviors, experiences with medical care, substance use, and psychological distress); *Motivation:* (1) Theoretical willingness to self-collect and return each of the 5 different types of biological specimens, (2) Actual return of each type of specimen, defined as the receipt of prepaid shipping envelopes back at the laboratories within 6 weeks of box delivery to participants, and (3) Reasons for not returning certain types of specimens; and *Behavioral skills:* (1) Perceived ability to self-collect and return each of the 5 different types of biological specimens and (2) Adequacy of specimens to conduct actual laboratory testing for HIV, gonorrhea, and chlamydia (determined by testing returned finger-stick blood samples, pharyngeal swabs, rectal swabs, and urine specimens) and potential laboratory testing for PrEP drug levels (determined by visually inspecting returned hair samples).

Phase 1 Analytic Plan

Descriptive statistics (means, medians, and interquartile ranges for continuous variables, and counts and proportions for categorical variables) will be used to characterize the demographic and behavioral characteristics of the sample (including HIV and bacterial STI testing history) using software for quantitative data analysis (SAS). Analyses paralleling the IMB model include— *Information:* (1) Proportions of participants who are aware of commercially available self-collection kits for home-based HIV testing will be estimated, (2) Scores for the knowledge of transmission and prevention of HIV and other STIs will be formulated for each participant, and (3) Variations in knowledge levels across categories of selected characteristics will be assessed using chi-square tests for homogeneity; *Motivation:* (1) Frequency distributions of participants' theoretical willingness to self-collect and return each of the 5 different types of biological specimens will be plotted, (2) Proportions of each type of specimen actually returned will be calculated and compared with theoretical willingness using Fisher exact tests across categories of selected characteristics (eg, age, race and ethnicity, educational level, sexual orientation, HIV and bacterial STI testing history, relationship status, sexual behaviors, experiences with medical care, substance use, and psychological distress), and (3) Reasons for not returning certain types of specimens will be tabulated, including manually reviewing and reassigning open-ended responses to appropriate pre-specified options; and *Behavioral skills:* (1) Frequency distributions of participants' perceived ability to self-collect and return each of the 5 different types of biological specimens will be plotted and (2) Proportions of each type of returned specimen that are deemed adequate for actual laboratory testing (for HIV, gonorrhea, and chlamydia) and potential laboratory testing (for PrEP drug levels) will be calculated. Frequencies and proportions of positive and negative HIV, gonorrhea, and chlamydia test results and linkage to care statistics will also be aggregated.

Phase 2 Procedures

The study staff will review participant records to identify 2 subsamples of individuals including those who (1) returned all 5 specimens and (2) returned some or no specimens. Up to 16 participants from each group (32 total) will be invited to participate in individual in-depth interviews to be conducted using VSee, a video-based teleconferencing software. VSee uses Federal Information Processing Standard Publication 140-2 certified encryption and does not stream data through a third party, promoting compliance with the Health Insurance Portability and Accountability Act. Within each subsample of 16 participants, quota sampling will be used to ensure equal numbers of black and white MSM including those of Hispanic ethnicity.

Invitations will be extended via email, phone calls, or texts depending upon participants' preferred communication methods indicated during registration. Participants will receive up to 3 reminders at weekly intervals requesting participation in an in-depth interview. The final reminder will advise participants that not contacting study staff within the next 7 days will suggest that they are no longer interested in continuing, unless they ask for an extension. Participants who do not respond within 7 days of receiving the final reminder (ie, by the end of 4 weeks after originally being requested to participate in an in-depth interview) will be thanked for their interest in the study and emailed information about the CDC HIV Risk Reduction Tool website.

Participants who agree to an in-depth interview will be given a range of dates and times to choose from and be emailed instructions to download the VSee app on their computers (to be used with a webcam) or their mobile phones (to be used with their phone's front-facing camera). One-on-one interviews will be conducted by the study staff using a desktop computer equipped with a webcam, and each session will be audio-recorded using a digital device to allow for verbatim transcription. Audio recordings of the interviews will be deleted on an ongoing basis as soon as the transcription is complete. Each in-depth interview will take approximately 45 min to complete, and participants will receive an incentive of US $40 in the form of an Amazon gift card.

Phase 2 Outcomes

Besides obtaining feedback on box packaging and its contents (including instructions for each specimen self-collection kit), specific domains that will be discussed in the in-depth interviews include— *Information:* (1) familiarity with the concept of biological specimen self-collection and (2) differentiating between rapid home HIV tests and specimen self-collection kits for laboratory testing; *Motivation:* (1) attitudes toward self-collecting and returning each type of biological specimen and (2) factors influencing the actual return or the failure to return certain specimens; and *Behavioral skills:* (1) ease of self-collection from different anatomical sites and (2) personal and logistical barriers (both perceived and actual) encountered.

Phase 2 Analytic Plan

Transcribed in-depth interviews will be checked for accuracy, formatted, and imported into software for qualitative data analysis (MAXQDA). After reading some transcripts, initial codes will be created and categorized under the following theoretical constructs— *Information*: (eg, *familiarity with the concept of biological specimen self-collection*); *Motivation:*

(eg, *willingness to return a finger-stick blood sample*); and *Behavioral skill* s: (eg, *ease of self-collecting a hair sample*). The coding scheme will be systematically reviewed, compared, and refined in team meetings, until we establish an intercoder reliability of ≥0.90. All transcripts will be coded using our agreed-upon coding scheme, and inductive codes will be added throughout the iterative analytic process. New themes that emerge under each construct of the IMB model will be discussed, and the codebook will be adapted as necessary. Although our qualitative analysis will be primarily led by a phenomenological approach (inductive), it will be guided by an underlying conceptual framework (deductive). Thematic analysis will be used to continue analyzing the data until theoretical saturation and redundancy across relevant domains are reached.

Results

Project Caboodle! was funded by the National Institutes of Health (NIH) in May 2018, and participant recruitment began in March 2019. The study team has completed the process of designing a study logo and creating advertisements for social media platforms (see Figure 2).

The Web-based informed consent document, eligibility screener, baseline survey, and posttest survey have been programmed in Qualtrics and are being tested for any inadvertent errors. The step-by-step written instructions accompanied by color images have been finalized by the laboratories that will be conducting biological specimen testing (Emory University Clinical Virology Research Laboratory and the HAL at UCSF). The boxes containing 5 specimen self-collection kits (finger-stick blood sample, pharyngeal swab, rectal swab, urine specimen, and hair sample) affixed with unique identification stickers are being assembled at UMSN in Ann Arbor. UPS shipping procedures (for mailing out boxes to participants), FedEx shipping procedures (for the return of specimens by participants using prepaid shipping envelopes), and payment procedures for completing the incentivized surveys and in-depth interviews (involving Amazon gift cards) are being finalized.

Figure 2. Sample advertisements for Project Caboodle!. STI: sexually transmitted infection.

Discussion

Principal Findings

Project Caboodle! seeks to explore the acceptability and feasibility of self-collecting at home and returning by mail a bundle of 5 different types of biological specimens by young sexually active MSM residing in the United States. The potentially reduced time, expense, and travel associated with this strategy could facilitate a wider implementation of HIV and bacterial STI testing algorithms, as well as PrEP adherence monitoring in this high-risk group. Although some recent studies have focused on the acceptability of self-collecting and returning individual specimens for HIV or other STI testing (as opposed to a bundle) [75-87], none have incorporated PrEP adherence monitoring using hair samples. As hair is nonbiohazardous, easy-to-ship, and stable at room temperature, this specimen is particularly adaptable to self-collection. As PrEP moves from clinical trials into routine practice, these preliminary data can guide the future development of remote monitoring strategies for MSM. Our study will fill a critical gap in knowledge regarding feasibility by comparing the theoretical willingness and perceived ability to self-collect each of the 5 different types of biological specimens with the proportions of specimens actually returned and their adequacy for laboratory testing, respectively. High levels of theoretical willingness may overstate the actual rates of self-collection and return of certain specimens (eg, finger-stick blood, which is invasive to collect) because intentions do not always translate into behavior [107,108]. Obtaining this information is important, as naively assuming that positive intentions would translate into meaningful action could result in an inefficient deployment of limited resources in larger research studies and public health programs. Finally, gathering qualitative data on attitudes toward and barriers to

self-collecting and returning both invasive and noninvasive specimens is critical to developing novel prevention interventions for MSM, with the ultimate goal of reducing their disproportionate burden of HIV and other STIs.

The potential challenges and limitations associated with our study do not outweigh the importance of conducting this research. Using social media platforms to recruit a convenience sample of young, high-risk MSM will reduce the generalizability of our findings. However, given the increasing use of websites and mobile apps by MSM to find sex partners [109-113], we believe this is an important starting point. We recognize practical issues with survey completion such as low or nonresponse, the potential for nondelivery of study boxes, and the reengagement of participants who returned some or no biological specimens for our qualitative phase. Black MSM's general distrust of the research community and heightened perceptions of stigma [114] could result in differential return of specimens. We will not be able to validate the veracity of returned samples (ie, whether they belong to an enrolled participant), but we do not anticipate this to be a major issue as MSM can choose to not return some specimens if they so desire. Finally, although we are confident about our ability to collect data on acceptability and potential return, we acknowledge that assessing specimen adequacy for laboratory testing is dependent solely upon their actual return. We recognize that baldness or very short cropped hair may limit the ability of a participant to self-collect an adequate hair sample but hope to capture this information in our posttest survey that includes questions to elicit reasons for not returning each type of specimen.

Conclusions

Despite these limitations, the self-collection of biological specimens for HIV and bacterial STI testing as well as PrEP adherence monitoring and their return by mail for laboratory testing might offer a practical and convenient solution to improve comprehensive prevention efforts for high-risk MSM. Our results will provide formative data that can be used to plan HIV and bacterial STI prevention programs and remote PrEP monitoring strategies for other minorities at risk, such as transgender men and women, as well as cisgender women [115].

Conflicts of Interest

None declared.

References

1. Grey JA, Bernstein KT, Sullivan PS, Purcell DW, Chesson HW, Gift TL, et al. Estimating the population sizes of men who have sex with men in US states and counties using data from the American Community Survey. JMIR Public Health Surveill 2016;2(1):e14 [FREE Full text] [doi: 10.2196/publichealth.5365] [Medline: 27227149]

2. Purcell D, Johnson C, Lansky A, Prejean J, Stein R, Denning P, et al. Estimating the population size of men who have sex with men in the United States to obtain HIV and syphilis rates. Open AIDS J 2012;6:98-107 [FREE Full text] [doi: 10.2174/1874613601206010098] [Medline: 23049658]

3. Centers for Disease Control and Prevention. 2017. HIV among gay and bisexual men URL: https://www.cdc.gov/nchhstp/newsroom/docs/factsheets/cdc-msm-508.pdf [accessed 2019-03-28] [WebCite Cache ID 75wWqMhuh]

4. Centers for Disease Control and Prevention. 2016. Diagnoses of HIV infection in the United States and dependent areas, 2016 URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2016-vol-28.pdf [accessed 2019-03-28] [WebCite Cache ID 75wX3xucZ]

5. Centers for Disease Control and Prevention. 2016. Trends in US HIV diagnoses, 2005-2014 URL: https://www.cdc.gov/nchhstp/newsroom/docs/factsheets/hiv-data-trends-fact-sheet-508.pdf [accessed 2019-03-28] [WebCite Cache ID 75wXB3UKq]

6. Rosenberg ES, Grey JA, Sanchez TH, Sullivan PS. Rates of prevalent HIV infection, prevalent diagnoses, and new diagnoses among men who have sex with men in US States, metropolitan statistical areas, and counties, 2012-2013. JMIR Public Health Surveill 2016 May 17;2(1):e22 [FREE Full text] [doi: 10.2196/publichealth.5684] [Medline: 27244769]

7. Patton M, Kidd S, Llata E, Stenger M, Braxton J, Asbel L, et al. Extragenital gonorrhea and chlamydia testing and infection among men who have sex with men--STD Surveillance Network, United States, 2010-2012. Clin Infect Dis 2014 Jun;58(11):1564-1570 [FREE Full text] [doi: 10.1093/cid/ciu184] [Medline: 24647015]

8. Bernstein K, Marcus J, Nieri G, Philip S, Klausner J. Rectal gonorrhea and chlamydia reinfection is associated with increased risk of HIV seroconversion. J Acquir Immune Defic Syndr 2010 Apr 01;53(4):537-543. [doi: 10.1097/QAI.0b013e3181c3ef29] [Medline: 19935075]

9. Galvin S, Cohen M. The role of sexually transmitted diseases in HIV transmission. Nat Rev Microbiol 2004 Jan;2(1):33-42. [doi: 10.1038/nrmicro794] [Medline: 15035007]

XSL·FO

RenderX

10. Fleming D, Wasserheit J. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex Transm Infect 1999 Feb;75(1):3-17 [FREE Full text] [Medline: 10448335]

11. Wasserheit J. Epidemiological synergy. Interrelationships between human immunodeficiency virus infection and other sexually transmitted diseases. Sex Transm Dis 1992;19(2):61-77. [Medline: 1595015]

12. Pepin J, Plummer F, Brunham R, Piot P, Cameron W, Ronald A. The interaction of HIV infection and other sexually transmitted diseases: an opportunity for intervention. AIDS 1989 Jan;3(1):3-9. [Medline: 2496709]

13. DiNenno E, Prejean J, Irwin K. Recommendations for HIV screening of gay, bisexual, and other men who have sex with men - United States, 2017. MMWR 2017;66:830-832 [FREE Full text]

14. Workowski KA, Bolan GA, Centers for Disease ControlPrevention. Sexually transmitted diseases treatment guidelines, 2015. MMWR Recomm Rep 2015 Jun 5;64(RR-03):1-137 [FREE Full text] [Medline: 26042815]

15. Branson B, Handsfield H, Lampe M. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. MMWR Recomm Reports 2006;55(14):1-17 [FREE Full text]

16. Sharma A, Wang L, Dunville R, Valencia R, Rosenberg E, Sullivan P. HIV and sexually transmitted disease testing behavior among adolescent sexual minority males: analysis of pooled Youth Risk Behavior Survey data, 2005-2013. LGBT Health 2017 Dec;4(2):130-140. [doi: 10.1089/lgbt.2016.0134] [Medline: 28145800]

17. Centers for Disease Control and Prevention. 2016. HIV infection risk, prevention, and testing behaviors among men who have sex with men: National HIV Behavioral Surveillance, 20 U.S. cities, 2014 URL: http://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-hssr-nhbs-msm-2014.pdf [accessed 2019-03-28] [WebCite Cache ID 75wXFPFFk]

18. Prestage G, Brown G, Keen P. Barriers to HIV testing among Australian gay men. Sex Health 2012 Nov;9(5):453-458. [doi: 10.1071/SH12033] [Medline: 23380195]

19. Margolis AD, Joseph H, Belcher L, Hirshfield S, Chiasson MA. 'Never testing for HIV' among men who have sex with men recruited from a sexual networking website, United States. AIDS Behav 2012 Jan;16(1):23-29. [doi: 10.1007/s10461-011-9883-4] [Medline: 21279431]

20. Li X, Lu H, Raymond HF, Sun Y, Jia Y, He X, et al. Untested and undiagnosed: barriers to HIV testing among men who have sex with men, Beijing, China. Sex Transm Infect 2012 Apr;88(3):187-193. [doi: 10.1136/sextrans-2011-050248] [Medline: 22158932]

21. Mackellar DA, Hou S, Whalen CC, Samuelsen K, Sanchez T, Smith A, WHBS Study Group. Reasons for not HIV testing, testing intentions, and potential use of an over-the-counter rapid HIV test in an internet sample of men who have sex with men who have never tested for HIV. Sex Transm Dis 2011 May;38(5):419-428. [doi: 10.1097/OLQ.0b013e31820369dd] [Medline: 21183863]

22. Finlayson TJ, Le B, Smith A, Bowles K, Cribbin M, Miles I, Centers for Disease ControlPrevention (CDC). HIV risk, prevention, and testing behaviors among men who have sex with men--National HIV Behavioral Surveillance System, 21 U.S. cities, United States, 2008. MMWR Surveill Summ 2011 Oct 28;60(14):1-34 [FREE Full text] [Medline: 22031280]

23. Schwarcz S, Richards T, Frank H, Wenzel C, Hsu LC, Chin CS, et al. Identifying barriers to HIV testing: personal and contextual factors associated with late HIV testing. AIDS Care 2011 Jul;23(7):892-900. [doi: 10.1080/09540121.2010.534436] [Medline: 21424942]

24. Deblonde J, de Koker P, Hamers F, Fontaine J, Luchters S, Temmerman M. Barriers to HIV testing in Europe: a systematic review. Eur J Public Health 2010 Aug;20(4):422-432. [doi: 10.1093/eurpub/ckp231] [Medline: 20123683]

25. Sanchez T, Finlayson T, Drake A, Behel S, Cribbin M, Dinenno E, Centers for Disease ControlPrevention (CDC). Human immunodeficiency virus (HIV) risk, prevention, and testing behaviors--United States, National HIV Behavioral Surveillance System: men who have sex with men, November 2003-April 2005. MMWR Surveill Summ 2006 Jul 07;55(6):1-16 [FREE Full text] [Medline: 16826162]

26. CDC. HIV prevalence, unrecognized infection, and HIV testing among men who have sex with men - Five U.S. cities, June 2004 - April 2005. MMWR 2005;54(24):597-601 [FREE Full text]

27. Lopez-Quintero C, Shtarkshall R, Neumark Y. Barriers to HIV-testing among Hispanics in the United States: analysis of the National Health Interview Survey, 2000. AIDS Patient Care STDS 2005 Oct;19(10):672-683. [doi: 10.1089/apc.2005.19.672] [Medline: 16232051]

28. Spielberg F, Branson B, Goldbaum G, Lockhart D, Kurth A, Celum CL, et al. Overcoming barriers to HIV testing: preferences for new strategies among clients of a needle exchange, a sexually transmitted disease clinic, and sex venues for men who have sex with men. J Acquir Immune Defic Syndr 2003 Mar 1;32(3):318-327. [Medline: 12626893]

29. Kellerman S, Lehman S, Lansky A, Stevens MR, Hecht FM, Bindman AB, et al. HIV testing within at-risk populations in the United States and the reasons for seeking or avoiding HIV testing. J Acquir Immune Defic Syndr 2002 Oct 1;31(2):202-210. [Medline: 12394799]

30. Spielberg F, Kurth A, Gorbach P, Goldbaum G. Moving from apprehension to action: HIV counseling and testing preferences in three at-risk populations. AIDS Educ Prev 2001 Dec;13(6):524-540. [Medline: 11791784]

31. Hoots B, Torrone E, Bernstein K, Paz-Bailey G, NHBS Study Group. Self-reported chlamydia and gonorrhea testing and diagnosis among men who have sex with men - 20 US cities, 2011 and 2014. Sex Transm Dis 2018 Jul;45(7):469-475. [doi: 10.1097/OLQ.0000000000000786] [Medline: 29465659]

32.   Tenover F, Jo Baron E, Gaydos C. Self-collected specimens for infectious disease testing. Clin Microbiol Newsl 2017 Apr;39(7):51-56. [doi: 10.1016/j.clinmicnews.2017.03.004]

33.   Fajardo-Bernal L, Aponte-Gonzalez J, Vigil P, Angel-Müller E, Rincon C, Gaitán HG, et al. Home-based versus clinic-based specimen collection in the management of Chlamydia trachomatis and Neisseria gonorrhoeae infections. Cochrane Database Syst Rev 2015 Sep 29(9):CD011317. [doi: 10.1002/14651858.CD011317.pub2] [Medline: 26418128]

34.   Grant R, Lama J, Anderson P, McMahan V, Liu AY, Vargas L, iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]

35.   Anderson P, Glidden D, Liu A, Buchbinder S, Lama JR, Guanira JV, iPrEx Study Team. Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men. Sci Transl Med 2012 Sep 12;4(151):151ra125 [FREE Full text] [doi: 10.1126/scitranslmed.3004006] [Medline: 22972843]

36.   Elion R, Coleman M. The preexposure prophylaxis revolution: from clinical trials to routine practice: implementation view from the USA. Curr Opin HIV AIDS 2016 Jan;11(1):67-73 [FREE Full text] [doi: 10.1097/COH.0000000000000222] [Medline: 26599165]

37.   Blumenthal J, Haubrich R. Pre-exposure prophylaxis for HIV infection: how antiretroviral pharmacology helps to monitor and improve adherence. Expert Opin Pharmacother 2013 Sep;14(13):1777-1785 [FREE Full text] [doi: 10.1517/14656566.2013.812072] [Medline: 23800167]

38.   Grant R, Anderson P, McMahan V, Liu A, Amico KR, Mehrotra M, iPrEx study team. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829 [FREE Full text] [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]

39.   Marrazzo J, Ramjee G, Richardson B, Gomez K, Mgodi N, Nair G, et al. Tenofovir-based preexposure prophylaxis for HIV infection among African women. N Engl J Med 2015 Feb 5;372(6):509-518. [doi: 10.1056/nejmoa1402269]

40.   van Damme L, Corneli A, Ahmed K, Agot K, Lombaard J, Kapiga S, FEM-PrEP Study Group. Preexposure prophylaxis for HIV infection among African women. N Engl J Med 2012 Aug 2;367(5):411-422 [FREE Full text] [doi: 10.1056/NEJMoa1202614] [Medline: 22784040]

41.   Donnell D, Baeten J, Bumpus N, Brantley J, Bangsberg DR, Haberer JE, et al. HIV protective efficacy and correlates of tenofovir blood concentrations in a clinical trial of PrEP for HIV prevention. J Acquir Immune Defic Syndr 2014 Jul 01;66(3):340-348 [FREE Full text] [doi: 10.1097/QAI.0000000000000172] [Medline: 24784763]

42.   Rotheram-Borus M, Swendeman D, Chovnick G. The past, present, and future of HIV prevention: integrating behavioral, biomedical, and structural intervention strategies for the next generation of HIV prevention. Annu Rev Clin Psychol 2009;5:143-167 [FREE Full text] [doi: 10.1146/annurev.clinpsy.032408.153530] [Medline: 19327028]

43.   Kippax S, Reis E, de Wit J. Two sides to the HIV prevention coin: efficacy and effectiveness. AIDS Educ Prev 2011 Oct;23(5):393-396. [doi: 10.1521/aeap.2011.23.5.393] [Medline: 22010803]

44.   Sullivan P, Carballo-Diéguez A, Coates T, Goodreau SM, McGowan I, Sanders EJ, et al. Successes and challenges of HIV prevention in men who have sex with men. Lancet 2012 Jul 28;380(9839):388-399 [FREE Full text] [doi: 10.1016/S0140-6736(12)60955-6] [Medline: 22819659]

45.   Vermund S, Tique J, Cassell H, Pask ME, Ciampa P, Audet C. Translation of biomedical prevention strategies for HIV: prospects and pitfalls. J Acquir Immune Defic Syndr 2013 Jun 01;63(Suppl 1):S12-S25 [FREE Full text] [doi: 10.1097/QAI.0b013e31829202a2] [Medline: 23673881]

46.   Kubicek K, Arauz-Cuadra C, Kipke M. Attitudes and perceptions of biomedical HIV prevention methods: voices from young men who have sex with men. Arch Sex Behav 2015 Feb;44(2):487-497 [FREE Full text] [doi: 10.1007/s10508-014-0398-8] [Medline: 25633499]

47.   Amico K, Stirratt M. Adherence to preexposure prophylaxis: current, emerging, and anticipated bases of evidence. Clin Infect Dis 2014 Jul;59(Suppl 1):S55-S60 [FREE Full text] [doi: 10.1093/cid/ciu266] [Medline: 24926036]

48.   Koss C, Bacchetti P, Hillier S, Livant E, Horng H, Mgodi N, et al. Differences in cumulative exposure and adherence to tenofovir in the VOICE, iPrEx OLE, and PrEP demo studies as determined via hair concentrations. AIDS Res Hum Retroviruses 2017 Dec;33(8):778-783 [FREE Full text] [doi: 10.1089/aid.2016.0202] [Medline: 28253024]

49.   Gandhi M, Murnane P, Bacchetti P, Elion R, Kolber MA, Cohen SE, et al. Hair levels of preexposure prophylaxis drugs measure adherence and are associated with renal decline among men/transwomen. AIDS 2017 Oct 23;31(16):2245-2251 [FREE Full text] [doi: 10.1097/QAD.0000000000001615] [Medline: 28832411]

50.   Saberi P, Neilands T, Ming K, Johnson MO, Kuncze K, Koss CA, et al. Strong correlation between concentrations of antiretrovirals in home-collected and study-collected hair samples: implications for adherence monitoring. J Acquir Immune Defic Syndr 2017 Dec 1;76(4):e101-e103 [FREE Full text] [doi: 10.1097/QAI.0000000000001492] [Medline: 28657913]

51.   Koss C, Hosek S, Bacchetti P, Anderson PL, Liu AY, Horng H, et al. Comparison of measures of adherence to HIV preexposure prophylaxis among adolescent and young men who have sex with men in the United States. Clin Infect Dis 2018 Jan 6;66(2):213-219 [FREE Full text] [doi: 10.1093/cid/cix755] [Medline: 29020194]

52.   Abaasa A, Hendrix C, Gandhi M, Anderson P, Kamali A, Kibengo F, et al. Utility of different adherence measures for PrEP: patterns and incremental value. AIDS Behav 2018 Dec;22(4):1165-1173 [FREE Full text] [doi: 10.1007/s10461-017-1951-y] [Medline: 29090394]

XSL•FO

RenderX

53. Markowitz M, Grossman H, Anderson P, Grant R, Gandhi M, Horng H, et al. Newly acquired infection with multidrug-resistant HIV-1 in a patient adherent to pre-exposure prophylaxis. J Acquir Immune Defic Syndr 2017 Dec 1;76(4):e104-e106 [FREE Full text] [doi: 10.1097/QAI.0000000000001534] [Medline: 29076941]

54. Seifert S, Castillo-Mancilla J, Erlandson K, Morrow M, Gandhi M, Kuncze K, et al. Brief report: adherence biomarker measurements in older and younger HIV-infected adults receiving tenofovir-based therapy. J Acquir Immune Defic Syndr 2018 Dec 1;77(3):295-298 [FREE Full text] [doi: 10.1097/QAI.0000000000001596] [Medline: 29189417]

55. Gandhi M, Glidden D, Mayer K, Schechter M, Buchbinder S, Grinsztejn B, et al. Association of age, baseline kidney function, and medication exposure with declines in creatinine clearance on pre-exposure prophylaxis: an observational cohort study. Lancet HIV 2016 Dec;3(11):e521-e528 [FREE Full text] [doi: 10.1016/S2352-3018(16)30153-9] [Medline: 27658870]

56. Gandhi M, Glidden D, Liu A, Anderson PL, Horng H, Defechereux P, iPrEx Study Team. Strong correlation between concentrations of tenofovir (TFV) Emtricitabine (FTC) in hair and TFV diphosphate and FTC triphosphate in dried blood spots in the iPrEx open label extension: Implications for pre-exposure prophylaxis adherence monitoring. J Infect Dis 2015 Nov 01;212(9):1402-1406 [FREE Full text] [doi: 10.1093/infdis/jiv239] [Medline: 25895984]

57. Baxi S, Liu A, Bacchetti P, Mutua G, Sanders EJ, Kibengo FM, et al. Comparing the novel method of assessing PrEP adherence/exposure using hair samples to other pharmacologic and traditional measures. J Acquir Immune Defic Syndr 2015 Jan 1;68(1):13-20 [FREE Full text] [doi: 10.1097/QAI.0000000000000386] [Medline: 25296098]

58. Liu A, Yang Q, Huang Y, Bacchetti P, Anderson PL, Jin C, et al. Strong relationship between oral dose and tenofovir hair levels in a randomized trial: hair as a potential adherence measure for pre-exposure prophylaxis (PrEP). PLoS One 2014;9(1):e83736 [FREE Full text] [doi: 10.1371/journal.pone.0083736] [Medline: 24421901]

59. Thaden J, Gandhi M, Okochi H, Hurt C, McKellar M. Seroconversion on preexposure prophylaxis: a case report with segmental hair analysis for timed adherence determination. AIDS 2018 Jun 1;32(9):F1-F4 [FREE Full text] [doi: 10.1097/QAD.0000000000001825] [Medline: 29683856]

60. Cohen S, Sachdev D, Lee S, Scheer S, Bacon O, Chen M, et al. Acquisition of tenofovir-susceptible, emtricitabine-resistant HIV despite high adherence to daily pre-exposure prophylaxis: a case report. Lancet HIV 2019 Jan;6(1):e43-e50. [doi: 10.1016/S2352-3018(18)30288-1]

61. Smieja M, Castriciano S, Carruthers S, So G, Chong S, Luinstra K, et al. Development and evaluation of a flocked nasal midturbinate swab for self-collection in respiratory virus infection diagnostic testing. J Clin Microbiol 2010 Sep;48(9):3340-3342 [FREE Full text] [doi: 10.1128/JCM.02235-09] [Medline: 20610685]

62. Dhiman N, Miller R, Finley J, Sztajnkrycer MD, Nestler DM, Boggust AJ, et al. Effectiveness of patient-collected swabs for influenza testing. Mayo Clin Proc 2012 Jun;87(6):548-554 [FREE Full text] [doi: 10.1016/j.mayocp.2012.02.011] [Medline: 22551906]

63. Murray M, Schulz L, Furst J, Homme JH, Jenkins SM, Uhl JR, et al. Equal performance of self-collected and health care worker-collected pharyngeal swabs for group a streptococcus testing by PCR. J Clin Microbiol 2015 Feb;53(2):573-578 [FREE Full text] [doi: 10.1128/JCM.02500-14] [Medline: 25502528]

64. Sellors J, Lorincz A, Mahony J, Mielzynska I, Lytwyn A, Roth P, et al. Comparison of self-collected vaginal, vulvar and urine samples with physician-collected cervical samples for human papillomavirus testing to detect high-grade squamous intraepithelial lesions. CMAJ 2000 Sep 05;163(5):513-518 [FREE Full text] [Medline: 11006761]

65. Garrow S, Smith D, Harnett G. The diagnosis of chlamydia, gonorrhoea, and trichomonas infections by self obtained low vaginal swabs, in remote northern Australian clinical practice. Sex Transm Infect 2002 Aug;78(4):278-281 [FREE Full text] [Medline: 12181467]

66. Crucitti T, Van Dyck E, Tehe A, Abdellati S, Vuylsteke B, Buve A, et al. Comparison of culture and different PCR assays for detection of Trichomonas vaginalis in self collected vaginal swab specimens. Sex Transm Infect 2003 Oct;79(5):393-398 [FREE Full text] [Medline: 14573835]

67. Schachter J, Chernesky M, Willis D, Fine PM, Martin DH, Fuller D, et al. Vaginal swabs are the specimens of choice when screening for Chlamydia trachomatis and Neisseria gonorrhoeae: results from a multicenter evaluation of the APTIMA assays for both infections. Sex Transm Dis 2005 Dec;32(12):725-728. [Medline: 16314767]

68. Belinson J, Hu S, Niyazi M, Pretorius RG, Wang H, Wen C, et al. Prevalence of type-specific human papillomavirus in endocervical, upper and lower vaginal, perineal and vaginal self-collected specimens: Implications for vaginal self-collection. Int J Cancer 2010 Sep 01;127(5):1151-1157 [FREE Full text] [doi: 10.1002/ijc.25144] [Medline: 20039323]

69. Chernesky M, Jang D, Gilchrist J, Randazzo J, Elit L, Lytwyn A, et al. Ease and comfort of cervical and vaginal sampling for Chlamydia trachomatis and Trichomonas vaginalis with a new Aptima specimen collection and transportation kit. J Clin Microbiol 2014 Feb;52(2):668-670 [FREE Full text] [doi: 10.1128/JCM.02923-13] [Medline: 24478509]

70. Boggan J, Walmer D, Henderson G, Chakhtoura N, McCarthy SH, Beauvais HJ, et al. Vaginal self-sampling for HPV infection as a primary cervical cancer screening tool in a Haitian population. Sex Transm Dis 2015 Nov;42(11):655-659 [FREE Full text] [doi: 10.1097/OLQ.0000000000000345] [Medline: 26462192]

71. Arias M, Jang D, Gilchrist J, Luinstra K, Li J, Smieja M, et al. Ease, comfort, and performance of the HerSwab vaginal self-sampling device for the detection of Chlamydia trachomatis and Neisseria gonorrhoeae. Sex Transm Dis 2016 Feb;43(2):125-129. [doi: 10.1097/OLQ.0000000000000406] [Medline: 26760183]

72. Hernandez BY, McDuffie K, Goodman MT, Wilkens LR, Thompson P, Zhu X, et al. Comparison of physician- and self-collected genital specimens for detection of human papillomavirus in men. J Clin Microbiol 2006 Feb;44(2):513-517 [FREE Full text] [doi: 10.1128/JCM.44.2.513-517.2006] [Medline: 16455906]

73. Ogilvie GS, Taylor DL, Achen M, Cook D, Krajden M. Self-collection of genital human papillomavirus specimens in heterosexual men. Sex Transm Infect 2009 Jun;85(3):221-225. [doi: 10.1136/sti.2008.033068] [Medline: 19066196]

74. Hernandez BY, Wilkens LR, Unger ER, Steinau M, Markowitz L, Garvin K, et al. Evaluation of genital self-sampling methods for HPV detection in males. J Clin Virol 2013 Sep;58(1):168-175. [doi: 10.1016/j.jcv.2013.06.032] [Medline: 23880161]

75. Tobin K, Edwards C, Flath N, Lee A, Tormohlen K, Gaydos CA. Acceptability and feasibility of a Peer Mentor program to train young Black men who have sex with men to promote HIV and STI home-testing to their social network members. AIDS Care 2018 Jul;30(7):896-902 [FREE Full text] [doi: 10.1080/09540121.2018.1442553] [Medline: 29482342]

76. Pearson WS, Kreisel K, Peterman TA, Zlotorzynska M, Dittus PJ, Habel MA, et al. Improving STD service delivery: would American patients and providers use self-tests for gonorrhea and chlamydia? Prev Med 2018 Oct;115:26-30. [doi: 10.1016/j.ypmed.2018.08.007] [Medline: 30096329]

77. Grov C, Cain D, Whitfield THF, Rendina HJ, Pawson M, Ventuneac A, et al. Recruiting a U.S. national sample of HIV-negative gay and bisexual men to complete at-home self-administered HIV/STI testing and surveys: challenges and opportunities. Sex Res Social Policy 2016 Mar 01;13(1):1-21 [FREE Full text] [doi: 10.1007/s13178-015-0212-y] [Medline: 26858776]

78. Grov C, Cain D, Rendina HJ, Ventuneac A, Parsons JT. Characteristics associated with urethral and rectal gonorrhea and chlamydia diagnoses in a US national sample of gay and bisexual men: results from the One Thousand Strong panel. Sex Transm Dis 2016 Mar;43(3):165-171 [FREE Full text] [doi: 10.1097/OLQ.0000000000000410] [Medline: 26859803]

79. McRee A, Esber A, Reiter PL. Acceptability of home-based chlamydia and gonorrhea testing among a national sample of sexual minority young adults. Perspect Sex Reprod Health 2015 Mar;47(1):3-10 [FREE Full text] [doi: 10.1363/47e2715] [Medline: 25776809]

80. Fisher M, Wayal S, Smith H, Llewellyn C, Alexander S, Ison C, Brighton Home Sampling Kit Project Steering Group. Home sampling for sexually transmitted infections and HIV in men who have sex with men: a prospective observational study. PLoS One 2015;10(4):e0120810 [FREE Full text] [doi: 10.1371/journal.pone.0120810] [Medline: 25848769]

81. Sharma A, Stephenson R, White D, Sullivan P. Acceptability and intended usage preferences for six HIV testing options among internet-using men who have sex with men. SpringerPlus 2014;3(1):109. [Medline: 24600551]

82. Odesanmi T, Wasti S, Odesanmi O, Adegbola O, Oguntuase O, Mahmood S. Comparative effectiveness and acceptability of home-based and clinic-based sampling methods for sexually transmissible infections screening in females aged 14-50 years: a systematic review and meta-analysis. Sex Health 2013 Dec;10(6):559-569. [doi: 10.1071/SH13029] [Medline: 24160747]

83. Götz HM, Wolfers M, Luijendijk A, van den Broek IV. Retesting for genital Chlamydia trachomatis among visitors of a sexually transmitted infections clinic: randomized intervention trial of home- versus clinic-based recall. BMC Infect Dis 2013 May 24;13:239 [FREE Full text] [doi: 10.1186/1471-2334-13-239] [Medline: 23705624]

84. Reagan M, Xu H, Shih S, Secura G, Peipert J. A randomized trial of home versus clinic-based sexually transmitted disease screening among men. Sex Transm Dis 2012 Nov;39(11):842-847 [FREE Full text] [doi: 10.1097/OLQ.0b013e3182649165] [Medline: 23064532]

85. Xu F, Stoner B, Taylor S, Mena L, Tian LH, Papp J, et al. Use of home-obtained vaginal swabs to facilitate rescreening for Chlamydia trachomatis infections: two randomized controlled trials. Obstet Gynecol 2011 Aug;118(2 Pt 1):231-239. [doi: 10.1097/AOG.0b013e3182246a83] [Medline: 21775837]

86. Sharma A, Sullivan P, Khosropour C. Willingness to take a free home HIV test and associated factors among internet-using men who have sex with men. J Int Assoc Physicians AIDS Care (Chic) 2011;10(6):357-364 [FREE Full text] [doi: 10.1177/1545109711404946] [Medline: 21527425]

87. Graseck A, Secura G, Allsworth J, Madden T, Peipert J. Home compared with clinic-based screening for sexually transmitted infections: a randomized controlled trial. Obstet Gynecol 2010 Dec;116(6):1311-1318 [FREE Full text] [doi: 10.1097/AOG.0b013e3181fae60d] [Medline: 21099596]

88. Fisher J, Fisher W, Shuper P. The information-motivation-behavioral skills model of HIV preventive behavior. In: Emerging Theories in Health Promotion Practice and Research. San Francisco, CA, US: Jossey-Bass; 2009:21-64.

89. Fisher W, Fisher J, Harman J. Semantic Scholar. 2003. The information-motivation-behavioral skills model: A general social psychological approach to understanding and promoting health behavior URL: https://pdfs.semanticscholar.org/0fb8/76145ee4c823c57a3aa7570ff1490978ef97.pdf [accessed 2019-04-19] [WebCite Cache ID 77k9InSMk]

90. Sharma A, Chavez P, MacGowan R, McNaghten AD, Mustanski B, Gravens L, et al. Willingness to distribute free rapid home HIV test kits and to test with social or sexual network associates among men who have sex with men in the United States. AIDS Care 2017 Dec;29(12):1499-1503. [doi: 10.1080/09540121.2017.1313386] [Medline: 28393612]

91. Sharma A, Sullivan S, Stephenson R. Detailed knowledge about HIV epidemiology and transmission dynamics and their associations with preventive and risk behaviors among gay, bisexual, and other men who have sex with men in the United

XSL•FO

RenderX

States. JMIR Public Health Surveill 2017 Mar 6;3(1):e11 [FREE Full text] [doi: 10.2196/publichealth.7255] [Medline: 28264795]

92. Jaworski B, Carey M. Development and psychometric evaluation of a self-administered questionnaire to measure knowledge of sexually transmitted diseases. AIDS Behav 2007 Jul;11(4):557-574 [FREE Full text] [doi: 10.1007/s10461-006-9168-5] [Medline: 17016760]

93. Centers for Disease Control and Prevention. CDC HIV Risk Reduction Tool URL: https://wwwn.cdc.gov/hivrisk/ [accessed 2019-03-28] [WebCite Cache ID 77DW69v70]

94. Paz-Bailey G, Raymond H, Lansky A, Mermin J. Using the National HIV Behavioral Surveillance System to inform HIV prevention efforts in the United States. AIDS Behav 2014 Apr;18(Suppl 3):S233-S236. [doi: 10.1007/s10461-014-0738-7] [Medline: 24659359]

95. Mimiaga M, Case P, Johnson C, Safren S, Mayer K. Preexposure antiretroviral prophylaxis attitudes in high-risk Boston area men who report having sex with men: limited knowledge and experience but potential for increased utilization after education. J Acquir Immune Defic Syndr 2009 Jan 1;50(1):77-83 [FREE Full text] [doi: 10.1097/QAI.0b013e31818d5a27] [Medline: 19295337]

96. Bauermeister J, Meanley S, Pingel E, Soler J, Harper G. PrEP awareness and perceived barriers among single young men who have sex with men. Curr HIV Res 2013 Oct;11(7):520-527 [FREE Full text] [Medline: 24476355]

97. Mustanski B, Johnson A, Garofalo R, Ryan D, Birkett M. Perceived likelihood of using HIV pre-exposure prophylaxis medications among young men who have sex with men. AIDS Behav 2013 Jul;17(6):2173-2179 [FREE Full text] [doi: 10.1007/s10461-012-0359-y] [Medline: 23128980]

98. Mimiaga M, Reisner S, Bland S. Health system and personal barriers resulting in decreased utilization of HIV and STD testing services among at-risk black men who have sex with men in Massachusetts. AIDS Patient Care and STDs 2009;23(10):825-835. [doi: 10.1089/apc.2009.0086]

99. Schmidt A, Hickson F, Weatherburn P, Marcus U, EMIS Network. Comparison of the performance of STI screening services for gay and bisexual men across 40 European cities: results from the European MSM Internet Survey. Sex Transm Infect 2013 Nov;89(7):575-582. [doi: 10.1136/sextrans-2012-050973] [Medline: 23744961]

100. Bauermeister J. Sexual partner typologies among single young men who have sex with men. AIDS Behav 2015 Jun;19(6):1116-1128 [FREE Full text] [doi: 10.1007/s10461-014-0932-7] [Medline: 25358726]

101. Carballo-Diéguez A, Dolezal C. HIV risk behaviors and obstacles to condom use among Puerto Rican men in New York City who have sex with men. Am J Public Health 1996 Nov;86(11):1619-1622. [Medline: 8916531]

102. Kahle E, Sharma A, Sullivan S, Stephenson R. HIV prioritization and risk perception among an online sample of men who have sex with men in the United States. Am J Mens Health 2018 Dec;12(4):676-687 [FREE Full text] [doi: 10.1177/1557988318776581] [Medline: 29781331]

103. Thompson H, Valdimarsdottir H, Winkel G, Jandorf L, Redd W. The Group-Based Medical Mistrust Scale: psychometric properties and association with breast cancer screening. Prev Med 2004 Feb;38(2):209-218. [Medline: 14715214]

104. Björgvinsson T, Kertz S, Bigda-Peyton J, McCoy K, Aderka I. Psychometric properties of the CES-D-10 in a psychiatric sample. Assessment 2013 Aug;20(4):429-436. [doi: 10.1177/1073191113481998] [Medline: 23513010]

105. United Way 2-1-1. URL: http://www.211.org/ [accessed 2019-03-28] [WebCite Cache ID 77DWJYmSj]

106. AIDSVu. URL: https://aidsvu.org/ [accessed 2019-03-28] [WebCite Cache ID 77DWHJL15]

107. Norberg P, Horne D, Horne D. The privacy paradox: personal information disclosure intentions versus behaviors. J Consum Affair 2007;41(1):100-126. [doi: 10.1111/j.1745-6606.2006.00070.x]

108. McEachan R, Conner M, Taylor N, Lawton R. Prospective prediction of health-related behaviours with the theory of planned behaviour: a meta-analysis. Health Psychol Rev 2011 Sep;5(2):97-144. [doi: 10.1080/17437199.2010.521684]

109. Liau A, Millett G, Marks G. Meta-analytic examination of online sex-seeking and sexual risk behavior among men who have sex with men. Sex Transm Dis 2006 Sep;33(9):576-584. [doi: 10.1097/01.olq.0000204710.35332.c5] [Medline: 16540884]

110. Garofalo R, Herrick A, Mustanski B, Donenberg G. Tip of the Iceberg: young men who have sex with men, the internet, and HIV risk. Am J Public Health 2007 Jun;97(6):1113-1117. [doi: 10.2105/AJPH.2005.075630] [Medline: 17463378]

111. Rosser B, Oakes J, Horvath K, Konstan J, Danilenko G, Peterson J. HIV sexual risk behavior by men who use the Internet to seek sex with men: results of the Men's INTernet Sex Study-II (MINTS-II). AIDS Behav 2009 Jun;13(3):488-498 [FREE Full text] [doi: 10.1007/s10461-009-9524-3] [Medline: 19205866]

112. Landovitz R, Tseng C, Weissman M, Haymer M, Mendenhall B, Rogers K, et al. Epidemiology, sexual risk behavior, and HIV prevention practices of men who have sex with men using GRINDR in Los Angeles, California. J Urban Health 2013 Aug;90(4):729-739 [FREE Full text] [doi: 10.1007/s11524-012-9766-7] [Medline: 22983721]

113. Rendina H, Jimenez R, Grov C, Ventuneac A, Parsons J. Patterns of lifetime and recent HIV testing among men who have sex with men in New York City who use Grindr. AIDS Behav 2014 Jan;18(1):41-49 [FREE Full text] [doi: 10.1007/s10461-013-0573-2] [Medline: 23925515]

114. Malebranche D, Peterson J, Fullilove R, Stackhouse R. Race and sexual identity: perceptions about medical culture and healthcare among Black men who have sex with men. J Natl Med Assoc 2004 Jan;96(1):97-107. [Medline: 14746359]

115.   Sheth A, Rolle C, Gandhi M. HIV pre-exposure prophylaxis for women. J Virus Erad 2016 Jul 1;2(3):149-155 [FREE Full text] [Medline: 27482454]

Abbreviations

CDC: Centers for Disease Control and Prevention
FTC: emtricitabine
HAL: Hair Analytical Laboratory
IMB: Information-Motivation-Behavioral skills
IP: Internet Protocol
MSM: men who have sex with men
NIH: National Institutes of Health
PrEP: pre-exposure prophylaxis
STD-KQ: Sexually Transmitted Disease Knowledge Questionnaire
STI: sexually transmitted infection
TFV: tenofovir
UCSF: University of California, San Francisco
UMSN: University of Michigan School of Nursing
UPS: United Parcel Service

XSL•FO
RenderX

Protocol

Use of Big Data and Machine Learning Methods in the Monitoring and Evaluation of Digital Health Programs in India: An Exploratory Protocol

Diwakar Mohan[1], MD, DrPH; Jean Juste Harrisson Bashingwa[2], PhD; Pierre Dane[3], MS; Sara Chamberlain[4], MA; Nicki Tiffin[2,5], MPH, PhD; Amnesty Lefevre[1,3], MHS, PhD

[1]Department of International Health, Global Disease Epidemiology & Control Program, Johns Hopkins Bloomberg School Public Health, Baltimore, MD, United States

[2]Faculty of Health Sciences, Department of Integrative Biomedical Sciences, & Member of the Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

[3]Division of Epidemiology and Biostatistics, School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa

[4]British Broadcasting Corporation Media Action, New Delhi, India

[5]Centre for Infectious Disease Research in Africa, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

Corresponding Author:
Diwakar Mohan, MD, DrPH
Department of International Health
Global Disease Epidemiology & Control Program
Johns Hopkins Bloomberg School Public Health
615 N Wolfe Street
E5614
Baltimore, MD, 21205
United States
Phone: 1 4105024545
Email: dr.dmohan79@gmail.com

Abstract

Background: Digital health programs, which encompass the subsectors of health information technology, mobile health, electronic health, telehealth, and telemedicine, have the potential to generate "big data."

Objective: Our aim is to evaluate two digital health programs in India—the maternal mobile messaging service (Kilkari) and the mobile training resource for frontline health workers (Mobile Academy). We illustrate possible applications of machine learning for public health practitioners that can be applied to generate evidence on program effectiveness and improve implementation. Kilkari is an outbound service that delivers weekly gestational age–appropriate audio messages about pregnancy, childbirth, and childcare directly to families on their mobile phones, starting from the second trimester of pregnancy until the child is one year old. Mobile Academy is an Interactive Voice Response audio training course for accredited social health activists (ASHAs) in India.

Methods: Study participants include pregnant and postpartum women (Kilkari) as well as frontline health workers (Mobile Academy) across 13 states in India. Data elements are drawn from system-generated databases used in the routine implementation of programs to provide users with health information. We explain the structure and elements of the extracted data and the proposed process for their linkage. We then outline the various steps to be undertaken to evaluate and select final algorithms for identifying gaps in data quality, poor user performance, predictors for call receipt, user listening levels, and linkages between early listening and continued engagement.

Results: The project has obtained the necessary approvals for the use of data in accordance with global standards for handling personal data. The results are expected to be published in August/September 2019.

Conclusions: Rigorous evaluations of digital health programs are limited, and few have included applications of machine learning. By describing the steps to be undertaken in the application of machine learning approaches to the analysis of routine system-generated data, we aim to demystify the use of machine learning not only in evaluating digital health education programs but in improving their performance. Where articles on analysis offer an explanation of the final model selected, here we aim to

XSL•FO
RenderX

emphasize the process, thereby illustrating to program implementors and evaluators with limited exposure to machine learning its relevance and potential use within the context of broader program implementation and evaluation.

International Registered Report Identifier (IRRID): DERR1-10.2196/11456

KEYWORDS

Introduction

Machine learning is an application of artificial intelligence that aims to allow computers to learn automatically from the analysis of large, highly granular datasets, with minimal human intervention [1]. In machine learning, models are created using existing data to make predictions about future events. Applications of machine learning in global public health are emerging, particularly in the context of digital health solutions, which have the potential to generate "big data." Digital health encompasses the subsectors of health information technology, mobile health (mHealth), electronic health (eHealth), telehealth, and telemedicine.

Machine learning approaches in digital health have been mainly in the area of analyzing data generated by wearable sensors and accelerometers including efforts to predict and personalize monitoring systems for mobile patients [2], predict physical activity type and energy expenditure [3-5], and falls [6]. Beyond analyses of accelerometer data, studies have explored user engagement with different apps on mobile phones and tablets [7], and responses to patient feedback on health services [8]. Similar applications to social media data have sought to improve the detection of online illegal drug sales [9], depression [10,11], as well as explore vaccination sentiment trends and improve disease identification [12]. Analyses of geographic information system data have been used to map risk of exposure to disease [13]. Collectively these varied applications of machine learning have been classified by Mooney et al in three broad categories of (1) surveillance, including systems to monitor trends in disease incidence, health behaviors, and environmental conditions, (2) hypothesis-generating research, and (3) causal inference [1].

Evidence gathering on the effectiveness of digital health solutions is a growing field [14]. However, very few evaluations have sought to incorporate machine learning algorithms [8], and broader guidelines on the monitoring and evaluation of digital health solutions have stopped short of outlining options for predictive modeling [15]. Appropriate implementation of

machine learning methodologies can facilitate dynamic, real-time interventions to improve data collection to assess program effectiveness and can also inform how data are used prospectively to improve program implementation.

In this paper, we outline methods proposed for the application of machine learning to the evaluation of two large-scale mHealth initiatives in India, which have scaled up to over 13 states in India since their initiation in 2012-2013. Kilkari is an outbound service that delivers weekly gestational age–appropriate audio messages about pregnancy, childbirth, and childcare directly to families on their mobile phones, starting from the second trimester of pregnancy until the child is one year old. Accredited social health activists (ASHA) mobilize women in the community to attend outreach and primary health center activities where auxiliary nurse midwives collect and register details of mothers and their pregnancies and, after delivery, children born in their catchment areas. Mobile Academy is an interactive voice response (IVR) audio training course for ASHAs in India. The training material delivered over the phone is designed to refresh their knowledge of life-saving preventative health behaviors and improve their interpersonal communications skills.

Through the use of these two digital health examples—the maternal mobile messaging service (Kilkari) and the mobile training resource for frontline health workers (Mobile Academy)—we illustrate possible applications of machine learning that can be applied to generate evidence on effectiveness, as well as to more broadly improve program implementation. We intend this paper to be an easy reference for public health practitioners considering the applicability of machine learning for digital health solutions, rather than a comprehensive review of the field of machine learning. In the course of the paper, we seek to provide an explanation in layman's terms of the methods under consideration, for an audience unfamiliar with machine learning algorithms or advanced statistical methodologies. We also provide references throughout the text for further in-depth reading. Textbox 1 outlines the possible research questions that can be addressed by the study.

Textbox 1. Study aims.

Aim 1: Describe subscriber losses at different points during pregnancy and postpartum period from the program

Objective 1a. Determine differences in subscribers who remain subscribed to Kilkari throughout the duration of service versus those lost at different points along the continuum of program databases.

Objective 1b. Develop a classifier for identifying different categories of losses.

Aim 2: Facilitate the program's ability to identify and target ASHAs likely to perform poorly on digital health training programs and knowledge assessments

Objective 2a. Determine predictors of training course completion (overall and time to completion) by ASHAs based on performance on early modules of the course and other characteristics including reported motivation, knowledge, individual characteristics, and mobile literacy.

Objective 2b. Develop a classifier for the routine identification of ASHAs likely to perform poorly.

Aim 3: Understand the factors underpinning successful receipt of calls (are calls received?)

Objective 3a. Determine what proportion of calls successfully reach the end-user's device for Kilkari.

Objective 3b. Identify the proportion of content specific to infant feeding and family planning received by end-users for Kilkari.

Aim 4. Determine how users are listening to messages

Objective 4a. Determine predictors for exposure to Kilkari content based on user characteristics.

Objective 4b. Measure exposure to Kilkari content based on technological and behavioral (end-user engagement) performance.

Objective 4c. Develop a classifier for measuring listening levels.

Aim 5: Understand optimal message delivery options for maximal impact

Objective 5a. Determine message delivery options with the most success in engaging client and assess patterns of listening.

Objective 5b. Determine the effect of early listening patterns (time of day of listening, duration and frequency of listening, content listening patterns) on postpartum engagement and overall exposure.

Methods

Summary

We present the methods section in parts. We first present a detailed description of the data we plan to use as our source including the architecture of the databases and data elements. Program data are currently held in different databases located in Gurugram and call data records are held in the Mobile Network Operator's datacenter in Delhi. Next, we provide a description of the data munging (ie, data wrangling) and analysis methods including a brief description of the various machine algorithms under consideration.

Data Sources and Flow

Auxiliary nurse midwives collect and register details of pregnant women and, after delivery, of postpartum women and children born in their catchment areas. These data are captured in print registers and uploaded at the block level by data entry operators, forming the data in the pregnancy tracking databases. The data collected include personal identifiers such as geographic location, names of women and a mandatory mobile phone number, and where available, details of the pregnancy and childbirth. Data capture happens at two key time points: (1) the earliest is the registration of the woman at the time of the identification of pregnancy, and (2) following childbirth, when the details for delivery care are available. In actual practice,

these events may happen many days or months after the event (pregnancy registration or birth of child) has happened.

Figure 1 summarizes the databases and flow of data for both Mobile Academy and Kilkari. The following are existing databases:

1. State-based databases that pre-date existing Mother-Child Tracking System (MCTS) and are integrated with MCTS
2. MCTS database at the national level
3. Reproductive Child Health (RCH) database at the national level
4. Call data records captured by Mobile Network Operator stored separately in their databases and used mainly for billing purposes
5. Call data records of Kilkari subscribers and ASHAs' usage of Mobile Academy captured by the IVR system and stored in the IVR database in a data center contracted by the government
6. Mobile Technology for Community Health (MOTECH) database, which is integrated with the RCH and MCTS databases and integrates information from call data records with a small set of MCTS and RCH data for each user
7. Management Information System database, which extracts data from MOTECH and generates reports. The sampling frame for both Mobile Academy and Kilkari are derived from the MCTS and RCH data. Data from the RCH and MCTS databases are pulled into the MOTECH database on a predetermined schedule every day.

Figure 1. Summary of data flow for Kilkari and Mobile Academy.
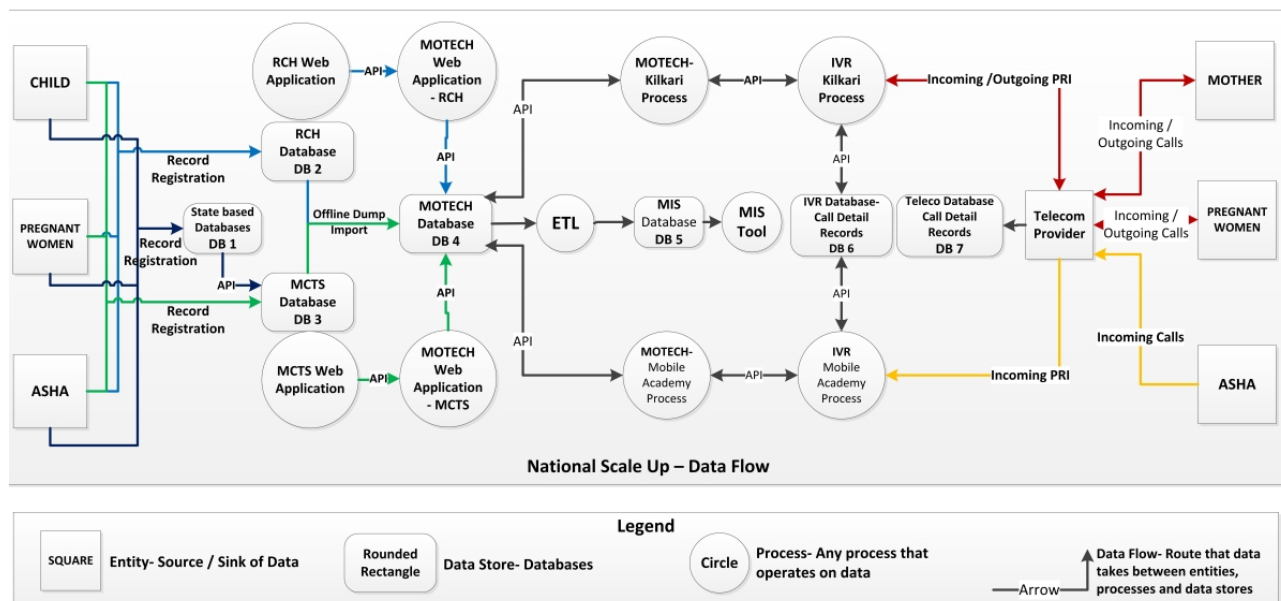
Ethical Considerations

The registration data on pregnant women and ASHAs are collected by the Ministry of Health and Family Welfare of the Government of India and the ministries of health of the states participating in the program. The data will be analyzed under a data sharing agreement with the Bill & Melinda Gates Foundation and Johns Hopkins University, University of Cape Town, and BBC Media Action. The Institutional Review Boards of Johns Hopkins School of Public Health, Sigma in New Delhi, India, and the University of Cape Town have provided the ethical certification for the study.

Data Processing

For the Kilkari program, the pregnant women or postpartum women's data are captured in the RCH and MCTS systems, or in state-based systems that then pass data to RCH or MCTS, and from there to the MOTECH system. Before the data are accepted by MOTECH, the system automatically runs validations to check that the mobile numbers are in the correct format, locations match location masters in the MOTECH database, and last menstrual period and date of birth are within the Kilkari timeframe. The MOTECH system uses the last menstrual period or the delivery date to determine the schedule of messages to be delivered. The MOTECH engine provides the list of phone numbers (clients) to be called each day to the IVR system, which then calls the numbers and plays the appropriate pre-recorded message, which is stored in the IVR system's content management system. If the call is not answered, then the IVR system attempts to call again at least 3 times every day for 4 days until the call is answered.

For the Mobile Academy program, details on ASHAs including their names, phone numbers, geographic location, and age are contained in either the RCH or MCT databases, or in state-based databases integrated with MCTS and used to register them to Mobile Academy. The MOTECH engine captures these data on ASHAs from the RCH or MCTS databases and following registration to Mobile Academy, ASHAs are eligible to call in

to the IVR system using the same phone number provided in the RCH database. The IVR system validates the phone number against the MOTECH system and then retrieves the "bookmark" information that details the status of the ASHA and her progress on the list of content expected to be covered. Based on this information, the appropriate content is delivered to the ASHA via the IVR system and the updated data return to the MOTECH database.

Analysis

The data from the databases (Figure 1) will be extracted onto secure password-protected hard drives from each server storage. Merging data files will be complex given the nature of identifiers across databases. An MCTS record does not have a beneficiary ID. Instead, it has a "Mother" (pregnancy) ID or a "Child" ID. In other words, MCTS tracks pregnancies and births, rather than women. When Kilkari first went live in October 2015, it mirrored the MCTS approach and generated subscription IDs for each pregnancy and then birth. However, the new RCH database does have a unique beneficiary ID, which enables the system to track an individual woman through her multiple pregnancies and the births of children. The architecture of the MOTECH database and Kilkari was changed in December 2016 to introduce a unique beneficiary ID and MOTECH was then integrated with RCH in mid-2017. There is an additional complexity, namely that MOTECH used to allow multiple Kilkari subscriptions on one mobile number, assuming a single phone could be shared by a number of women in a joint family. However, a decision was made to remove this feature in 2017 (July 28 for RCH and October 6 for MCTS) due to the complexity it created in analyzing system-generated data. Hence the analytic time horizon assumed in the analysis may span from 2017-2018 after the MCTS-RCH integration occurred and the aforementioned changes were made. The merging of datasets will occur in India, and only de-identified data will be stored on the hard drives and used in this analysis. As part of Study Aim 1, we will examine the quality of the data for completeness, including patterns and any geographic clustering in missingness.

Additional Data From Baseline Surveys of Evaluation of Kilkari and Mobile Academy

Analyses described in this section are being carried out as part of a larger external evaluation of Kilkari. We describe concurrent efforts to undertake a randomized controlled trial (RCT) in the state of Madhya Pradesh (MP) for Kilkari, inclusive of baseline surveys with pregnant and postpartum women, and ASHA workers. Once identified as part of baseline survey activities and randomized to receive Kilkari content (or no content at all), phone numbers will be fed directly into the MOTECH database for provision of program services. For pregnant women, additional data collected as part of baseline household surveys include demographic factors (age, education, parity, literacy), socioeconomic characteristics (household assets, conditions), health care seeking and practices, as well as data on digital literacy and phone access. These data can be linked to MOTECH, IVR, and call center records to provide additional data elements. Overall, these data as well as data on technology performance (receipt of messages) and user engagement (behavioral performance) with content will help estimate exposure to Kilkari used in the assessment of causality as part of the RCT. For ASHAs, baseline survey data will include similar data elements on demographic, socioeconomic, and mobile literacy and phone access as well as knowledge and work-related variables linked to reported motivation and satisfaction. Overall, these added data elements can be linked to IVR and call record data for this subpopulation of Mobile Academy and Kilkari users in four districts of MP where the RCT is underway.

Data Processing and Analysis

Descriptive statistics, including univariate plots like histograms, will be used to understand the distribution of each variable, including skewness and outliers. Multivariate plots like scatterplots and

locally weighted scatterplot smoothing (LOWESS) lines will be used to understand the relationships between different variables. Efforts to prepare the data are divided into two parts: splitting data into training and testing groups, and data processing.

Splitting Into Training and Testing

To avoid overfitting models that work well for the data in hand but fail to predict well with other datasets, the data will be split into three components. This is possible due to the large size of the dataset. The training set will comprise 60%, the test dataset 20%, and the validation dataset 20% of the data. The test set will be used to test and fine-tune the accuracy of predictive models, and the final selected model will be applied to the validation dataset. We anticipate having data from 2017, 2018, and 2019 and will ensure equal representation by random sampling. To ensure that the data are controlled for time as a confounder, subsets will be equally represented across different time periods.

Processing

Data processing is the act of preparing the data from its raw format into a usable format by the machine learning models.

Indications for data processing will include (1) making the data easier to use; new indicators will need to be created to facilitate their use as predictors, (2) reducing computational cost of many algorithms by decreasing the number of variables, especially correlated and collinear variables, (3) removing noise due to outliers, and (4) making the results easier to understand.

Algorithms

The most common methods by which algorithms learn about data to make predictions are supervised, unsupervised, and semisupervised learning [1]. Supervised learning trains algorithms using example input and output data, previously labeled by humans. Data may be labeled—a term used to denote that the outcome (or class) is known (eg, ASHA has completed the training module or not completed the training module)—or unlabeled. In contrast, unsupervised learning is concerned with uncovering structure and patterns within complex datasets based on information that is neither classified nor labeled. In unsupervised machine learning, the algorithms learn to infer structure based on unlabeled input data using clustering techniques. Semisupervised learning is a hybrid analytic technique, applied in contexts where the majority of data points are missing outcome information and yet prediction remains the goal [1].

In this program context, supervised machine learning algorithms are expected to be the primary analytic method employed because analyses are focused on classification using predictors and available data are expected to be labeled. The transformation of variables may be achieved by a variety of techniques including the creation of composite indicators and box-cox transformations. Unsupervised machine learning techniques, including dimensionality reduction techniques such as principal components analysis or K-means clustering, will be carried out as appropriate. Principal component analysis uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The first principal component has the largest possible variance and accounts for the highest proportion of the variance in the data, with each succeeding component accounting for the highest variance possible after accounting for the previous components. K-means clustering is a way to use data to uncover natural groupings within a heterogeneous population (Table 1). To uncover patterns, the algorithm starts by first assigning data points into random groups. The group centers are then calculated, and the group memberships are re-assigned based on the distances between each data point and the group centers. This process is repeated until there are no changes in the group memberships from the previous iteration [16]. In its application to Mobile Academy, K-means clustering will be used to detect patterns in ASHA engagement with training content, including training initiation and completion. Among Kilkari users, K-means clustering will be used to assess patterns in exposure to content by user characteristics based on data elements available in the RCH, including parity, age, and geographic area.

Table 1. Sample of data elements by source for Kilkari & Mobile Academy.

Database	Description	Sample of data elements anticipated for use in analyses	Variables	Boxes with reference to Figure 1
MCTS and RCH databases	• National databases on reproductive, maternal newborn and child health care seeking among mothers and children <5 years of age • Physically located in the NIC data center in Delhi	• District, Mother & child unique IDs • Pregnancy IDs • ASHA IDs • Last menstrual period • Date of birth of child	• Geographic identifiers • Unique episode / beneficiary identifier	DB2, DB3
MOTECH database	• Program database containing data on women and children as well as registered ASHAs • Database works in conjunction with two algorithms, Kilkari and Mobile Academy, which function as "engines" for running the program • Physically located in Railtel data center, Gurugram, Haryana	Kilkari: • Activation date • Deactivation date • Calls answered • Duration of content heard per message Mobile Academy: • Course start date • Course completion date • Chapter wise scores • Notification retry count • Course passing date	Kilkari: • Duration of enrollment in the program • Status of enrollee Mobile Academy: • Time taken to finish course • Performance score of ASHA • Number of tries by ASHA to complete modules	DB4
IVR database	• Records when calls are triggered, what happens after the call is triggered (ie, does it get answered or does it fail, and if it fails why has it failed, eg, network errors, device errors) and whether the call needs to be retried; if yes, how many times. Records similar data for incoming calls. • Physically located in Railtel data center, Gurugram, Haryana	Kilkari: • Content file name • Message duration Mobile Academy: • Content identifier • Duration of listening	Kilkari: • Message identifier • Duration of listening to message Mobile Academy: • Content identifier • Duration of listening	DB6
Call records database	• Operator database on call handling • Physically located in the Reliance data center, New Delhi	• Call status • Call failure reasons • Call start time • Call stop time • Additional elements	• Duration of listening to content	DB7

Training of Algorithms

Once data have been processed, testing of algorithms will be carried out. Table 2 summarizes the algorithms proposed for training along with their intended applications to Mobile Academy and Kilkari. To determine the model with the best fit, we will explore several machine learning approaches in turn. Models will be fit on the training set, and the fitted model used to predict the responses for the observations in the validation set. The preferred analytic approaches will be selected based on their ability to minimize the total error of the classification, where the latter is defined as the probability that a solution will classify an object under the wrong category. We describe each approach considered below in lay terminology, along with indications for use, and its proposed application in the evaluations of Mobile Academy and Kilkari.

Table 2. Summary of algorithms proposed for testing and their intended application to Mobile Academy and Kilkari.

Algorithm	Description	Intended evaluation application	
		Mobile Academy	Kilkari
Supervised			
1 Logistic regression	Classification (nonlinear model)	Classification of ASHA workers by user characteristics and patterns of training initiation by ASHAs	User characteristics associated with exposure to messaging content and duration in MP where face-to-face survey data have been collected
2 Linear discriminant analysis	Classification (linear model). It is a linearization of Gaussian naïve Bayes.	Classification of ASHA workers by user characteristics and patterns of training initiation, completion, and performance	User characteristics associated with exposure to messaging content and duration
3 Support vector machines (SVMs)	SVMs are techniques based on the calculation of the maximum margin hyperplane for the classification problems	Classification of ASHA workers by user characteristics and patterns of training initiation and completion	User characteristics associated with exposure to messaging content and duration
4 Classification and regression trees	Predictive model that consists of leaves that represent the target and branches that represent conjunctions of inputs features. Considered a subset of decision trees. Random forests operate by constructing multiple decision trees during training and aggregating their results to avoid overfitting by single trees.	Classification of ASHA workers by user characteristics and patterns of training initiation and completion	User characteristics associated with exposure to messaging content and duration
5 Naïve Bayes	Classification model based on probabilities	Classification of ASHA workers by user characteristics and patterns of training initiation and completion	User characteristics associated with exposure to messaging content and duration
6 Neural Networks (NNs)	NNs are powerful models for machine learning. They are a generalization of linear and nonlinear models	Classification of ASHA workers by user characteristics and patterns of training initiation and completion	User characteristics associated with exposure to messaging content and duration
Unsupervised			
7 K-means	K-means clustering is a way to use data to uncover natural groupings within a heterogeneous population	Grouping of ASHA workers by user characteristics and patterns of training initiation and completion	User characteristics associated with exposure to messaging content and duration

Our choice of methods will include a mix of algorithms based on their strengths and weaknesses and the objective of the process. A comprehensive comparison of supervised learning methods is provided in literature [17,18]. SVM and NNs are expected to perform better with continuous data while the Naïve Bayes method and decision trees perform better with discrete/categorical variables. Naïve Bayes and decision trees have good tolerance to missing values, while NNs and SVM do not. NNs and Naïve Bayes have difficulty handling irrelevant and redundant attributes (ie, extra variables with no useful information or variables with too many categories and too few numbers), while SVM and decision trees are insensitive towards them. Variables with high correlation negatively affect the performance of both Naïve Bayes and NNs, whereas SVM are relatively robust to correlated variables. While Naïve Bayes is robust to noise, NNs are sensitive to poor measurement of variables and susceptible to overfitting. NNs and SVM perform well with multidimensional data and when there is a nonlinear relationship between predictor and outcome. Naïve Bayes requires less memory for both training and validation phase, whereas NN requires large memory allocation across all phases. SVM and NNs usually outperform other methods while Naïve Bayes may yield less accurate results. Table 3 compares the strengths and weakness of different supervised machine learning methods.

Testing and Validation

To facilitate decision making on the optimal analytic approach, three steps will be undertaken: (1) develop the correct model for each algorithm using the training dataset, (2) apply the final model for each algorithm on the test dataset, and (3) apply the best performing algorithm on the validation dataset.

In Step 1, algorithms will be run using the training dataset comprising 60% of the total sample from across all states for which data are available. For each algorithm, iterative testing will be run to select the best model that fits the data. The emerging results will then be assessed for model fit and accuracy. Table 4 summarizes the four proposed metrics for assessing the performance of each model.

XSL•FO

RenderX

Table 3. Performance comparisons of learning algorithms modified from Kotisiantis et al [17,18] (++++ represents the best and + the worst performance).

Model attributes	Decision trees	NNs	Naïve Bayes	SVM	Linear discriminant analysis	Logistic regression
Accuracy in general	++	+++	+	++++	+++	++
Speed of learning with respect to number of attributes and number of instances	+++	+	+++	+		
Speed of classification	++++	++++	++++	++++	++++	++
Tolerance to missing values	+++	+	++++	++	+++	+
Tolerance to irrelevant attribute	+++	+	++	++++	+++	+
Tolerance to redundant attributes	++	++	+	+++	++	+
Tolerance to highly interdependent attributes	++	+++	+	+++	++	+
Dealing with discrete/ binary/ continuous attributes	++++	+++ (not discrete)	+++ (not continuous)	+++ (not discrete)	++++	+++
Tolerance to noise	++	++	+++	++	+++	++
Dealing with danger of overfitting	++	+	+++	++	++	+
Attempts for incremental learning	++	+++	++++	++		
Explanation ability/ transparency of knowledge/ classification	++++	+	++++	+	+	+
Model parameter handing	+++	+	++++	+	++	+

Table 4. Metrics for assessing the performance of each model.

Model metrics	Formula[a]	Description
Accuracy	(TP+TN)/(TP+TN+FN)	Proportion of cases correctly classified
Precision or positive predictive value	TP/(TP+FP)	Fraction of relevant instances among the retrieved instances
Sensitivity or recall	TP/(TP+FN)	Fraction of relevant instances retrieved over total amount of relevant instances
Area under receiver operating characteristic curve (AUROC)	Area covered by the function of true positive rate/ false positive rate	The curve is plotted and area between the curve and the 45° line is multiplied by 2 to give the AUROC. A value of 1 represents perfect accuracy while 0.5 means the prediction is worthless.

[a]TP: true positive, TN: true negative, FP: false positive, FN: false negative

To illustrate the definition of performance metrics for Mobile Academy, we define true positives (TP) as the number of correctly classified ASHAs who have completed the training, and true negatives (TN) as the number of correctly classified ASHAs who have not completed the training. False positives (FP) are defined as the number of ASHAs incorrectly classified as having completed the training, while false negatives (FN) are the number of ASHAs incorrectly classified as not having completed the training.

Results from the performance metrics will help define the final model for each algorithm. In Step 2, these final models for each algorithm will be applied to the test dataset, which comprises approximately 20% of the total data. Using the same performance metrics, the models with the best fit and accuracy will be applied to the validation dataset as part of Step 3. Ultimately, predictions for Mobile Academy will aim to determine the probability of the ASHA finishing the course in a predetermined time frame and the possible score/performance of the individual ASHAs. For Kilkari, we will determine predictors for exposure to Kilkari content based on user

characteristics, as well as explore the effect of early listening patterns on postpartum engagement and overall exposure.

Results

The project has obtained the necessary approvals for the use of data in accordance with global standards for handling personal data. The results are expected to be published in August 2019.

Discussion

Principal Considerations

This paper presents the testing of a range of machine learning approaches to be incorporated as part of the evaluation of two large digital health programs in India—Mobile Academy and Kilkari. By utilizing machine learning approaches, we aim to improve the use of data for generating evidence on program reach and exposure as well as factors underpinning uptake for both programs. We will start by measuring dropped cases and missing data along the continuum of the databases. As part of this analysis, we aim to understand differentials in the

XSL•FO
RenderX

characteristics of individuals lost along the continuum of databases and in turn, missed by the Mobile Academy or Kilkari programs. We then consider the technological performance as captured by four stakeholders: (1) government data systems (MCTS, RCH, webservices), (2) program's call delivery and receipt systems, (3) mobile network operator (network coverage and quality), and (4) user device characteristics (switched on, within range of network). Finally, we explore user engagement with the program (behavioral performance) including patterns in the data that could predict key program performance metrics, including training completion for Mobile Academy and user coverage and exposure for Kilkari.

Analyses to measure dropped cases and missing data along the continuum of the databases, starting with the MCTS and RCH databases, are anticipated to generate insights into user relay of information to the Government of India including the provision of phone numbers by health workers and women. At present, there are no reliable estimates of the proportion of pregnant women with accurate mobile numbers, last menstrual periods, or dates of birth (for themselves or child) covered by the databases. More broadly, these analyses will help improve understanding of the quality of the RCH and MCTS data and underlying sampling frames used for Mobile Academy, Kilkari, and a range of other programs, including the characteristics of individuals included compared to those excluded. The predictors identified will help identify groups of women who are more likely to be missed by the program and ASHAs less likely to complete their course. These findings will also be used to inform the sampling of respondents for qualitative research to explore issues in depth and possibly identify opportunities for improving program targeting like registration of users.

Analyses to understand the technological performance of the program will build off of those conducted on the MOTECH platform in Ghana [19] and of MomConnect in South Africa [8]. In the former, IVR message delivery trends suggested that 25% or less of expected mobile health information messages were received by pregnant women [19]. While 20% delivery rates of successful outbound dialing calls are standard in the mobile industry, limitations in the timeliness of problem identification represent a missed opportunity for improving program exposure using, for example, call retry logics and systems. In South Africa, over 80% of short messaging service (SMS) messages were successfully delivered as part of MomConnect. SMS as a delivery channel has a much higher success rate than outbound dialing calls but suffers from other weaknesses in countries where illiteracy rates are high and local language fonts are not widely available on devices. In the case of MomConnect, however, challenges in the use of unstructured supplementary service data meant that 26% of initiated registrations did not convert into successful registrants [8]. These two examples reinforce the need to understand the user journey and follow the flow of data to understand whether the technology performs as intended.

Beyond understanding the technological performance of the program, and the wider telecommunications network and device landscape that it operates within, analyses will aim to measure user engagement, including predictors for exposure to content based on user characteristics. In the context of Kilkari, we will additionally plan to explore the effect of early listening patterns (eg, time of day of listening, duration and frequency of listening, content listening patterns) as a predictor for postpartum engagement and overall exposure. For Mobile Academy, we will use a mixture of unsupervised and supervised machine learning techniques to generate predictors of training course completion by ASHAs based on performance on early modules of the course. These will be externally validated using a range of data elements on other ASHA characteristics obtained from the broader evaluation, including reported motivation, knowledge, individual characteristics, and mobile literacy.

Overall, the proposed analyses are anticipated to complement primary data collection activities proposed as part of the summative evaluation of Kilkari and Mobile Academy. Findings emerging from this analysis will provide program implementers with tools for improving predictions of success and performance and provide insights into strategic use and collection of data. Elsewhere, deployments of similar programs including MomConnect in South Africa, Aponjan in Bangladesh, as well as other maternal messaging programs may be able to apply some of the same approaches described [8].

Limitations

Ethical issues related to identifiers are important considerations for analyses described here. We will de-identify the databases at the point of data download, and all data will be secured in a password-protected hard storage device with access controlled by the Principal Investigator of the study. Once data are accessed, we note that findings will be only as reliable as the quality of underlying data. Completeness of the coverage of data and the data elements can be limited in low-resource settings. In recognition of this challenge, we will assess the quality of data for completeness and timeliness. The inherent design of the messaging program means that only those with mobile phones can be part of the sample, which could introduce selection bias in the association between outcome and predictors. Confounding will be an issue due to the many unmeasured variables associated with exposure such as mobile phone ownership, and registration into the pregnancy tracking database, and outcomes such as active listening to messages and adoption of healthy behaviors. Call answering does not automatically mean listening to the message by the intended client (eg, ASHAs or pregnant women) and as such, our analyses will be limited in its ability to identify the listener (information bias). However, the potential benefits of having other family members listen to information content could be immense. Beyond challenges with the measurement of exposure, the absence of a complete set of predictors will be present due to the limitations of such large-scale data gathering processes (incomplete model).

Conclusions

This paper aims to provide a survey of approaches to the applications of machine learning to improve the implementation and evaluation of digital health programs. The two digital health examples described represent two of the largest digital health programs, based on number of active users, currently being implemented globally. Developing classifiers based on the above machine learning approaches will help identify gaps in data (Aim 1), target potential slow learners (Aim 2), measure

exposure to program (Aim 3), characterize exposure levels in the population (Aim 4), and evaluate program impact (Aim 5). The scale of implementation and associated generation of data on user engagement with program content provides opportunities for big data analytics and more specifically, the use of machine learning approaches to improve the generation of evidence on program reach and exposure as well as factors underpinning uptake for both Mobile Academy and Kilkari.

Authors' Contributions

AL conceived the overall evaluation study with input from DM and SC. DM and AL wrote the draft of the manuscript with assistance from JJHB. NT, SC, PD, and NM provided extensive feedback on the drafts of the manuscript.

Conflicts of Interest

SC is employed by BBC Media Action and involved in the implementation of both programs from which data in this study will be drawn.

References

1. Mooney S, Pejaver V. Big Data in Public Health: Terminology, Machine Learning, and Privacy. Annu Rev Public Health 2018 Dec 01;39:95-112 [FREE Full text] [doi: 10.1146/annurev-publhealth-040617-014208] [Medline: 29261408]
2. Clifton L, Clifton D, Pimentel M, Watkinson P, Tarassenko L. Predictive monitoring of mobile patients by combining clinical observations with data from wearable sensors. IEEE J Biomed Health Inform 2014 May;18(3):722-730. [doi: 10.1109/JBHI.2013.2293059] [Medline: 24808218]
3. Ellis K, Godbole S, Marshall S, Lanckriet G, Staudenmayer J, Kerr J. Identifying Active Travel Behaviors in Challenging Environments Using GPS, Accelerometers, and Machine Learning Algorithms. Front Public Health 2014;2:36 [FREE Full text] [doi: 10.3389/fpubh.2014.00036] [Medline: 24795875]
4. Kerr J, Patterson R, Ellis K, Godbole S, Johnson E, Lanckriet G, et al. Objective Assessment of Physical Activity: Classifiers for Public Health. Med Sci Sports Exerc 2016 May;48(5):951-957 [FREE Full text] [doi: 10.1249/MSS.0000000000000841] [Medline: 27089222]
5. Parkka J, Ermes M, Korpipaa P, Mantyjarvi J, Peltola J, Korhonen I. Activity Classification Using Realistic Data From Wearable Sensors. IEEE Trans. Inform. Technol. Biomed 2006 Jan;10(1):119-128. [doi: 10.1109/TITB.2005.856863]
6. Fahmi P, Viet V, Deok-Jai C. Semi-supervised fall detection algorithm using fall indicators in smartphone. 2012 Presented at: 6th International Conference on Ubiquitous Information Management and Communication; Feb. 20-22, 2012; Kuala Lumpur, Malaysia. [doi: 10.1145/2184751.2184890]
7. Rahman Q, Janmohamed T, Pirbaglou M, Ritvo P, Heffernan J, Clarke H, et al. Patterns of User Engagement With the Mobile App, Manage My Pain: Results of a Data Mining Investigation. JMIR Mhealth Uhealth 2017 Jul 12;5(7):e96. [doi: 10.2196/mhealth.7871] [Medline: 28701291]
8. LeFevre AE, Dane P, Copley CJ, Pienaar C, Parsons AN, Engelhard M, et al. Unpacking the performance of a mobile health information messaging program for mothers (MomConnect) in South Africa:. BMJ Glob Health 2018;3(Suppl 2):e000583 [FREE Full text] [doi: 10.1136/bmjgh-2017-000583] [Medline: 29713510]
9. Mackey T, Kalyanam J, Katsuki T, Lanckriet G. Twitter-Based Detection of Illegal Online Sale of Prescription Opioid. Am J Public Health 2017 Dec;107(12):1910-1915. [doi: 10.2105/AJPH.2017.303994] [Medline: 29048960]
10. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. 2013 Presented at: Int AAAI Conf Weblogs Soc Media (ICWSM); July 8-11, 2013; Boston, MA p. 128.
11. Yang W, Mu L. GIS analysis of depression among Twitter users. Applied Geography 2015 Jun;60:217-223. [doi: 10.1016/j.apgeog.2014.10.016]
12. Du J, Xu J, Song H, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):69 [FREE Full text] [doi: 10.1186/s12911-017-0469-6] [Medline: 28699569]
13. Wang J, McMichael A, Meng B, Becker N, Han W, Glass K, et al. Spatial dynamics of an epidemic of severe acute respiratory syndrome in an urban area. Bull World Health Organ 2006 Dec;84(12):965-968 [FREE Full text] [Medline: 17242832]

XSL•FO
RenderX

14. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]

15. LeFevre A, Agarwal S, Zeller K, Vasudevan L, Healey K, Tamrat T, et al. M&E for digital health: A pragmatic workbook to assist with monitoring and evaluation of digital health interventions. In. Geneva: World Health Organization; 2016:A.

16. Armstrong J, Zhu M, Hirdes J, Stolee P. K-means cluster analysis of rehabilitation service users in the Home Health Care System of Ontario: examining the heterogeneity of a complex geriatric population. Arch Phys Med Rehabil 2012 Dec;93(12):2198-2205. [doi: 10.1016/j.apmr.2012.05.026] [Medline: 22705468]

17. Kotsiantis S, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques. In: Emerging artificial intelligence applications in computer engineering. 2007 Presented at: Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies; 2007; Athens p. 160.

18. Kotsiantis S, Zaharakis I, Pintelas P. Machine learning: a review of classification and combining techniques. Artif Intell Rev 2007 Nov 10;26(3):159-190. [doi: 10.1007/s10462-007-9052-3]

19. LeFevre AE, Mohan D, Hutchful D, Jennings L, Mehl G, Labrique A, et al. Mobile Technology for Community Health in Ghana: what happens when technical functionality threatens the effectiveness of digital health programs? BMC Med Inform Decis Mak 2017 Dec 14;17(1):27 [FREE Full text] [doi: 10.1186/s12911-017-0421-9] [Medline: 28292288]

Abbreviations

ASHAs: accredited social health activists
IVR: Interactive Voice Response
MA: Mobile Academy
MCTS: Maternal and Child Tracking System
mHealth: mobile health
MP: Madhya Pradesh
NN: neural networks
RCH: reproductive child health
SVM: support vector machines

XSL•FO
RenderX

Protocol

Impact of a Search Engine on Clinical Decisions Under Time and System Effectiveness Constraints: Research Protocol

Anton van der Vegt[1], BEng, BSc; Guido Zuccon[1], PhD; Bevan Koopman[2], PhD; Anthony Deacon[3], MD

[1]School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, Australia

[2]Australian eHealth Research Centre, The Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia

[3]School of Medicine, University of Queensland, St Lucia, Australia

Corresponding Author:
Anton van der Vegt, BEng, BSc
School of Information Technology and Electrical Engineering
The University of Queensland
L4, General Purpose South (78)
St Lucia,
Australia
Phone: 61 733651111
Email: a.vandervegt@uq.net.au

Abstract

Background: Many clinical questions arise during patient encounters that clinicians are unable to answer. An evidence-based medicine approach expects that clinicians will seek and apply the best available evidence to answer clinical questions. One commonly used source of such evidence is scientific literature, such as that available through MEDLINE and PubMed. Clinicians report that 2 key reasons why they do not use search systems to answer questions is that it takes too much time and that they do not expect to find a definitive answer. So, the question remains about how effectively scientific literature search systems support time-pressured clinicians in making better clinical decisions. The results of this study are important because they can help clinicians and health care organizations to better assess their needs with respect to clinical decision support (CDS) systems and evidence sources. The results and data captured will contribute a significant data collection to inform the design of future CDS systems to better meet the needs of time-pressured, practicing clinicians.

Objective: The purpose of this study is to understand the impact of using a scientific medical literature search system on clinical decision making. Furthermore, to understand the impact of realistic time pressures on clinicians, we vary the search time available to find clinical answers. Finally, we assess the impact of improvements in search system effectiveness on the same clinical decisions.

Methods: In this study, 96 practicing clinicians and final year medical students are presented with 16 clinical questions which they must answer without access to any external resource. The same questions are then represented to the clinicians; however, in this part of the study, the clinicians can use a scientific literature search engine to find evidence to support their answers. The time pressures of practicing clinicians are simulated by limiting answer time to one of 3, 6, or 9 min per question. The correct answer rate is reported both before and after search to assess the impact of the search system and the time constraint. In addition, 2 search systems that use the same user interface, but which vary widely in their search effectiveness, are employed so that the impact of changes in search system effectiveness on clinical decision making can also be assessed.

Results: Recruiting began for the study in June 2018. As of the April 4, 2019, there were 69 participants enrolled. The study is expected to close by May 30, 2019, with results to be published in July.

Conclusions: All data collected in this study will be made available at the University of Queensland's UQ eSpace public data repository.

International Registered Report Identifier (IRRID): DERR1-10.2196/12803

XSL•FO

RenderX

Introduction

Clinicians are routinely faced with medical questions related to their patient interactions [1]. Studies conducted with primary care physicians show that on average between 0.07 and 1.85 questions are generated per patient encounter [2], or a little under 1 question per hour [3]. Of these questions, many are often left unanswered, as demonstrated by 3 studies in the United States [3-5] where 63.76% (702/1101), 44.91% (477/1062), and 70.2% (207/295) of the medical questions raised by the clinicians were left unanswered. Clinicians are expected to seek and apply the best evidence to answer their clinical questions, according to an evidence-based-medicine approach to clinical decision making [6,7]. Search engines provide a means for clinicians to access scientific literature while on the job. However, physicians suggest that lack of time and the belief that the system will not provide a definitive answer are 2 of the primary barriers to pursuing an answer [4,5]. So, the question remains: how effective are scientific literature search engines at supporting clinicians in making better clinical decisions. This study aims to address this question.

Study Aims

The overall aim of this study is to examine the suitability of using a search engine to search scientific literature to enable time-pressured clinicians to make better clinical decisions. To support this assessment, the following 3 research questions (RQs) will be addressed:

> *RQ1: Does the use of a Web-based scientific literature search system enable clinicians to make better clinical decisions?*
>
> *RQ2: How does time pressure impact clinical decision quality?*
>
> *RQ3: Does a significantly better search system, as measured by standard information retrieval (IR) evaluation measures, translate to better and faster clinical decisions?*

Significance of This Study

This study will inform both health care providers, with regard to system selection to suit their use case, and system designers, with regard to evidence selection and search effectiveness requirements. It will also contribute a rich data collection for future research purposes, including specifically:

- Clinician query sets to analyze the impact of query quality on clinical decision making.
- Clinician evidence sets (ie, actual text selected from the literature by clinicians to provide evidence for their clinical answers) to analyze their relationships with clinical questions, clinician queries, search retrieval snippet cues, clicked documents, and answer quality.
- Clinician document relevance ratings to analyze the relationship with snippet design, clicked documents, judged relevance, selected evidence, and answer quality.
- Clinician search engine results page (SERP) interaction, including read time and clicks to help identify patterns of search behavior and how this relates to clinical decision quality.

- Search and answer time breakdown to identify where time is spent during search for evidence.

Sources of Evidence

IR systems can use one or more of many different sources of evidence to help clinicians answer their clinical questions, including scientific literature, best practice information, guidelines, or synthesized information, such as that generated by UpToDate (Wolters Kluwer) [8]. Haynes identified the *5S* levels of organization of evidence from health care research [9], which depicts a pyramid of health care evidence with journal studies at the base followed by syntheses, synopses, summaries, and finally systems, such as computerized decision support systems, at the top. Of interest to this research is the use of scientific medical literature (SML), such as that found in the MEDLINE (US National Library of Medicine) [10] database or accessed via PubMed (US National Library of Medicine). It includes original research and meta studies, such as systematic reviews, and is represented by the bottom 2 layers of Haynes' pyramid of evidence. SML is widely used across the medical research community and the public [11], but it is also a common source of evidence used by clinicians [12,13] to support their clinical queries.

Physician Preference Versus Suitability?

Although SML is used widely by clinicians, larger studies conducted across medical institutions suggest that it is not a preferred source of evidence for busy clinicians. In particular, Ellsworth et al [14] found in a survey of 450 clinicians across the Mayo Clinic that 56.8% (255/450) of respondents preferred synthesized information sources versus 12.9% (58/450) who preferred original research. Hoogendam et al [15] studied the clinical evidence preferences of 70 clinicians in a Dutch academic medical center over the course of 18 months. Their study found that while answering 1305 patient-related questions, clinicians chose to use UpToDate 78.49% (883/1125) of the time rather than PubMed. Hoogendam et al asserted that the time required to find an answer was the most likely explanation for this bias, noting that clinicians spent, on average, less than 5 min pursuing a question.

Clinician preference for synthesized evidence, rather than SML, is at best an implicit indicator of the suitability of SML search systems for their clinical needs. However, the time clinicians have available for answering their questions, and therefore the time needed to search for a definitive answer, is likely to be an important facet of SML system suitability to be incorporated within our study.

Previous Search System Studies

It is difficult to find conclusive evidence supporting SML as the sole source of evidence for clinicians under strict time constraints. Dunn et al [12] analyzed surveys from 14,544 clinicians examining the impact of evidence search on patient care. They found that 75.33% (10,956/14,544) of respondents used more than one evidence source and that journals (print and Web) and MEDLINE were the top 2 sources used. They concluded that these sources are an effective component in providing clinical answers; however, the use of UpToDate and other evidence sources made it difficult to evaluate MEDLINE

in a stand-alone context. McKibbon et al [13] assessed the successful answering of 46 clinical questions across 23 clinicians. The physicians took on average 13 min to answer each question and could reference multiple data sources of their choice, including PubMed and MEDLINE; however, their correct answer rate only improved from 18 (39% [18/46]) questions correct presearch to 19 (41% [19/46]) correct after using a search system. Westbrook et al [16], on the contrary, found more extensive improvement with the use of a search system. They studied the answer accuracy of 8 clinical questions presented to 75 clinicians, including nurses and doctors. The study was designed to simulate realistic time pressures, so participants were given 10 min per question; however, this limit was not enforced. Although participants completed the 8 questions within 80 min, it was unclear whether some questions took longer than 10 min to complete. The study showed that the introduction of a clinical evidence search system improved the correct answer rate from 174 (29.0% [124/600]) correct questions without the system to 298 (49.7% [298/600]) correct with the system. The search system comprised 6 sources of evidence, PubMed included.

These studies show that an evidence search system can be effective to help clinicians make better clinical decisions and that SML may be a helpful component of a broader range of evidence sources; however, they do not confirm whether an SML search system is suitable as a stand-alone system for the same task. Studies conducted where SML was the sole source of evidence include the ones by Hersh et al [17,18]. In their first study, 19 medical students and 8 nursing students answered 3 medical questions each [17]. The correct answer rate improved from 39 (45% [39/87]) correct answers to 66 (76% [66/87]) after searching MEDLINE alone. This is a much higher increase than found in the study by Westbrook et al; perhaps attributable to the questions asked, some of which were examination style, and the 1-hour timeframe to complete the questions. In the second study, 45 medical and 21 nurse practitioner students answered a total of 324 questions [18]. The use of MEDLINE-only search improved correctness from 104 (32.1% [104/324]) correct to 150 (46.3% [150/324]) overall; however, the nursing students showed a small improvement of just 3 percentage points. These studies [17,18] do focus on SML alone; however, the longer allowable answer timeframes and the conflicting results motivate the authors of this study to more tightly control the user study, similar to Westbrook et al, but with enforced time limits and a single evidence source.

Time Constraints and Time Pressure

According to Ordonez and Benson [19], time constraints exist whenever there is a deadline for a task; however, for the task performer to be time pressured, the time constraint must induce stress such that they feel the need to cope with the limited time. In our study, time pressure will be induced by specifying to the participant, and enforcing, a time limit for searching the SML for an answer.

In the field of psychology, experiments have revealed many coping mechanisms that impact the task performer's decisions [20-23]. Many of these coping mechanisms are relevant to clinical decision making, for example, Wright [23] found that

under significant time pressure, subjects changed their decision-making strategy, used fewer information attributes to make their decision and were more reliant on negative attributes, that is, those that had negative consequences. In Edland and Svenson's review of the literature of time-pressured decision making [21], they noted that time pressure can lead to a shallower search for information across alternatives. Svenson and Benson found that task performers will also change their decision strategy when put under time pressure [20].

Some of these behaviors have been explored in the IR field. Chang and Wei explored the impact of time constraints on users' search strategy [24] and found significant differences between users with or without a time constraint: users under time constraints tended to view less documents and spend more time on the search engine results page. Crescenzi et al [25,26] confirmed that searchers under time-constrained conditions reported significantly greater time pressure, felt that the tasks were more difficult, and felt less satisfied with their performance. This outcome prompts the question of whether or not this lower satisfaction in performance correlates to poorer decisions. The influence of time pressure within the clinical setting has been studied by Tsiga et al [27]. In their study of 34 general practitioners, practicing within a town in Greece. They found that under time pressure, clinicians asked less questions regarding symptoms and conducted less thorough physical examinations for a given clinical scenario. This study will examine the impact of time pressure on clinical decisions. Time pressure is a major barrier to using an evidence search system [4,5], and the time-consuming nature of using an SML system [28], such as PubMed, may suggest it is inappropriate under certain time constraints. By varying the time available to search for evidence, this study will explore the relationship between the time a clinician has available to search for answers and the quality of their clinical decisions.

Search System Effectiveness

A less obvious factor that may also impact the suitability of SML search systems for time-pressured clinicians is the effectiveness of the search engine. Intuitively, a more effective system that provides more relevant literature for the clinician's question is more likely to speed up the answer process and, therefore, present SML as a more suitable evidence source. Studies conducted outside of health have shown that search system effectiveness can impact user search behavior, performance, and satisfaction [29-32]. In particular, Allan et al [31] varied the system effectiveness, as measured by binary preference, and captured the time it took participants to find answer facets to specific questions. They found that for specific bands of improved system effectiveness, user performance also improved, including reduced time on task, less errors, and an increased rate of finding new, correct answers. This is in contrast with the study by Turpin et al [29] who found no significant relationship between system effectiveness, as measured by mean average precision (MAP) and user performance for a simple precision-based task and only a weak relationship for a simple recall-based task.

System effectiveness was implicitly excluded in the health domain studies above by using the same search system

throughout each study [17,18,33]. To our knowledge, our study will be the first to research the impact of search system effectiveness on clinician decision making.

In summary, the aim of this study is to examine the suitability of using a search engine to search scientific literature to enable time-pressured clinicians to make better clinical decisions. The impact of both time pressure and search system quality on clinical decision making will be assessed.

Methods

Study Design

A total of 96 participants consisting of practicing clinicians and final year medical students are provided with 16 clinical scenarios, each with a single question. Figure 1 depicts the study steps. The participants must firstly answer the questions without any supporting evidence. In the second stage of the study, the same set of clinicians are provided with the same 16 clinical scenarios and an SML search system. A bespoke best-match

SML search system, called Taskiir, was used to avoid any experience variation from using the well-known PubMed interface, as described by Yoo and Mosa [34]. The participants will be constrained to one of 3, 6, or 9 min to search for suitable evidence and complete the task. The time allocated to each user for each task will change depending on the timing cohort they are assigned to (see Task Order and System Rotation section for details). In total, 2 SML search systems with the same user interface, but with significantly different search performance, will be provided to the participants for alternating questions. In this way, the presearch and postsearch correct answer rate by participant and by system will be captured.

To enable comparison with previous studies, much of the method employed by Westbrook et al [33] is replicated, including the use of 6 of the 8 clinical questions used in that study. The main differences with the Westbrook et al study are (1) the varied and strict time limits set to search and answer each question; (2) the use of medical literature only for evidence, rather than the 6 sources they used; and (3) the use of 2 search systems with different search performance.

Figure 1. Process flow diagram of study shows both stages of the study. Stage 1 is untimed and the clinician has no access to any support resource. In stage 2, each question is timed and the participant is allocated a search engine to use for each question. If time runs out, the participant is brought directly to the task completion page.
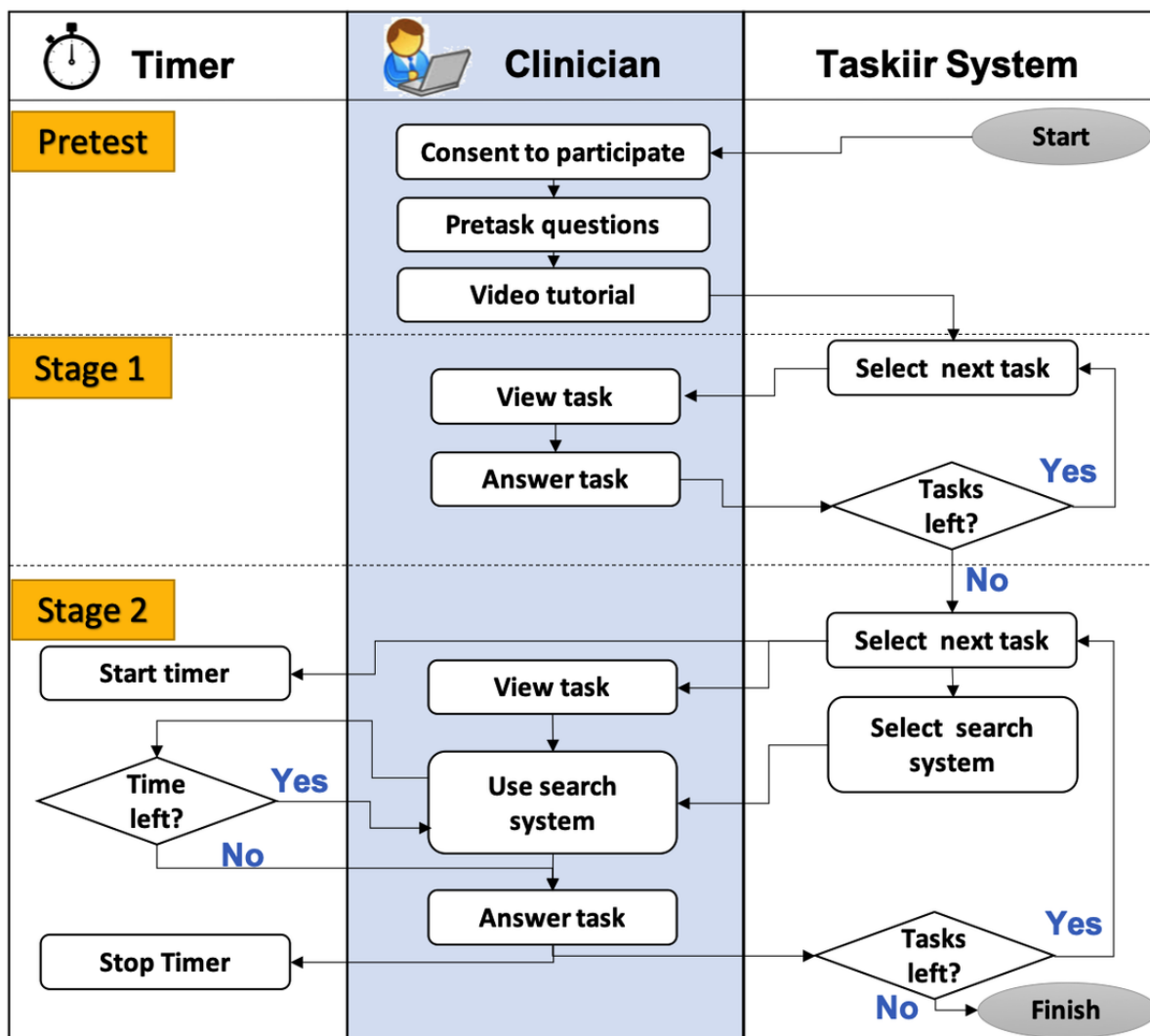
Participants

A convenience sample of 96 practicing clinicians and final year medical students, including nurses, general practitioners, and hospital physicians, will be asked to participate. The practicing clinical participants must be Australian registered clinicians residing in Australia. All participants must have access to a computer with an internet connection. Participants will be offered a small honorarium (Aus $50 gift card) to complete the assessment and will be recruited via mail, email, and Web-based noticeboards directed to medical student societies, clinical departments in hospitals, public health area networks, and medical faculties at Australian universities.

Procedures

Participants will be asked to complete a 2-hour, Web-based assessment of a medical SML search system called Taskiir. After voluntary consent is received, the participants are allocated their login details via email. In the email, the participant is advised that they can perform the study in multiple sittings, within a 2-week period, at a time to suit them and that they must use their laptop/computer (not iPad) to access the study on the Web. They were also encouraged to ask for help, via email, if they had any queries or problems. After testing the system with clinicians, we found that trying to complete all 16 questions in a single sitting was too onerous for some people, either because they did not have 2-hour time blocks available or they found the workload too mentally fatiguing. The system was reconfigured so that after completing any task, the participant could stop and resume again at the next task. All such pauses were recorded by the system.

After Initial Login

The participant is asked 7 questions to capture demographic data, search, and medical experience (see Multimedia Appendix 1) as well as sleep information. A 5- to 10-min video tutorial follows where the study is described in more detail and the participant is shown how to use the SML search engine. The tutorial emphasizes that the participant must answer the question without the aid of other people or by looking at other resources. Once complete, the participant is shown specific instructions (provided in Multimedia Appendix 2 that again reinforce the participant's obligation to perform the test alone, before they are permitted to move onto the 2-stage assessment.

In Stage 1

A total of 16 clinical tasks are presented to the participant, one at a time. To complete each task, the clinician must answer a single question within a few minutes, although this time limit is not enforced. In addition, 14 of the 16 tasks require the participant to select 1 of 4 answers (yes, no, conflicting evidence, and do not know) and the other tasks require a 1- to 2-word answer. At the end of the last task, the system will move the participant to stage 2 of the study.

In Stage 2

The participant must complete the same 16 tasks in the same order as stage 1; however, the participant must now use Taskiir to help them answer the question and to find evidence to support their answer. Evidence is collected by the participant selecting

text, images, or both from the source documents they read. The time allocated to search for each task is set according to the timing cohort the participant belongs to and will be one of 3, 6, or 9 min. The participant is told of the time allocation at the start of each question and a minute-by-minute countdown timer is always visible to the participant; warnings are given 30 seconds before time-out. At time-out, the screen is blocked, and the participant is taken to the task completion screen to enter their final details. Other methods of communicating the time limit were trialed during development of the system. In the end, the above method was chosen because it provided a balance between (1) making the participant aware of the time allocated for each question, (2) avoiding time anchoring (where the participant incorrectly assumed all questions are allocated the same time as the first question), (3) keeping them updated with the time remaining so they do not run out of time without warning, and (4) not distracting the participant with time information (eg, using a second-by-second countdown timer that diverted too much attention away from the task).

As this is the first such study measuring the impact of time variation, a few time limits covering a wide range are required to generate significant differences in the outcomes. A useful starting point to establish these time limits is the average completion time of 6.1 min per question, reported in the Westbrook study [33]. From the same study, the SD, based on the average completion times for each of the 8 questions, across 2 systems, is 3.1 min. Therefore, time limits are set at the average question answer time (6 min) and approximately 1 SD either side of this (3 and 9 min). These limits should induce time pressures for 84% of questions with a 3-min time limit, 50% of questions with a 6-min time limit, and 16% of questions with a 9-min time limit. From previous studies, realistic answer timeframes for busy clinicians should be below 5 min [3,15,35], so the 3 proposed time limit cohorts will encompass this pragmatic indicator of search time suitability.

The timer is stopped during the system search for documents to eliminate the system search time variation or other network/system delays that may bias the overall search time available. System search time starts when the participant clicks the search button and ends when the screen is populated with the search results and is available for use. The question timer will be stopped for each search conducted, including a *Move Next* or *Move Previous* on the search screen. Participants will be told that search time is excluded from the timing to alleviate any additional time stress they may feel because of a perceived or actual slow system.

A control group of participants that could use the information system without time constraints was considered; However, it was decided that numerous similar previous studies, such as that of Westbrook [16], had already generated results that could be compared with the outcomes of this study. Allocating test participants to a control group without time constraints would reduce the statistical power of any test results achieved here and expanding the participant set was not feasible for time and cost reasons.

Data Capture

Immediately after initial login, participant information is captured as per the table in the Multimedia Appendix 1. Data capture then occurs on both the presearch and postsearch answer screens, as listed in Table 1. All system interactions will also be captured including (1) overall time spent searching for and answering each question; (2) dwell times before first query, on the SERP screen, on the document viewing screen, and on the answer screen; (3) the participant's search query terms and resulting SERP; (4) documents selected from the SERP; (5) evidence selected by the participant from the documents they are viewing; and (6) relevance ratings by the participant of the documents they view (essential, helpful, duplicate-essential, duplicate-helpful, and not helpful). Multimedia Appendix 3 itemizes the search interaction times and how these relate to the study timings. Although desirable, it is not possible to question the participant regarding the utility of each search system because the user is not made aware of which search system is in use for each task.

Anonymity of the data collected is maintained by (1) identifying users and all of their interactions with a random user identification within the data capture system, (2) having no participant identification information stored in the same system database, and (3) capturing only generic participant information (see Multimedia Appendix 1) that could not be used to identify an individual.

Availability of Data and Material

The datasets generated and/or analyzed during this study as well as access to the software for the Taskiir search system are currently not publicly available because the study is still underway and therefore not complete.

At the completion of the study, the following research data will be made available on the University of Queensland's publicly accessible eSpace data repository [36]:

1. Excel spreadsheet containing all data reported by the user (as specified in the *Data Capture* section above) by task including answers to all study questions and task responses. Overall task timings will also be provided here.
2. mySQL database (anonymized) containing the raw data captured, including detailed user-task interaction timings, search terms, SERP results, SERP clicks, document selections, document relevance selections, and evidence selection text.
3. Auxiliary Excel/text files containing summarized subsets of (2), as required for further research and analysis.

Clinical Tasks

The criteria for task selection was that each task must (1) have answers able to be found in the literature, (2) be able to be answered with yes/no/conflicting information or a single-term/phrase response, (3) be credible to a practicing clinician, and (4) have nonobvious answers. Overall, 6 of the 16 clinical questions are those produced and used by Westbrook et al [16] and are reproduced here in Table 2. The tasks consist of real-life scenarios and a clinical question for each scenario. Westbrook et al derived the tasks using clinical experts and designed them to be clinically relevant and of mixed complexity. In addition, 4 questions are sourced from Hersh et al [18], which are also clinical questions and used for the same purposes as this study. Overall, 3 questions are modified from the text retrieval conference (TREC) 2015, clinical decision support (CDS) topic set [37]. These questions were provided with diagnoses, which our medical physician (DA, MBBS), modified into a question of a similar format to the other questions. Finally, our medical physician also devised a further 3 other clinical questions for the purposes of this test. Moreover, 2 general practitioners trialed all questions for suitability.

Table 1. Data capture on the presearch and postsearch answer screens.

No	Data: purpose	Measurement	Presearch	Postsearch
1	Answer: Decision quality	Select (Yes/no/conflicting evidence/do not know) or type answer depending on the question	Yes	Yes
2	Confidence in answer: impact of the system on answer confidence	How confident are you in your answer? (1=no confidence, 2=a little confident, 3=moderately confident, 4=very confident, and 5=certain)	Yes	Yes
3	Perceived difficulty: relationship with time constraints and answer quality	How would you rate the difficulty of this clinical question? (presearch) and How would you rate the difficulty of the search for evidence for this task? (postsearch). (1=very easy, 2=easy, 3=neither easy nor difficult, 4=difficult, and 5=very difficult)	Yes	Yes
4	Perceived impact of time constraint on decision: relationship with decision quality and confidence	How would you rate the time you had available to make your decision? (1=not nearly enough time, 2=nearly enough time, 3=just enough time, 4=more than enough time, and 5=much more than enough time)	N/A	Yes
5	Perceived impact of time constraint on decision: relationship with decision quality and confidence	How would you rate the time you had available to collect evidence? (1=not nearly enough time, 2=nearly enough time, 3=just enough time, 4=more than enough time, and 5=much more than enough time)	N/A	Yes
6	Perceived impact of time constraint on participant's stress level: relationship with decision quality and confidence	How much stress did you feel due to time pressure? (1=none, 2=a little, 3=a moderate amount, 4=a lot, and 5=more than a lot)	N/A	Yes

XSL•FO
RenderX

Table 2. Task specifications including the full task scenario supplied to the participant as well as the relevant reference from the corpus that supports the answer.

Question	Source
Cytobrush Pap Smear: Is the Cytobrush superior to a spatula for obtaining cells for Pap smears, in terms of technical quality (eg, percentage of interpretable smears)? [18]	[38,39][a]
Glue Ear: A mother brings her 15-month-old son who has been seen three times in the past year for glue ear. She has heard that this can lead to learning and developmental problems and thinks her child may need surgery. His hearing is normal. Does current evidence support the need for the insertion of tympanostomy tubes to avoid developmental problems in this child? [16]	[40]
Asthma Inhaler: What is the best delivery device for effective administration of inhaled medication to a 5-year-old child during a moderately severe acute asthma attack? [16]	[41]
Nicotine Replacement Therapy after heart attack: A patient staying in hospital had a myocardial infarction two days ago and is now threatening to sign himself out. You suspect this is due to nicotine withdrawal. The patient wishes to stop smoking and seeks your advice on whether he can start nicotine replacement therapy. Is nicotine replacement therapy appropriate for this patient? [16]	[42,43]
Glucosamine sulfate: A 58-year-old woman with long-standing pain of osteoarthritis in knees, hips, and hands asks about the benefits of glucosamine sulfate. Does existing evidence demonstrate that glucosamine has a disease modifying role in osteoarthritis? [16]	[44]
Brown snake: A man is bitten by a brown snake and is taken to the hospital emergency department. There is clear evidence of envenoming (poisonous effects of venom). The hospital has run out of brown snake antivenom, so the patient must be given polyvalent snake antivenom (which contains antivenom for all Australian snakes). Should epinephrine be given with the antivenom to prevent anaphylaxis? [16]	[45,46]
Osteomyelitis diabetic foot: What anaerobic microorganism is most commonly found in osteomyelitis associated with diabetic foot? [16]	[47][a]
Ultrasound for Deep Vein Thrombosis (DVT): Is ultrasound the best diagnostic test available to exclude the presence of lower extremity deep vein thrombosis? [18]	[48,49][a]
Protein-losing nephropathy: Does dietary protein effect the level of proteinuria in patients with diabetic (a type of protein-losing) nephropathy? [18]	[50,51][a]
Bladder Cancer: Is there evidence of an association between petroleum product exposure and bladder cancer? [18]	[52][a]
Loin pain: A 48-year-old man presents with severe right sided loin pain and is diagnosed with a 4 mm distal ureteric calculus. Has Tamsulosin been shown to increase the chances of the calculus passing?[b]	[53,54][a]
Breast cancer: Is oestrogen receptor positivity a better prognostic factor than human epidermal growth factor receptor 2 (HER2) overexpression for patients with breast cancer?[b]	[55][a]
Dementia: Are the clinical effects of Mematine, when used as a sole agent in the treatment of Alzheimer's Dementia, greatest in the "mild" stage of the disease?[b]	[56][a]
Paroxysmal nocturnal hemoglobinuria: Is flow cytometry the most accepted laboratory investigation to confirm a suspected diagnosis of Paroxysmal Nocturnal Hemoglobinuria? [AD modified TREC CDS 2015 [37], Q14].	[57,58][a]
Anaemia: Is the efficacy and side effect profile of oral iron polymaltose and oral ferrous sulfate equivalent when used for the treatment of iron deficiency anaemia among children? [AD modified TREC CDS 2015 [37], Q27].	[59][a]

[a]Answer provided by author, Dr AD (MBBS).

[b]Question derived by author, Dr AD (MBBS).

In Westbrook et al's study, 6 sources of evidence were available to search by the clinicians; however, only medical literature was provided in this study, as this was the source of evidence under investigation. To ensure that at least 1 relevant document existed in the corpus for each task, our medical physician searched through the corpus, using the search system, to identify 1 or more relevant documents. The resulting relevant PubMed sources are listed for each question in Table 2. The answers are excluded in this protocol to avoid any chance of participants viewing the answers before completing the study. However, they will be provided together with the results data.

Corpus

The clinical information corpus used is the TREC 2014 and 2015 document collection [37,60]. This consists of a snapshot of the Open Access Subset of PubMed Central taken on January 21, 2014. It contains a total of 733,138 articles. The corpus was preprocessed according to the method employed by [61], including the removal of all HTML/XML tags, all numbers and all nonalphabetical characters. The corpus was then indexed with Galago (the Lemur Project) [62] version 3.12 using a Porter stemmer and stop words removal. After indexing, all very rare terms were also removed, that is, all terms with 3 or less occurrences in the corpus.

Custom Search System

A custom search engine and interface, together called Taskiir, is employed for the evidence search process (see Figure 2). Similar to normal commercial search engines, Taskiir allows the participant to write their query and perform a best match

search of documents in the corpus. A snippet, highlighting matching query terms, is then provided in the SERP, which shows up below the query. Users can then select documents of interest to view the full text. While viewing the full text document, the participant can also select (with their mouse) any text or graphics that they want to use as evidence for their final answer. The participant can view their evidence or complete the task at any time. Instructions on using the system are provided on each page, and a mandatory walk-through tutorial is provided before starting the study.

To investigate the impact of search system effectiveness on clinical decision making (RQ2), Taskiir utilizes 2 search algorithms: (1) A state-of-the-art system, which is an improved version of the TREC 2015 CDS Task A winning system [61]. The TREC 2015 CDS track was targeted to identify the state-of-the-art IR system because the topics in Task A were of a similar clinical nature to the Westbrook tasks and the search corpus was the same as that used in this study. The 2 improvements made over the winning system include the removal of negated Unified Medical Language System (UMLS) terms from the UMLS query expansion terms as well as a change to the pseudorelevance feedback term weighting (from 0.75 to 0.5). All improvements resulted from tuning parameters on the CDS 2014 test collection and testing on the 2015 collection to avoid data overfitting. (2) A baseline document retrieval system consisting of a BM25 algorithm, which is a widely adopted best-match retrieval method. It is the default, out-of-the-box method employed by many search engines, including the very popular Elasticsearch (Elasticsearch BV) [63] and Lucene (Apache) systems [64]. The parameters were set to default values ($K=1.2$; $b=0.75$).

Information Retrieval Evaluation Measures

Document retrieval performance figures for both systems are shown in Table 3. The measures depicted were the standard set chosen for the TREC 2014 and 2015 CDS task. IR system performance measures are usually calculated for a ranked retrieval of 1000 documents MAP, for example, is the average of all precision values taken at each rank where a relevant document is found. Precision at a given rank is the number of relevant documents found up to that rank divided by the rank. MAP is useful because it provides a single measurement of system performance across all queries. However, because MAP is only averaged across relevant rank positions, results can be biased toward a system retrieving fewer relevant documents but at lower rank positions. Precision at rank position 10 is simply the precision calculated at rank position 10. It is useful to identify high-precision systems that provide many relevant

documents in the first 10 documents retrieved. This is often pertinent to a clinical search where clinicians have little time to view many documents. R-precision (R-prec) is the ratio r/R where r is the number of relevant documents retrieved by the system up to ranking R and R is the number of judged relevant documents for that query. Unlike MAP, R-prec takes into consideration the number of relevant documents that could be found and, therefore, is helpful for search tasks where recall is important. R-prec is a useful measure for systems that need to return many or all of the relevant documents, for example, in clinical cases that require physicians to seek alternatives, say for treatments. Discounted cumulative gain (DCG) sums the gain at each rank position (ie, the relevance grading value) multiplied by a discount factor that takes into consideration that lower ranked documents are less likely to be read. Normalized discounted cumulative gain (nDCG) compares the DCG with an ideal DCG for each rank, so that scores are normalized between 0 and 1. nDCG is designed to promote systems that provide more relevant documents higher up in the ranking.

One problem with all these standard measures is the underlying assumption that all relevant documents within the test collection are identified for each query. This is rarely the case because of cost limitations. In the measures above, unjudged documents are considered as nonrelevant; however, this may not be the case. To account for unjudged documents, Aslam et al derived 2 new measures, inferred nDCG and inferred average precision, which have become accepted methods of evaluating system retrieval performance when relevance judgements are incomplete [65].

Sample Size

The 2 largest and most similar studies [16,18], both commenced with a presearch correct answer rate of around 30% (29% and 32%) and a postsearch rate around 50% (50% and 46%). Using this as our basis, we wanted to be able to discriminate between the postsearch correct answer rate between each of the 3 time-constrained cohorts. Therefore, to derive the sample size, we estimated that the correct answer rates might vary evenly by 10 percentage points between each group, starting at no improvement. This creates 3 datasets with average correct answer rates of 30%, 40%, and 50% for the 3-, 6- and 9-min cohorts, respectively. Applying a 2-proportion statistical comparison (ie, a 2-sample, 2-sided equality) [66], between each pair of answer rates and setting statistical power to 90%, error rate to 5%, and equal sample sizes per cohort, the minimum sample size required is 514 per cohort, which equates to 32 people per cohort sitting 16 tasks or 96 people in total.

Figure 2. Screenshot of the Taskiir custom search system interface. Shows the task in the top left, search query box in the top right, and search results below.

Table 3. Comparison of document retrieval performance figures, across the text retrieval conference (TREC) 2015 test collection, for systems used in this study and the winning TREC 2015 system.

System	Inferred normalized discounted cumulative gain	Inferred average precision	Precision at rank position 10	R-precision	Mean average precision
Wayne State University (WSU) system[a]	0.2928	0.0777	0.4633	0.2329	0.1851
State-of-the-art system	0.3159	0.0849	0.4800	0.2401	0.1930
BM25 system	0.2168	0.0461	0.3600	0.1717	0.1114
State-of-the-art versus BM25 (%)	+46[b] (P=1.2E-05)	+84[b] (P=0.0056)	+33[b] (P=4.1E-04)	+40[b] (P=4.3E-05)	+73[b] (P=7.9E-05)

[a]As per TREC 2015 CDS, task A, automatic runs listed in [37] for task summary.

[b]Significance using paired t test

Task Order, System Rotation, and Task Timing

Task order and system rotation is set as per the table in Multimedia Appendix 4 for each participant to minimize confounding factors. The design is as follows:

1. A Latin square experimental design is constructed for 16 tasks and 16 participants to minimize the impact of user fatigue on specific tasks.
2. To minimize task order effects, each column of the square is randomized.
3. To incorporate a within-subject design across the system variable, 2 sets of the Latin square derived in (2) are required with alternating use of systems. The first tranche of 16 participants will start their first task with the state-of-the-art system, whereas the second tranche will start with BM25 system. In this way, across the 32 participants, each system will be used equally across all tasks and will experience the same task-order pattern.

The search time allowed for each task is controlled by applying a time limit for each task the participant performs. Participants are randomly assigned to 1 of 3 timing cohorts. The time constraint by task number is specified for each cohort in Multimedia Appendix 5. The rotation of task timing ensures that:

1. the maximum duration for search in stage 2 is fixed to 96 min for all participants

XSL•FO
RenderX

2. each task is conducted under all time constraints an equal number of times (32 per cohort)

3. a within-subject design across the time constraint variable such that each participant performs 4 to 6 tasks per time constraint

4. task time constraints are applied in the same random order according to the task order rotation Latin square, specified above

Statistical Analyses

To assess the impact of introducing the SML search system on clinical decision quality (RQ-1), each participant's answer, both presearch and postsearch, will be coded to right (R) or wrong by comparing the participant's answer with an expert judged assessment (gold answer) of each task. Samples for which (1) no evidence is captured and (2) no relevant documents are marked (either as essential or helpful), by the participant for their postsearch answer, will be discarded, as the value of the search system cannot be confirmed in these cases. Therefore, the decision quality is defined by the correct answer rate (number of right answers/total sample count (N)). A further detailed analysis will be performed of the collected evidence to identify tasks where the literature may contradict the gold answers. Where this occurs, the task answers will be reviewed by experts and overall correct rates adjusted.

To assess the significance of any change in the proportion of right or wrong answers, the McNemar test will be employed because it is a nonparametric test suited to a binary result, with samples taken at 2 points in time. Nonparametric is a better model to assume, given that the data distribution is unlikely to be regular because of the different medical groupings of participants. The sign test, which is also a nonparametric test, will be used to identify any significant changes to the correct rate. To assess any differences between the participant groups (nurses, doctors, and students), a Chi-square analysis will be performed. The participant's confidence in their answers will be assessed presearch and postsearch to identify any significant changes relating to search intervention, also using chi-square analysis.

To assess the impact of time constraints on clinical decision quality (RQ-2), the analyses above will be repeated with a breakdown by time constraint category, that is, 3, 6 and 9 min. In addition, an analysis of time-outs by constraint category will be conducted to assess the impact of time constraints on task completion. Time-outs are defined as samples where at the postsearch answer stage (1) the task timer reaches the constraint duration and (2) the participant provides no evidence to support their answer. It is assumed that in a time-out scenario, the participant was unable to complete the task. Significant differences by time-constraint category will be analyzed using the chi-square analysis. An analysis of variance (ANOVA) will be performed across confidence, difficulty, participant-perceived time impact assessments (impact of time on answer, evidence capture, and stress), and search behaviors, such as SERP dwell time, number of queries issued, number of documents opened, and the quantity of evidence items selected. To gain an understanding of the impact of providing a time constraint on the decision-making process, both the average time to search and the average proportion of available search time used will be evaluated and compared for the different task-timing samples. Tombros et al [67] used this proportional figure as a further gauge of participant stress and it can be compared with the reported stress by the participants.

To assess the impact of search engine performance on clinical decision quality (RQ-3), a similar set of analyses will be performed as that for RQ-1, except broken down by search system (state-of-the-art and BM25). In addition, the same ANOVA methods employed for time-constraint categories in RQ-2 analysis will be performed. In addition, an ANOVA will be performed across system categories and system time constraints to identify cases where system performance effects may matter most. The impact of search engine performance on clinical decision time (RQ-3) will also be assessed by evaluating the postsearch task completion times for those tasks that were completed (ie, relevant documents and/or evidence identified). This is measured in 2 ways: (1) search time only and (2) search time plus time spent filling in the answer form. Differences in search times between the systems will be assessed using the chi-square analysis. Finally, a participant-derived performance assessment of the 2 systems can be constructed by building a graded query relevance (QREL) listing (standard format for representing relevance assessments in IR), by query, based on all participants' relevance ratings. Using this QREL, a recomparison of the 2 systems can be evaluated and compared using the formal TREC evaluation results to provide better insight into any changes observed (or not) in the clinical decision and timing results for the 2 systems.

There are a number of potentially confounding factors within the experimentation. A covariant analysis (repeated measure ANOVA) will be performed on the task number, task at total duration point (for fatigue), and time transitions (eg, 3-min task to 6-min task and 3-min task to 9-min task).

Results

Recruiting began for the study in June 2018. As of April 4, 2019, there were 69 participants enrolled. The study is expected to close by May 30, 2019, with results to be published in July 2019.

Discussion

The study is currently underway, and results will be reported at the conclusion of participant testing.

Authors' Contributions

AV, GZ, and BK were responsible for identifying the RQs and designing the study. AV developed the custom search program, Taskiir, including the search engine and test interface. AV drafted the study protocol with comments and contributions from GZ and BK. AD was responsible for all expert medical input, including the selection of all medical tasks, formulation of some task, and answer confirmation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Pre-task questions and answer options.

[PDF File (Adobe PDF File), 57KB - resprot_v8i5e12803_app1.pdf]

Multimedia Appendix 2

Post tutorial instructions provided to the user prior to task assessment.

[PNG File, 507KB - resprot_v8i5e12803_app2.png]

Multimedia Appendix 3

User interaction time capture . Taskiir user-interaction data capture detail for Stage two of the study, when the participant can use a search engine to help them to complete their task. Column 2 identifies all captured variables as either time-stamped events or calculated variables. Column 3 identifies where the event is triggered or how the variable is calculated.

[PDF File (Adobe PDF File), 61KB - resprot_v8i5e12803_app3.pdf]

Multimedia Appendix 4

Task Presentation Order : Latin square design of task presentation order. The presentation order is from left to right. The numbers in the table represent the task numbers. The subjects are denoted in the first column from S1 to S16. Two such squares (32 subjects) form a timing cohort. System selection is alternated for each column of the square starting with the State of Art system in column one. Note that the task order for the first 16 subjects are the same as for the second 16 subjects, however the search system used for each task is switched, i.e., so that the BM25 system is used for column 1 questions.

[PDF File (Adobe PDF File), 35KB - resprot_v8i5e12803_app4.pdf]

Multimedia Appendix 5

Task timing selection for each question. Timing cohorts of 32 people are identified in the top row as C1, C2 and C3. Each cohort will conduct the search for each task, as listed in the first column (T1, T2...T16), within the time constraint specified in minutes in the table.

[PDF File (Adobe PDF File), 14KB - resprot_v8i5e12803_app5.pdf]

References

1. Ely JW, Osheroff JA, Gorman PN, Ebell MH, Chambliss ML, Pifer EA, et al. A taxonomy of generic clinical questions: classification study. Br Med J 2000 Aug 12;321(7258):429-432 [FREE Full text] [Medline: 10938054]
2. Coumou HC, Meijman FJ. How do primary care physicians seek answers to clinical questions? A literature review. J Med Libr Assoc 2006 Jan;94(1):55-60 [FREE Full text] [Medline: 16404470]
3. Ely JW, Osheroff JA, Ebell MH, Bergus GR, Levy BT, Chambliss ML, et al. Analysis of questions asked by family doctors regarding patient care. Br Med J 1999 Aug 07;319(7206):358-361 [FREE Full text] [Medline: 10435959]

XSL•FO

RenderX

4. Ely JW, Osheroff JA, Chambliss ML, Ebell MH, Rosenbaum ME. Answering physicians' clinical questions: obstacles and potential solutions. J Am Med Inform Assoc 2005;12(2):217-224 [FREE Full text] [doi: 10.1197/jamia.M1608] [Medline: 15561792]

5. Gorman PN, Helfand M. Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered. Med Decis Making 1995;15(2):113-119. [doi: 10.1177/0272989X9501500203] [Medline: 7783571]

6. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the evidence cart. J Am Med Assoc 1998 Oct 21;280(15):1336-1338. [Medline: 9794314]

7. Haynes RB, Sackett DL, Gray JM, Cook DJ, Guyatt GH. Transferring evidence from research into practice: 1. The role of clinical care research evidence in clinical decisions. ACP J Club 1996;125(3):A14-A16. [Medline: 8963526]

8. Wolters Kluwer. UpToDate URL: https://www.uptodate.com/home [accessed 2019-05-06] [WebCite Cache ID 78A8XVbl0]

9. Haynes R. Of studies, syntheses, synopses, summaries, and systems: the "5S" evolution of information services for evidence-based healthcare decisions. Evid Based Med 2006 Dec;11(6):162-164. [doi: 10.1136/ebm.11.6.162-a] [Medline: 17213159]

10. U.S. National Library of Medicine. MEDLINE®: Description of the Database URL: https://www.nlm.nih.gov/bsd/medline.html [accessed 2019-05-06] [WebCite Cache ID 78A9FGqXU]

11. U.S. National Library of Medicine. Key MEDLINE® Indicators URL: https://www.nlm.nih.gov/bsd/bsd_key.html [accessed 2019-05-06] [WebCite Cache ID 78AEZkz1d]

12. Dunn K, Marshall JG, Wells AL, Backus JEB. Examining the role of MEDLINE as a patient care information resource: an analysis of data from the Value of Libraries study. J Med Libr Assoc 2017 Oct;105(4):336-346. [doi: 10.5195/jmla.2017.87] [Medline: 28983197]

13. McKibbon KA, Fridsma DB. Effectiveness of clinician-selected electronic information resources for answering primary care physicians' information needs. J Am Med Inform Assoc 2006;13(6):653-659. [doi: 10.1197/jamia.M2087] [Medline: 16929042]

14. Ellsworth MA, Homan JM, Cimino JJ, Peters SG, Pickering BW, Herasevich V. Point-of-care knowledge-based resource needs of clinicians. Appl Clin Inform 2015;6(2):305-317. [doi: 10.4338/aci-2014-11-ra-0104] [Medline: 26171077]

15. Hoogendam A, Stalenhoef AF, Robbé PF, Overbeke AJ. Answers to questions posed during daily patient care are more likely to be answered by UpToDate than PubMed. J Med Internet Res 2008;10(4):e29 [FREE Full text] [doi: 10.2196/jmir.1012] [Medline: 18926978]

16. Westbrook JI, Coiera EW, Gosling AS. Do online information retrieval systems help experienced clinicians answer clinical questions? J Am Med Inform Assoc 2005;12(3):315-321 [FREE Full text] [doi: 10.1197/jamia.M1717] [Medline: 15684126]

17. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Rose L, Friedman CP. Factors associated with successful answering of clinical questions using an information retrieval system. Bull Med Libr Assoc 2000 Oct;88(4):323-331 [FREE Full text] [Medline: 11055299]

18. Hersh WR, Crabtree MK, Hickam DH, Sacherek L, Friedman CP, Tidmarsh P, et al. Factors associated with success in searching MEDLINE and applying evidence to answer clinical questions. J Am Med Inform Assoc 2002;9(3):283-293 [FREE Full text] [Medline: 11971889]

19. Ordóñez L, Benson IL. Decisions under time pressure: how time constraint affects risky decision making. Organ Behav Hum Decis Process 1997 Aug;71(2):121-140. [doi: 10.1006/obhd.1997.2717]

20. Svenson O, Benson L. Framing and time pressure in decision making. In: Time Pressure and Stress in Human Judgment and Decision Making. Boston: Springer; 1993:133-144.

21. Edland A, Svenson O. Judgment and decision making under time pressure. In: Time Pressure and Stress in Human Judgment and Decision Making. Boston: Springer; 1993:27-40.

22. Johnson E, Payne J, Bettman J. Adapting to time constraints. In: Time Pressure and Stress in Human Judgment and Decision Making. Boston: Springer; 1993:103-116.

23. Wright P. The harassed decision maker: time pressures, distractions, and the use of evidence. J Appl Psychol 1974;59(5):555-561. [doi: 10.1037/h0037186]

24. Liu C, Wei Y. The impacts of time constraint on users' search strategy during search process. 2016 Dec 27 Presented at: Proceedings of the Association for Information Science and Technology banner; 2016; Copenhagen, Denmark p. 1-9. [doi: 10.1002/pra2.2016.14505301051]

25. Crescenzi A, Capra R, Arguello J. Time pressure, user satisfaction and task difficulty. In: Proceedings of the 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries. 2014 May 08 Presented at: ASIST '13; 2013; Montreal, Quebec, Canada p. 1-4. [doi: 10.1002/meet.14505001121]

26. Crescenzi A, Kelly D, Azzopardi L. Impacts of time constraints and system delays on user experience. In: Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval. New York: Association for Computing Machinery, Inc; 2016 Presented at: CHIIR '16; March 13-17, 2016; Carrboro p. 141-150. [doi: 10.1145/2854946.2854976]

27. Tsiga E, Panagopoulou E, Sevdalis N, Montgomery A, Benos A. The influence of time pressure on adherence to guidelines in primary care: an experimental study. BMJ Open 2013;3(4):e002700 [FREE Full text] [doi: 10.1136/bmjopen-2013-002700] [Medline: 23585394]

XSL•FO

RenderX

28. Cook DA, Sorensen KJ, Hersh W, Berger RA, Wilkinson JM. Features of effective medical knowledge resources to support point of care learning: a focus group study. PLoS One 2013;8(11):e80318 [FREE Full text] [doi: 10.1371/journal.pone.0080318] [Medline: 24282535]

29. Turpin A, Scholer F. User Performance versus Precision Measures for Simple Search Tasks. In: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. 2006 Presented at: SIGIR '06; August 06-11, 2006; Seattle, Washington, USA p. 11-18. [doi: 10.1145/1148170.1148176]

30. Hersh W, Turpin A, Price S. Do batch and user evaluations give the same results? In: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval. 2000 Presented at: SIGIR '00; July 24-28, 2000; Athens, Greece p. 17-24. [doi: 10.1145/345508.345539]

31. Allan J, Carterette B, Lewis J. When Will Information Retrieval Be "Good Enough"? - User Effectiveness As a Function of Retrieval Accuracy. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. 2005 Presented at: SIGIR'05; 2005; Salvador, Brazil p. 433-440. [doi: 10.1145/1076034.1076109]

32. Moffat A, Thomas P. Users Versus Models?: What Observation Tells Us About Effectiveness Metrics. In: Proceedings of the 22nd ACM international conference on Information & Knowledge Management. 2013 Presented at: CIKM '13; October 27-November 01, 2013; San Francisco, California, USA p. 13. [doi: 10.1145/2505515.2507665]

33. Westbrook J, Gosling A, Coiera E. The impact of an online evidence system on confidence in decision making in a controlled setting. Med Decis Making 2005;25(2):178-185. [doi: 10.1177/0272989X05275155] [Medline: 15800302]

34. Yoo I, Mosa A. Analysis of PubMed user sessions using a full-day PubMed query log: a comparison of experienced and nonexperienced PubMed users. JMIR Med Inform 2015 Jul 02;3(3):e25 [FREE Full text] [doi: 10.2196/medinform.3740] [Medline: 26139516]

35. Ramos K, Linscheid R, Schafer S. Real-time information-seeking behavior of residency physicians. Fam Med 2003 Apr;35(4):257-260. [Medline: 12729309]

36. The University of Queensland Australia. 2019. Assessment of clinical decision making using scientific medical literature under time pressure URL: https://espace.library.uq.edu.au/view/UQ:3da8527 [accessed 2019-05-08] [WebCite Cache ID 78E8O0hgZ]

37. Simpson MS, Voorhees EM, Hersh W, Roberts K. NIST. 2015. Overview of the TREC 2015 Clinical Decision Support Track URL: https://trec.nist.gov/pubs/trec24/papers/Overview-CL.pdf [accessed 2019-05-08] [WebCite Cache ID 78DARpYjX]

38. Kaur P, Kushtagi P. Plastic spatula with narrow long tip provides higher satisfactory smears for Pap test. J Cytol 2013 Jul;30(3):159-161 [FREE Full text] [doi: 10.4103/0970-9371.117654] [Medline: 24130406]

39. Del Pup L, Lino DP, Canzonieri V, Vincenzo C, Serraino D, Diego S, et al. What sampling device is the most appropriate for vaginal vault cytology in gynaecological cancer follow up? Radiol Oncol 2012 Jun;46(2):166-169 [FREE Full text] [doi: 10.2478/v10019-012-0019-x] [Medline: 23077454]

40. Capo H, Kupersmith MJ, Berenstein A, Choi IS, Diamond GA. The clinical importance of the inferolateral trunk of the internal carotid artery. Neurosurgery 1991 May;28(5):733-7; discussion 737. [doi: 10.1097/00006123-199105000-00018] [Medline: 1876255]

41. Uchida Y, Nomura A, Ohtsuka M, Hamada M, Hasegawa S. [Bronchoconstriction and neuropeptides]. Kokyu To Junkan 1989 May;37(5):493-497. [Medline: 2749010]

42. Bullock CB, Pridham KF. Sources of maternal confidence and uncertainty and perceptions of problem-solving competence. J Adv Nurs 1988 May;13(3):321-329. [Medline: 3417926]

43. Mortensen SA, Aabo K, Jonsson T, Baandrup U. Clinical and non-invasive assessment of anthracycline cardiotoxicity: perspectives on myocardial protection. Int J Clin Pharmacol Res 1986;6(2):137-150. [Medline: 3459718]

44. Lederberg J. Medical science, infectious disease, and the unity of humankind. JAMA 1988 Aug 05;260(5):684-685. [Medline: 3392795]

45. Shakhnazarov AM, Alkadarskiĭ AS. [Adenomatosis of the proximal segment of the common hepatic duct]. Arkh Patol 1988;50(2):64-65. [Medline: 3377702]

46. Rhodin JA. Perfusion and superfusion fixation effects on rat mesentery microvascular beds. Intravital and electron microscope analyses. J Submicrosc Cytol 1986 Jul;18(3):453-470. [Medline: 3091849]

47. Deffez JP, Rostand B, Brethaux J, Allain P, Hazen M. [Capdepont dysplasia and vertical excess of the lower half of the face. Genetic syndrome or fortuitous association?]. Rev Stomatol Chir Maxillofac 1986;87(3):131-137. [Medline: 3464066]

48. Eckstein JW. Some measures of medical education. J Iowa Med Soc 1976 Apr;66(4):124-128. [Medline: 1262723]

49. Townsend MB. Creating a better work environment. Measuring effectiveness. J Nurs Adm 1991 Jan;21(1):11-14. [Medline: 1987337]

50. Chiappino G, Friedrichs KH, Forni A, Rivolta G, Todaro A. Alveolar and lung fibre levels in non-occupationally exposed subjects. IARC Sci Publ 1989(90):310-313. [Medline: 2744830]

51. Saint-Maurice C. [Pediatric anesthesia-resuscitation--the future?]. Cah Anesthesiol 1987 Sep;35(5):389-390. [Medline: 3664345]

52. Kalden JR. Clinical autoimmunity. Immunol Today 1989 Aug;10(8):S31-S33. [Medline: 2803497]

53. Ferguson HW. Biophysical profile scoring: the fetal Apgar. Am J Nurs 1988 May;88(5):662-663. [Medline: 3364475]

54. Baker EJ. Valvoplasty, angioplasty and embolotherapy in congenital heart disease. Int J Cardiol 1986 Aug;12(2):139-145. [Medline: 2943682]

55. Hicks CB, Boslego JW, Brandt B. Evidence of serum antibodies to Neisseria gonorrhoeae before gonococcal infection. J Infect Dis 1987 Jun;155(6):1276-1281. [doi: 10.1093/infdis/155.6.1276] [Medline: 2883240]

56. Araujo LI, Maseri A. Whole heart distribution of myocardial perfusion, metabolism and myocardial viability by positron emission tomography. Br Med Bull 1989 Oct;45(4):922-932. [doi: 10.1093/oxfordjournals.bmb.a072374] [Medline: 2695219]

57. Pettinato G, Insabato L, De Chiara A, Ruocco V, Bucci L. Fine-needle aspiration cytology of angiolymphoid hyperplasia with eosinophilia: a case report with electron microscopy and immunohistochemistry. Diagn Cytopathol 1989;5(1):88-94. [Medline: 2721357]

58. Spitzer ED, Travis SJ, Kobayashi GS. Comparative in vitro activity of LY121019 and amphotericin B against clinical isolates of Candida species. Eur J Clin Microbiol Infect Dis 1988 Feb;7(1):80-81. [Medline: 3132386]

59. Clark CJ, Cochrane LM. Assessment of work performance in asthma for determination of cardiorespiratory fitness and training capacity. Thorax 1988 Oct;43(10):745-749 [FREE Full text] [doi: 10.1136/thx.43.10.745] [Medline: 3206382]

60. Simpson M, Voorhees E, Hersh W. NIST. 2014. Overview of the TREC 2014 Clinical Decision Support Track URL: https://pdfs.semanticscholar.org/fcf8/1b7641c0cd7be089051018a53fabfa685da0.pdf [accessed 2019-05-08] [WebCite Cache ID 78DAVuMGP]

61. Balaneshin-kordan S, Kotov A, Xisto R. WSU-IR at TREC 2015 Clinical Decision Support Track: Joint Weighting of Explicit and Latent Medical Query Concepts from Diverse Sources. 2015 Presented at: Proceedings of the 2015 Text Retrieval Conference; November 2015; Detroit, USA p. 593.

62. Galago. URL: https://www.lemurproject.org/galago.php [accessed 2019-05-06] [WebCite Cache ID 78B0PaFy0]

63. Elasticsearch. URL: https://www.elastic.co/products/elasticsearch [accessed 2019-05-06] [WebCite Cache ID 78B1ETyqO]

64. Lucene. Welcome to Apache Lucene URL: https://lucene.apache.org/ [accessed 2019-05-06] [WebCite Cache ID 78B1h4PSW]

65. Aslam J, Pavlu V, Yilmaz E. A statistical method for system evaluation using incomplete judgments. 2006 Presented at: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval; August 06-11, 2006; Seattle, Washington, USA p. 541-548. [doi: 10.1145/1148170.1148263]

66. Shao J, Wang H, Chow SC. Sample Size Calculations in Clinical Research. Boca Raton: CRC Press; Aug 27, 2007:89-89.

67. Tombros A, Ruthven I, Jose J. How users assess web pages for information seeking. J Am Soc Inf Sci 2005 Feb 15;56(4):327-344. [doi: 10.1002/asi.20106]

Abbreviations

ANOVA: analysis of variance
CDS: clinical decision support
DCG: discounted cumulative gain
IR: information retrieval
MAP: mean average precision
nDCG: normalized discounted cumulative gain
QREL: query relevance
RQ: research question
R-prec: R-precision
SERP: search engine results page
SML: scientific medical literature
TREC: text retrieval conference
UMLS: Unified Medical Language System

XSL•FO
RenderX

Protocol

Validation of the InnoWell Platform: Protocol for a Clinical Trial

Tracey A Davenport[1,2], BA (Hons), eMBA; Haley M LaMonica[1,2], BSc, MA, PhD, ABPP-CN; Lisa Whittle[1,2], BSc (Psych); Amelia English[2], BSc (Hons); Frank Iorfino[1,2], BSc (Psych), MBMSc, PhD; Shane Cross[1], BPsy (Hons), MPsych (Clinical), PhD; Ian B Hickie[1], AM, MD, FRANZCP, FASSA

[1]Brain and Mind Centre, The University of Sydney, Camperdown, Australia

[2]InnoWell Pty Ltd, Camperdown, Australia

Corresponding Author:
Tracey A Davenport, BA (Hons), eMBA
Brain and Mind Centre
The University of Sydney
94 Mallett Street
Camperdown, 2050
Australia
Phone: 61 0404839897
Email: tracey.davenport@sydney.edu.au

Abstract

Background: New electronic health technologies are being rapidly developed to improve the delivery of mental health care for both health professionals and consumers and better support self-management of care. We developed a Web-based platform (the InnoWell Platform) that supports the prevention, early intervention, treatment, and continuous monitoring of mental health and maintenance of well-being in people aged 2 years and older. The platform is a customizable digital tool kit that operates through existing service providers who utilize the system to provide their consumers with access to evidence-based assessments and feedback, intervention options, and outcome monitoring. It does this by collecting, storing, and reporting personal and health information back to consumers and their health professionals to promote collaborative care partnerships that aim to improve the management of mental ill health and maintenance of well-being

Objective: The aim of this study was to describe the research protocol for a naturalistic prospective clinical trial wherein all consumers presenting for care to a traditional face-to-face or Web-based mental health service in which the InnoWell Platform is being offered as part of standard clinical care will be given the opportunity to use the platform.

Methods: The Web-based platform is a configurable and customizable digital tool that assists in the assessment, monitoring and management of mental ill health, and maintenance of well-being. It does this by collecting, storing, and reporting health information back to the person and his or her clinician to enable transformation to person-centered care. The clinical trial will be conducted with individuals aged 2 years and older presenting to participating services for care, including persons from the veteran community, Aboriginal and Torres Strait Islander people, people from culturally and linguistically diverse backgrounds, the lesbian, gay, bisexual, transgender, and intersex community, and those from broader education and workforce sectors, as well as people with disabilities, lived experience of comorbidity, complex disorders, and suicidality.

Results: Project Synergy was funded in June 2017, and data collection began in November 2018 at a youth mental health service. At the time of this publication, 5 additional services have also begun recruitment, including 4 youth mental health services and a veteran's service. The first results are expected to be submitted in 2020 for publication.

Conclusions: This clinical trial will promote access to comprehensive, high-quality mental health care to improve outcomes for consumers and health professionals. The data collected will be used to validate a clinical staging algorithm designed to match consumers with the right level of care and reduce the rate of suicidal thoughts and/or behaviors and suicide by suggesting pathways to care that are appropriate for the identified level of need, while simultaneously enabling a timely service response.

Trial Registration: Australian New Zealand Clinical Trial Registry ACTRN12618001676202; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=374632 (Archived by WebCite at http://www.webcitation.org/78TOi5jwl)

International Registered Report Identifier (IRRID): DERR1-10.2196/13955

XSL•FO
RenderX

Introduction

Background

As we move further into the twenty-first century, major revolutions in technology are transforming the way we live our lives—from the way we socialize, work, access information, and receive services. The mental health sector, like many others, is undergoing immense change in the face of this revolution, whereby new technologies are being developed at a rapid pace, many of which challenge the way people receive and manage their health care [1]. This is evident in the dramatic increase in the number of mobile apps, internet-based resources, and platforms that target mental health problems [2]. Although it remains a major priority to develop and assess effective Web-based interventions, an equally important task is to determine how these technologies can effectively integrate with existing health services in ways that improve the delivery of mental health care for the service, consumers' health professionals, and presenting consumers [3].

The Australian Federal Government initiated the National Mental Health Strategy in 1992 [4], promoted by 5-yearly government plans outlining priority areas for investment and reform [5-9]. Despite the efforts of the last 25 years, the most recent report from the National Mental Health Commission (NMHC) continues to highlight fundamental shortcomings within the mental health system, including persistent stigma, poor experiences of care by individuals with lived experience, families and support people, delays in service provision, fragmented services, and inefficient and ineffective use of resources [10]. The NMHC review emphasizes the key concepts of person-centered care, personalized care options, regionalization of mental health services and suicide prevention, adoption of new technologies, and systematic evaluation to drive future investments [11]. In relation to the latter, the Fifth National Mental Health and Suicide Prevention Plan now includes 24 national key performance indicators that can be reported on to track the progress of the plan [9], including prevalence of mental illness, changes in mental health consumers' clinical outcomes, rates of employment and social, community, and family participation among individuals with mental illness, and rates of suicide. Technology is increasingly recognized as a way to support and drive mental health service reform. Electronic health (eHealth) is broadly defined by the World Health Organization as the use of information and communication technologies for health-related purposes, such as service delivery [12]. Various models of eHealth services have been shown to be successful, including stand-alone systems for symptom prevention and self-help, consumer-assisted care, such as peer support, virtual clinics offering professional care, and stepped-care systems for integrated care [13]. The Australian Government published its "E-mental health strategy for Australia" in 2012 to promote its broader mental health reform agenda, emphasizing the ability of eHealth apps, etools, and platforms to deliver evidence-based interventions on the Web as a complement to traditional face-to-face services to improve

health outcomes [14]. Importantly, eHealth services can improve access to mental health care by overcoming issues of distance, cost, and stigma.

Stage-Based Stepped Care

In addition to revolutions in technology to support service delivery, there have also been advances in service models that aim to match consumers to the right level of care (ie, right care, right time). Clinical staging models are commonly employed in medical settings for this purpose, and they have more recently been implemented in clinical psychiatry [15]. Staging models consider the spectrum of mental ill health, and they aim to place consumers on that continuum, from those with risk factors with symptoms or impairment through to those with persistent and recurrent syndromes [15]. The emphasis is then on matching consumers at various stages to interventions that are appropriate to that stage [16]. As highlighted in Textbox 1, a staged-based stepped-care model combines the principles of clinical staging with the objectives of stepped care, including the use of low-intensity interventions for those at early stages and offering more intensive interventions to consumers at higher stages, while monitoring outcomes to increase or decrease service intensity as needs change [3].

Suicide Prevention

The Zero Suicide in Health Care International Declaration was developed collaboratively by international leaders in mental health and addiction services with the aim of making suicide a "never event" [17]. This global initiative includes crucial recommendations and targets to promote suicide prevention worldwide, including fostering a safety-oriented culture to reduce the rates of suicide for consumers under care, employing an evidence-based approach to enhancing routine care and driving system changes that will result in improved outcomes and better care for those identified as being at risk, investing in training for health professionals aimed at improving identification, assessment, and management of suicidal thoughts and/or behaviors and risk factors thereof (eg, chronic pain, substance misuse), and systematically identifying, assessing, and monitoring suicidality in the broad consumer base for a service for the purposes of triaging individuals to the appropriate intensity of care. The abovementioned goals are synonymous with the core principles of the InnoWell Platform, which include increasing access to standardized, broad-based assessment, identifying and tracking consumer needs, matching those needs with personalized care options without having to wait for an appointment, enhancing the quality of the care provided to consumers, and supporting and guiding health professionals at all levels of experience to foster skill and professional development. Furthermore, the platform's suicide escalation protocol is specifically designed to facilitate the detection of suicidal thoughts and/or behaviors and suggest pathways to care that are appropriate for the identified level of need, while simultaneously enabling a timely response from health professionals for those consumers reporting high suicidality [18]. Suicide prevention was highlighted as one of the priority

XSL·FO

RenderX

areas in the Fifth National Mental Health and Suicide Prevention Plan [9]. Government agencies, service providers, and community agencies have all been identified as being key to reducing suicide rates and improving health outcomes. Despite efforts to improve suicide prevention in Australia, there has been no reduction in suicide rates over the past decade [19].

Textbox 1. Key features of stage-based stepped care.

- Provide broad and holistic initial screening, followed by more targeted mental health assessment for those who endorse screening questions

- The intensity of the intervention should be matched to the consumer's level of need as determined by clinical stage

- Provide parallel interventions for risk factors associated with poor outcomes (eg, unemployment, alcohol, and/or other substance misuse)

- Employ proactive monitoring of treatment progress and outcomes

The InnoWell Platform

Through a process of participatory design with lived experience, health professionals, and service staff (including administration and management), we have developed a Web-based platform (the InnoWell Platform) that supports the prevention, early intervention, treatment, and continuous monitoring of mental ill health and maintenance of well-being in people aged 2 years and older [20]. The platform is a customizable digital tool kit that operates through existing service providers who utilize the system to provide their consumers with access to evidence-based assessments and feedback, intervention options, and outcome monitoring. It does this by collecting, storing, and reporting personal and health information back to consumers and their health professionals to promote collaborative care partnerships that aim to improve the management of mental ill health and maintenance of well-being. This platform uses multiple sources of information to develop a comprehensive understanding of the consumers' needs and track their progress over time. This primarily involves Web-based self-reported psychometric measures from both consumers and their health professionals, as well as objective behavioral data collected via third-party integrations. Textbox 2 highlights the functionality built into the InnoWell Platform.

Specifically, in relation to suicide prevention, the InnoWell Platform employs a suicide escalation protocol derived from previous work [18]. In the current platform, suicidal thoughts and/or behaviors are assessed using 2 self-report surveys: Suicidal Ideation Attributes Scale [21] and Columbia-Suicide Severity Rating Scale [22]. Suicidal ideation without plan or intent in the past month triggers a pop-up for the consumer in the platform, providing details for crisis support services.

Self-report of suicidal ideation with a plan and/or intent and/or a suicide attempt within the past 3 months triggers both the pop-up for the consumer and sends a notification to the consumer's health professional, requiring a clinical care action (eg, telephone contact, action safety plan).

Primary Objective

The primary objective of this clinical trial (Australian New Zealand Clinical Trial Registry ACTRN12618001676202) is to validate the algorithms used to derive stage-based stepped care on data collected by the platform and previously published staging criteria [15]. To do this, 4 independent ratings of stage-based stepped care will be compared: (1) platform generated, (2) health professional, (3) multidisciplinary team via consensus, and (4) expert-clinician reference group (comprising 3 or more academic health professionals with "area of research expertise" in stage-based stepped care for mental ill health). This expert group will review the clinical data of a subset of randomly generated consumers and independently allocate clinical stage.

Secondary Objective

In accordance with the Fifth National Mental Health and Suicide Prevention Plan [9], the secondary objective of this clinical trial is to reduce the rate of suicidal thoughts and/or behaviors by promoting access to comprehensive services and reducing barriers to care, as well as improving service quality and evidence-based clinical interventions. It is hypothesized that use of the platform within traditional mental health services will facilitate earlier identification and rapid service response (eg, earlier appointment with a health professional) to high levels of suicidal thoughts and/or behaviors.

Textbox 2. InnoWell Platform functionality.

- Multidimensional assessment across a range of domains (eg, sleep, anxiety, mood, physical health)
- Suicidal thoughts and/or behaviors and at-risk mental state identification and subsequent escalation to required intervention
- Dashboard of results across the range of biopsychosocial domains
- Algorithms to determine severity of needs across these biopsychosocial domains
- Algorithms to allocate clinical stage and suggested levels of care
- Real-time data tracking and interactive progress report
- Shared care planning between treating health professional and consumer
- Video-visit with a health professional
- Support person input and health information sharing
- Coordination of care across multi-disciplinary services
- Aggregate service performance indicator dashboard

Methods

Study Design

This paper employs a naturalistic prospective clinical trial design wherein all consumers presenting for care to a service utilizing the InnoWell Platform as part of their standard clinical care will be offered the opportunity to use the platform.

Trial Site and Participating Centers

The InnoWell Platform is a Web-based app, and all data are collected electronically. As the sponsor and locality of the research team, the University of Sydney is considered the physical trial site. At the time of this publication, participants were being recruited from the following participating centers: headspace services (Camperdown, Coffs Harbour, Lismore, and Port Macquarie in New South Wales, and Edinburgh North in South Australia) and Open Arms—Veterans and Families Counselling in Surry Hills, New South Wales. Future participating centers may include services for children and their families, adult staged-care services, older persons mental health, general practice, as well as Aboriginal and Torres Strait Islander designed and controlled services.

Participants

Eligibility Criteria

All consumers aged 2 years and older presenting for care to a traditional face-to-face or Web-based mental health service utilizing the platform as part of their standard clinical care will be eligible to participate in the clinical trial. This will include persons from the following populations: children and their families, young people, adults, older adults, the Veteran community, Aboriginal and Torres Strait Islander peoples, people from culturally and linguistically diverse backgrounds, the lesbian, gay, bisexual, transgender, and intersex community, and those from broader education and workforce sectors, as well as persons with disabilities, lived experience of comorbidity (including alcohol or other substance misuse), complex disorders, and/or suicidality. Owing to the nature of its research design, this clinical trial does not have defined exclusion criteria.

Sample Size

The clinical trial does not have an upper or lower limit on the number of participants.

Screening Procedures

It will be standard clinical care for all consumers presenting to a traditional face-to-face or Web-based mental health service utilizing the InnoWell Platform to be directed to the platform for assistance in assessment, management, and monitoring of their mental ill health and maintenance of well-being.

On contacting the service either by telephone (or short message service), on the Web (eg, email, chat), or in person, the consumer will be introduced to the clinical trial via an ethics-approved script. Although this script is generic and adaptable to different populations, settings, and services, it informs the consumer of what is involved in using the platform and notifies them that the data collected by the platform will be shared with their service provider to promote person-centered collaborative care and shared decision making. If the consumer is interested and responds "Yes," they will receive a unique email invitation to sign up to the platform. When the participants receive the email invitation, they will be required to accept the invitation and be directed to "Sign up" by creating an account. If they do not sign up immediately, they will be sent a reminder via short message service. The participants will be required to review and accept both the "Privacy" and "Terms of use" within the platform before proceeding to a "Research data sharing" screen, which is where details about this clinical trial are contained.

Informed Consent

Opt Out Process

As part of the standard "Terms of use" of the platform, participants aged 14 years and older will be informed that their deidentified health data collected by the platform will be used for research purposes unless they "opt out." If participants decide to "opt out," they can indicate this by ticking a box. This action will be noted in the platform's database, and it will automatically switch their access to the platform to read only, preventing them from entering any new data. If participants do

not "opt out," their deidentified data will be accessible by members of the research team. Importantly, participants will be able to edit their data sharing permissions at any time, such that any future data collected about them will not be accessible to researchers if, for example, they choose to withdraw from the clinical trial. The "opt out" consent approach described above is an efficient procedure without violating the option of providing choice. The approach taken considers a participant's willingness rather than refusal to participate in the clinical trial and provides the necessary information to make an informed decision. Finally, the risk to the participants can be considered very low, as their data are deidentified (ie, all personal identifiers removed, including name, date of birth, and email address).

Parental Consent Requirements and Process

For children between the ages of 2 and 11 years, consent for the deidentified data to be used for research purposes will be provided solely by the parents/guardians. Child assent will not be sought. The parents/guardians will also serve as the "consumer" in these instances, answering the surveys in relation to the child. For consumers aged 12 and 13 years, parents/guardians will again be required to provide consent for the young person's deidentified data to be used for research purposes. In addition, as per the standard process described above, the young person will be given the option to "opt out" of sharing his or her data through the standard "Terms of use" provided when participants access the platform. For deidentified data to be used for research purposes, both parents/guardians and the young person (if aged 12 or 13 years) need to have provided permission to do so (ie, if the parent consents but the young person "opts out," the data will not be used and vice versa). Parents/guardians of children aged 13 years or younger will receive a standard introduction email to the platform sent by the service, which will include brief information regarding the research, as well as a link to further details, and a page where parents/guardians can indicate consent (or not) for the use of their young person's data for research purposes. This parental consent is specifically related to the storage and use of their young person's data for this clinical trial. Thus, if a parent/guardian chooses not to consent to the use of data for this clinical trial, there will be no impact on the young person's standard clinical care; however, their access to the platform will be automatically switched to read only.

The InnoWell Platform Description

The InnoWell Platform is being embedded within traditional face-to-face and Web-based mental health services, and it is being offered to consumers presenting to those services as part of standard clinical care. It is a Web-based platform, and it can be accessed via traditional computing and mobile devices. As outlined in Textbox 2, the platform allows consumers to complete Web-based clinical assessments to understand their needs; explore their personalized dashboard of results, including current symptoms, level of functioning, and health history (see Figure 1); select from recommended care options (eg, fact sheets, apps, etools, and other Web-based systems) to support their mental health and well-being (see Figure 2); track their progress (in real time); share their dashboard and plan with their health professional(s) to support care.

Participant Procedures

Participants will use the platform of their own accord as part of their standard clinical care with the mental health service. No other participation is required. All survey data are part of the functionality of the platform, which participants complete as part of standard clinical care through their service. The surveys provide assessment across a range of biopsychosocial domains, including psychological distress, suicidal thoughts and/or behaviors, social and occupational functioning, depressed mood, anxiety, sleep-wake cycle, social connectedness, psychosis-like experiences, mania-like experiences, alcohol use, tobacco use, self-harm, physical health, posttraumatic stress, eating behaviors, and body image. These domains, along with the associated surveys configured in the platform, are determined by services, in conjunction with the researchers to meet the needs of their consumers. Table 1 provides an example of the surveys that may be included across a sampling of biopsychosocial domains.

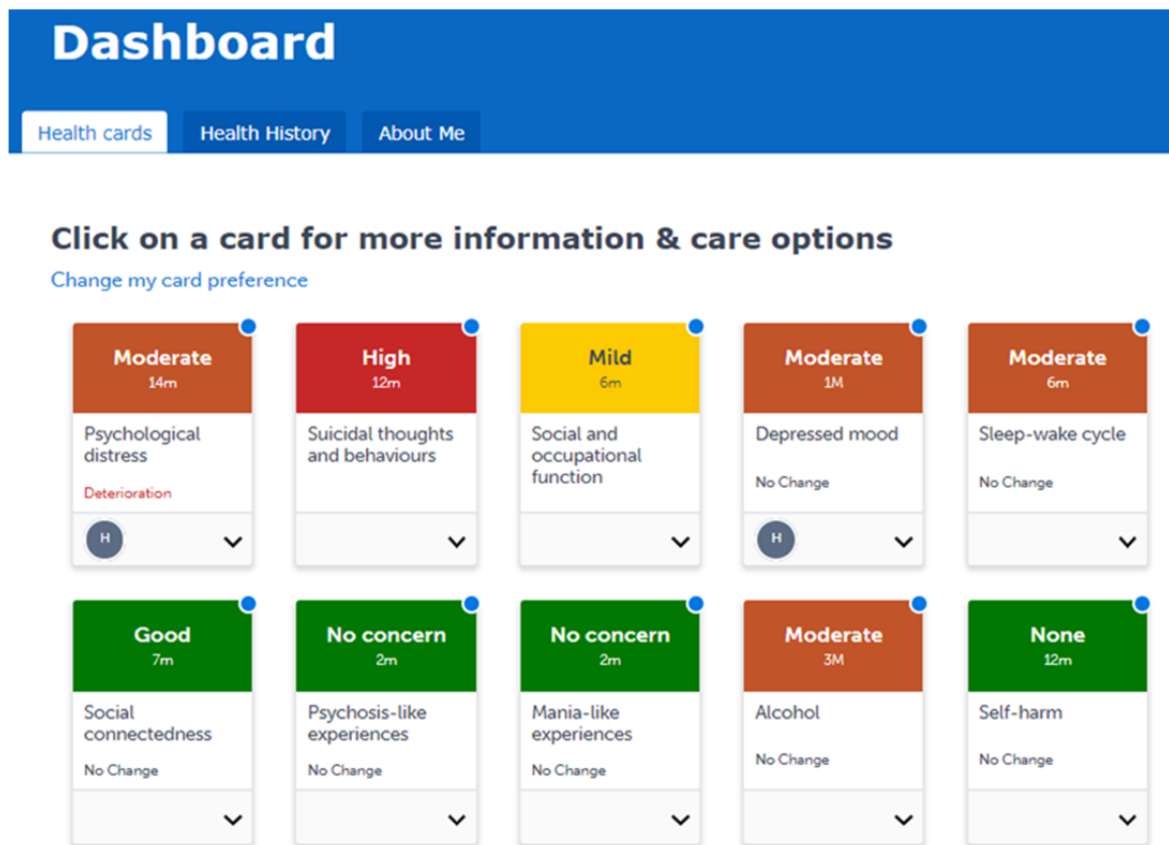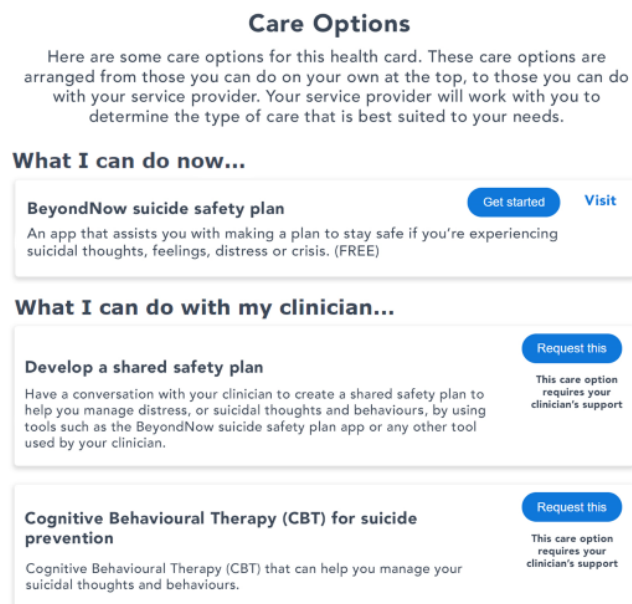**Figure 1.** Sample personalized dashboard of results.

Figure 2. Sample clinical and nonclinical care options.

Table 1. Example domains and surveys.

Example domain	Survey
Anxiety	• Overall Anxiety Severity and Impairment Scale [23] • Generalized Anxiety Disorder Scale [24]
Depression	• Quick Inventory of Depressive Symptomatology [25] • Patient Health Questionnaire [26]
Functioning	• Instrumental Activities of Daily Living and Physical Self-Maintenance Scale [27] • Activities of Daily Living Index [28]
Trauma	• Primary Care Posttraumatic Stress Disorder Screen for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [29] • Posttraumatic Stress Disorder Checklist [30]
Suicide behaviors	• Columbia-Suicide Severity Rating Scale [22]

Completion of the surveys serves to populate a participant's dashboard, which comprises health cards reflecting the scores (eg, mild, moderate, and high) generated from the participant's responses (see Figure 1). The abbreviations under the scores reflect the time since the survey was last completed by the participant. For example, "1M" indicates the survey was completed one month earlier, and "14m" indicates the survey was completed 14 min earlier. By clicking on a health card in their dashboard, the participants can learn more about the biopsychosocial domain (eg, *This card tracks your symptoms of depression. Some signs of depression include feelings of worthlessness or losing interest in enjoyable activities.*). Links to psychoeducational resources are also available for further details about symptoms and treatment options. Each health card also contains a range of different care options, commonly known as interventions, to help the participant in managing a particular area of health. Care options are divided into 2 types: clinical and nonclinical (Figure 2). Clinical care options require a health professional's involvement, such as individual therapy and group therapy. In contrast, a participant can immediately access and begin using nonclinical care options, such as apps, etools, or other systems, without the support of a health professional.

Given the emphasis on suicide prevention, the InnoWell Platform has been built to include immediate access to crisis support options. At any time when engaging with the platform, a participant has access to the "I need help now" button (see Figure 3), which provides contact details for crisis services (eg, Lifeline, Beyond Blue, and Kids Helpline). As described previously, these support options also appear as a pop-up directly after the participant scores low, moderate, or high on the suicidal thoughts and/or behaviors health card. When a participant scores high on this health card, a notification of suicide risk is also triggered for the treating health professional requiring clinical action.

Figure 3. The "I need help now" button.

Configuration

The InnoWell Platform is highly configurable, allowing each service to adapt the technology to the needs of its consumer base. Although there is a set of core biopsychosocial domains that all services will be encouraged to assess (ie, psychological distress, suicidal thoughts and/or behaviors, psychosis-like experiences, mania-like experiences, social and occupational functioning, self-harm, tobacco use, alcohol use, social connectedness, depression, anxiety, and posttraumatic stress), the measures used to evaluate these areas can be tailored to the service. Similarly, the algorithms used to derive the thresholds for each domain will be based on the most appropriate psychometrics for that population (eg, young people, veterans, and older adults). As the platform is implemented across multiple health services and accumulates more data, these algorithms will become digitally smart.

Education and Training Requirements

As stated previously, the implementation of clinical staging models is relatively new to clinical psychiatry [15], and therefore the education and training requirements of health professionals will need to be scoped to ensure consistency of application in each health service. As required, the training will cover the theoretical and scientific underpinnings of the model, the clinical assessment requirements, detailed criteria used to assign stage-based stepped care, the application of the model to subsyndromal, prodromal, or mixed syndromes, and the clinical utility of the model for the purposes of planning, implementing, and monitoring treatment. The method of training delivery may vary from service to service to ensure access and broad distribution.

Outcomes

Primary Outcome

The platform-generated assignment of stage-based stepped care derived from an algorithm calculated from data collected by the platform will be compared with that allocated by the health professional working directly with the consumer, the within-service multidisciplinary consensus team, as well as by an independent expert-clinician reference group, all of which will rely on previous published clinical staging criteria [15]. For example, to meet criteria for Stage 2 or higher for depression, the disorder needs to have features indicative of more severe disorders, including psychomotor retardation, agitation, impaired cognitive function, severe circadian dysfunction, psychotic features, brief hypomanic periods, severe neurovegetative changes, or severe suicidality. Although a clinician will assess these features as part of standard clinical care to assign clinical stage, the platform will rely on the scores on several health cards (ie, depression, mania-like experiences, psychotic-like experiences, alcohol use, and eating behaviors and body image) for this purpose. The primary outcome will be the reliability of the ratings across these 4 methods of assigning clinical stage as assessed using Cohen kappa coefficient.

Secondary Outcome

Service-level performance data regarding improvements in safety and clinical quality will be derived from the platform,

with an emphasis on identification and rapid response to consumers endorsing suicidal thoughts and/or behaviors (see Figure 4). Outcomes will include the percentage of consumers endorsing suicidal thoughts and/or behaviors across 4 severity levels (ie, none, low, medium, and high) at service entry, the number of suicide escalations (ie, notifications to health professionals) for a specific period and during an entire episode of care, and time between suicide escalation and an action by a health professional (eg, call to consumer, schedule follow-up appointment, and contact emergency services).

Data Collection, Management, and Security

All collected data are stored in the platform database that resides in the Google Cloud Platform (in Sydney, New South Wales) and the encrypted backup database that resides in the Amazon Web Services platform (in Sydney, New South Wales). The researchers' access deidentified clinical trial data using straight-through digital processing methods such that a copy of clinical trial data (at any point in time) can be automatically moved from the platform database to a specified secure University of Sydney network server in one go.

This trial will be conducted in accordance with the Privacy Act of 1988 [31]. Upon consenting to participate in the clinical trial, participants will be assigned a unique identifier (Version 4 UUID), which is automatically generated by the platform and stored within its database. A Version 4 UUID is a universally unique identifier that is generated using random numbers. Importantly, the unique identifier will be used to identify a participant in all research studies of the InnoWell Platform should they be a consumer at more than 1 participating center. In addition, researchers will assign all participants with a Study ID number within deidentified clinical trial datasets. These Study ID numbers will be maintained separately from the data

collection process and used for statistical analysis purposes only. All data are backed up hourly in a secure, long-term storage on Google's Infrastructure.

Data Analysis

Scientific validity will be evaluated using interrater reliability statistical methods, such as Cohen kappa coefficient, to quantify the degree of agreement within and among the 4 ratings of stage-based stepped care (ie, platform-generated, health professional allocated, in-service multidisciplinary consensus, and independent expert-clinician reference group). The clinical performance, including aspects of safety, of the platform will be evaluated using aggregate data derived from clinical rating scales completed by the consumer and health professional to measure and track health domain intensity, frequency, quality, and change over time, and aggregate service-level performance data, including safety and clinical quality (eg, time to first assessment, service accessibility, wait time for clinical intervention, service efficiency, and user satisfaction). Within- and between-group analyses using multivariate statistics (eg, multivariate analyses of variance, Kruskal-Wallis test) will be computed to evaluate differences in clinical outcomes among services and among population groups. In addition, reliable change and effect-size scores will be calculated to determine clinical improvement over time for consumer data and then aggregated for service-level outcome data. Analytical validity will be assessed using trends only (ie, descriptive statistics) as a means to determine user engagement and overall effectiveness of the technical performance of the platform.

Ethics Approval

This trial has been approved by the Northern Sydney Local Health District Human Research Ethics Committee (NSLHD HREC), reference number HREC/17/HAWKE/480.

Figure 4. Hypothetical service-level performance data related to safety.

Notifications for Suicide thoughts and behaviour

Number of service notifications upon first questionnaire completion

16 Individuals   24.24% Percent of individuals

Number of service notifications over the time period filtered

18 Individuals   27.27% Percent of individuals

Average time for a clinician to respond to a notification

1 Days   19:28 Hours : Minutes

Results

InnoWell is a for-profit-business that has built the technology platform underpinning this clinical trial. InnoWell is a joint venture between The University of Sydney and PricewaterhouseCoopers (PwC) Australia that aims to transform mental health through person-centered care. Beyond June 2020, InnoWell will support this clinical trial financially, as it represents the core research and development work necessary to continuously and iteratively evolve the platform. The first results are expected to be submitted for publication in 2020.

Discussion

Person-Centered Care

At its core, the InnoWell Platform is designed to promote person-centered health care and assist clinical practices that place consumers as equal partners in health care decision making. The platform is also intended to help minimize the variability in care provision between health professionals and services by supporting clinical judgement with data. In addition, the platform aims to maximize the use of available resources and minimize duplication of services and wastage of time for all users, including consumers, health professionals, and service staff.

Feasibility of Implementation With Service

Before implementing this clinical trial, a thorough assessment of the feasibility of deploying the platform within each service (or participating center) is required [32,33]. This includes an understanding of the basic service attributes (eg, type of service, qualifications of health professionals, service location(s), and information and communication technology systems), as well as the identified needs of the organization that will be met by the platform. In other words, what problem(s) will the technology address? The fit of the InnoWell Platform to the values, priorities, and strategic plan for the service and the readiness of the service to adopt the technology also need to be examined. As the platform is customizable, it is necessary to identify in what way the technology needs to be modified or adapted to fit the service and its consumers. Finally, key to this process is the engagement of critical stakeholders who can drive the adoption of the platform, identify facilitators, mitigate barriers to implementation, and champion the technology among frontline staff whose practice will be most affected. The effective deployment of the platform will be supported by

implementation officers embedded within the services. These people will ensure that research protocols are adhered to and report any adjustments that might be necessary; they will monitor, catalogue, and report on the progress of the implementation to key trial leads; they will collect data at the participating center, relating to facilitators or barriers of implementation from clinical and administrative staff; they will assist with preparation of and delivery of training to relevant personnel regarding information essential to site-specific implementation; they will ensure that the platform accurately and consistently reflects the research data; they will ensure that all participants' safety and well-being are first priority, by liaising with technology specialists, the researchers, and the relevant participating center, and by following established ethical protocols.

Ongoing Development of Functionality

Qualitative and quantitative data will be gathered from all users of the platform, including consumers, as well as health professionals and service staff, to inform the iterative redevelopment of the platform, including site-specific customization of any new required functionality, as well as (re)configuration of content, questionnaires, and algorithms. Furthermore, the data collected during iterative user-testing sessions (covered under site-specific ethics approvals from relevant Human Research Ethics Committees) will help to monitor, evaluate, and provide ongoing feedback about the quality, acceptability, and usability of the platform for continuous improvement of the platform.

Future Research

This clinical trial will be run in parallel with a series of participating center-specific impact evaluation studies. These adjunctive studies will focus on gathering data through Web-based questionnaires, as well as workshops and semistructured interviews with service staff (ie, health professionals, managers, and administrators) relating to the impact of embedding the InnoWell Platform in their service. Topics to be covered will include digital readiness and competence, the impact of the platform on the service, the social return on investment, and the quality, usability, and acceptability of the platform. In addition, workshops will also offer the opportunity for researchers to introduce and provide brief training to service staff relating to new functionality that will be incorporated into the platform at quarterly intervals. Separate ethics approvals will be sought for these studies.

Authors' Contributions

Authors IH, TD, and SC were integral in securing funding to support this study. The study was designed by TD, HL, LW, and AE. HL prepared the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Conflicts of Interest

IH was an inaugural Commissioner on Australia's NMHC (2012-18). He is the codirector, Health and Policy, at the Brain and Mind Centre (BMC), The University of Sydney. The BMC operates an early-intervention youth services at Camperdown, under contract to headspace. IH has previously led community-based and pharmaceutical industry-supported (Wyeth, Eli Lily, Servier, Pfizer, and AstraZeneca) projects focused on the identification and better management of anxiety and depression. He was a member of the Medical Advisory Panel for Medibank Private until October 2017, a Board Member of Psychosis Australia Trust, and a member of Veterans Mental Health Clinical Reference group. He is the Chief Scientific Advisor to and an equity shareholder at Innowell. Innowell has been formed by The University of Sydney and PwC Australia to deliver the Aus $30 million Australian Government–funded "Project Synergy." Project Synergy is a 3-year program for the transformation of mental health services through the use of innovative technologies.

References

1. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
2. Anthes E. Nature. 2016. Pocket psychiatry: mobile mental-health apps have exploded onto the market, but few have been thoroughly tested URL: https://www.nature.com/polopoly_fs/1.19694!/menu/main/topColumns/topLeftColumn/pdf/532020a.pdf?origin=ppub [accessed 2019-05-19] [WebCite Cache ID 78VAj7mdv]
3. Cross SP, Hickie I. Transdiagnostic stepped care in mental health. Public Health Res Pract 2017 Apr 27;27(2):- [FREE Full text] [doi: 10.17061/phrp2721712] [Medline: 28474049]
4. Department of Health and Family Services. 1992. National Mental Health Strategy URL: http://www.health.gov.au/internet/publications/publishing.nsf/Content/mental-pubs-i-midrev2-toc~mental-pubs-i-midrev2-2~mental-pubs-i-midrev2-2-nat [accessed 2019-02-14] [WebCite Cache ID 76Bt3WMue]
5. The Department of Health. Canberra, Australia: Commonwealth of Australia; 1992. First National Mental Health Plan URL: http://www.health.gov.au/internet/publications/publishing.nsf/Content/mental-pubs-i-midrev2-toc~mental-pubs-i-midrev2-2~mental-pubs-i-midrev2-2-fir [accessed 2019-02-14] [WebCite Cache ID 76BsxSouW]
6. The Department of Health. Canberra, Australia: Commonwealth of Australia; 1998. Second National Mental Health Plan URL: http://www.health.gov.au/internet/publications/publishing.nsf/Content/mental-pubs-i-midrev2-toc~mental-pubs-i-midrev2-2~mental-pubs-i-midrev2-2-sec [accessed 2019-02-14] [WebCite Cache ID 76BsnPvoM]
7. The Department of Health. Canberra, Australia: Commonwealth of Australia; 2003. Third National Mental Health Plan 2003-2008 URL: http://www.health.gov.au/internet/publications/publishing.nsf/Content/mental-pubs-n-infopri2-toc~mental-pubs-n-infopri2-pt1~mental-pubs-n-infopri2-pt1-4 [accessed 2019-02-14] [WebCite Cache ID 76BrAaIo7]
8. Health Ministers of Australia. Department of Health. Canberra, Australia: Commonwealth of Australia; 2009. Fourth National Mental Health Plan - An agenda for collaborative government action in mental health 2009-2014 URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/9A5A0E8BDFC55D3BCA257BF0001C1B1C/$File/plan09v2.pdf [accessed 2019-02-14] [WebCite Cache ID 76Bpfzvod]
9. National Mental Health Strategy. Canberra, Australia: Commonwealth of Australia; 2017. The Fifth National Mental Health and Suicide Prevention Plan URL: https://apo.org.au/sites/default/files/resource-files/2017/10/apo-nid114356-1220416.pdf [accessed 2019-02-14] [WebCite Cache ID 76BqcACdw]
10. National Mental Health Commission. 2014. The National Review of Mental Health Programmes and Services Sydney URL: https://www.mentalhealthcommission.gov.au/our-reports/our-national-report-cards/2014-contributing-lives-review.aspx [accessed 2019-05-20] [WebCite Cache ID 78VCjDrgH]
11. Hickie IB. Time to implement national mental health reform. Med J Aust 2015 Jun 01;202(10):515-517. [Medline: 26021354]
12. World Health Organization, International Telecommunication Union. 2012. National eHealth strategy toolkit URL: https://apps.who.int/iris/handle/10665/75211 [accessed 2019-02-14] [WebCite Cache ID 76AIVl1wP]
13. Zwar N, Harris M, Griffiths R, Roland M, Dennis S, Davies G. Australian Primary Health Care Research Institute. Kensington, Australia: The University of New South Wales School of Public Health and Community Medicine; 2006. A systematic review of chronic disease management URL: https://openresearch-repository.anu.edu.au/bitstream/1885/119226/3/final_25_zwar_pdf_85791.pdf [accessed 2019-02-14] [WebCite Cache ID 76BtdjlU3]
14. Department of Health and Ageing, Australian Government. 2012. E-Mental Health Strategy for Australia URL: https://www.health.gov.au/internet/main/publishing.nsf/Content/7C7B0BFEB985D0EBCA257BF0001BB0A6/$File/emstrat.pdf [accessed 2019-02-14] [WebCite Cache ID 76BtklPu5]

XSL•FO

RenderX

15. Hickie IB, Scott J, McGorry PD. Clinical staging for mental disorders: a new development in diagnostic practice in mental health. Med J Aust 2013 May 20;198(9):461-462. [Medline: 23682872]

16. Cross SP, Hermens DF, Scott EM, Ottavio A, McGorry PD, Hickie IB. A clinical staging model for early intervention youth mental health services. Psychiatr Serv 2014 Jul;65(7):939-943. [doi: 10.1176/appi.ps.201300221] [Medline: 24828746]

17. International Initiative for Mental Health Leadership. Suicide Prevention Resource Center. Atlanta: Zero Suicide; 2016. Zero Suicide: International Declaration for Better Healthcare URL: https://zerosuicide.org/international-declaration/ [accessed 2019-02-14] [WebCite Cache ID 76Aab5fpH]

18. Iorfino F, Davenport TA, Ospina-Pinillos L, Hermens DF, Cross S, Burns J, et al. Using new and emerging technologies to identify and respond to suicidality among help-seeking young people: a cross-sectional study. J Med Internet Res 2017 Jul 12;19(7):e247 [FREE Full text] [doi: 10.2196/jmir.7897] [Medline: 28701290]

19. Australian Bureau of Statistics. Canberra: Commonweatlh of Australia; 2015. Causes of Death, Australia URL: https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/3303.0~2015~Main%20Features~Australia's%20leading%20causes%20of%20death,%202015~3 [accessed 2019-05-21] [WebCite Cache ID 78XymmnW5]

20. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a mental health eClinic to improve access to and quality of mental health care for young people: using participatory design as research methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]

21. van Spijker BA, Batterham PJ, Calear AL, Farrer L, Christensen H, Reynolds J, et al. The suicidal ideation attributes scale (SIDAS): community-based validation study of a new scale for the measurement of suicidal ideation. Suicide Life Threat Behav 2014 Aug;44(4):408-419. [doi: 10.1111/sltb.12084] [Medline: 24612048]

22. Posner K, Brown GK, Stanley B, Brent DA, Yershova KV, Oquendo MA, et al. The Columbia-suicide severity rating scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. Am J Psychiatry 2011 Dec;168(12):1266-1277 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111704] [Medline: 22193671]

23. Norman SB, Cissell SH, Means-Christensen AJ, Stein MB. Development and validation of an Overall Anxiety Severity And Impairment Scale (OASIS). Depress Anxiety 2006;23(4):245-249. [doi: 10.1002/da.20182] [Medline: 16688739]

24. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

25. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry 2003 Sep 01;54(5):573-583. [doi: 10.1016/S0006-3223(02)01866-8] [Medline: 12946886]

26. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

27. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontologist 1969;9(3):179-186. [doi: 10.1093/geront/9.3_Part_1.179] [Medline: 5349366]

28. Mahoney F, Barthel D. Functional Evaluation: the Barthel Index. Md State Med J 1965 Feb;14:61-65 [FREE Full text] [Medline: 14258950]

29. Prins A, Bovin MJ, Smolenski DJ, Marx BP, Kimerling R, Jenkins-Guarnieri MA, et al. The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): development and evaluation within a veteran primary care sample. J Gen Intern Med 2016 Oct;31(10):1206-1211 [FREE Full text] [doi: 10.1007/s11606-016-3703-5] [Medline: 27170304]

30. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD Checklist (PCL). Behav Res Ther 1996 Aug;34(8):669-673. [doi: 10.1016/0005-7967(96)00033-2] [Medline: 8870294]

31. Australian Government. Federal Register of Legislation. Privacy Act 1988 URL: https://www.oaic.gov.au/privacy-law/privacy-act/ [accessed 2019-02-14] [WebCite Cache ID 76AcdHTnh]

32. Meyers DC, Durlak JA, Wandersman A. The quality implementation framework: a synthesis of critical steps in the implementation process. Am J Community Psychol 2012 Dec;50(3-4):462-480. [doi: 10.1007/s10464-012-9522-x] [Medline: 22644083]

33. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]

Abbreviations

BMC: Brain and Mind Centre
eHealth: electronic health
NMHC: National Mental Health Commission
PwC: PricewaterhouseCoopers

XSL•FO
RenderX

Protocol

Quality of Life and Clinical Outcome After Traumatic Spleen Injury (SPLENIQ Study): Protocol for an Observational Retrospective and Prospective Cohort Study

Claudia PAM Raaijmakers[1,2], MD[‡]; Paul NM Lohle[2], MD, PhD; Paul Lodder[3,4], MSc; Jolanda de Vries[1,3,5], PhD

[1]ETZ Hospital (Elisabeth-TweeSteden Ziekenhuis), Trauma TopCare, Tilburg, Netherlands

[2]ETZ Hospital (Elisabeth-TweeSteden Ziekenhuis), Department of Radiology, Tilburg, Netherlands

[3]Tilburg University, Department of Medical and Clinical Psychology, Tilburg, Netherlands

[4]Tilburg University, Department of Methodology and Statistics, Tilburg, Netherlands

[5]ETZ Hospital (Elisabeth-TweeSteden Ziekenhuis), Department of Medical Psychology, Tilburg, Netherlands

[‡]SPLENIQ study group

Corresponding Author:
Claudia PAM Raaijmakers, MD
ETZ Hospital (Elisabeth-TweeSteden Ziekenhuis)
Trauma TopCare
PO Box 90151
Tilburg, 5000 LC
Netherlands
Phone: 31 132212065
Email: c.raaijmakers@etz.nl

Abstract

Background: Little is known about the effect of a splenic rupture on the quality of life (QOL) of patients, although the spleen is one of the most frequently injured organs in blunt abdominal trauma. It is essential to obtain more knowledge about QOL after traumatic spleen injury so that this can be taken into account when choosing treatment.

Objective: The primary objective of the SPLENic Injury and Quality of life (SPLENIQ) study is to determine QOL after treatment for traumatic spleen injury. The secondary objective is to investigate clinical and imaging outcome in relation to QOL.

Methods: A combination of a retrospective single-center and a prospective multicenter observational cohort study will be conducted. Patients in the retrospective study have had a splenic injury after blunt abdominal trauma and were admitted for treatment to the ETZ Hospital (Elisabeth-TweeSteden Ziekenhuis) in Tilburg between January 2005 and February 2017. Concerning the prospective cohort study, patients with splenic injury admitted to 1 of the 10 participating hospitals between March 2017 and December 2018 will be asked to participate. The follow-up period will be 1 year regarding QOL, clinical symptoms, and imaging. Patients in the retrospective study will complete 2 questionnaires: World Health Organization QOL assessment instrument-Bref (WHOQOL-Bref) and 12-Item Short-Form Health Survey (SF-12). Patients in the prospective study will complete 5 questionnaires at 1 week, 1 month, 3 months, 6 months, and 12 months after treatment: WHOQOL-Bref, SF-12, Euroqol 5-Dimensional 5-Level (EQ-5D-5L) questionnaire, Institute for Medical Technology Assessment (iMTA) Productivity Cost Questionnaire (iPCQ), and iMTA Medical Consumption Questionnaire (iMCQ). In both the retrospective and prospective study, patients treated with splenic artery embolization will undergo magnetic resonance imaging (MRI). The retrospective group will undergo MRI once, and the prospective group will undergo MRI 1 month and 1 year after treatment. Treatment of splenic injury depends on the severity of the splenic injury, the hemodynamic condition of the patient, and the hospital's or doctor's preference. This study is observational in nature without randomization. Concerning the retrospective data, multivariate analysis of covariance will be done. With regard to the prospective data, mixed linear modeling will be performed.

Results: This project was funded in April 2015 by ZonMw. The results of the retrospective study will be expected in March 2019. With regard to the prospective study, inclusion of patients was completed in December 2018 and data collection will be completed in December 2019. The first results will be expected in 2019.

Conclusions: To our knowledge, this is the first study that examines QOL in patients with a traumatic spleen injury. The SPLENIQ study responds to the shortage of information about QOL after treatment for traumatic spleen injury and may result in the development of a patient-oriented protocol.

XSL•FO
RenderX

Introduction

Background

Morbidity and mortality are the most commonly used outcome parameters in trauma care literature. However, most patients survive their trauma, and depending on the severity of the trauma, they will be limited in daily life, both physically and mentally. Although quality of life (QOL) is recognized as an important outcome measure, it is still a highly neglected aspect in trauma care studies, including studies with splenic injury patients. Multiple studies that did measure QOL have shown that severely injured patients suffer from psychological complaints and decreased QOL [1-3]. These factors have a major social and economic impact because trauma often involves young patients who frequently are unable to return to work, to reintegrate back into society, or to retrieve their previous activity level [4-11]. In case of splenic injury, treatment choices have to be made in which it is currently unknown what the effects for these patients will be in the short, medium, and long term. When more knowledge becomes available about QOL, it will be useful to determine the choice of treatment.

In blunt abdominal trauma, where the spleen is one of the most frequently injured organs, much is known about morbidity and mortality [12]. Internal bleeding caused by abdominal organ injury is one of the main causes of death after trauma, and a missed splenic rupture is the most common cause of preventable death in trauma patients [13,14]. Presently, the standard of care in hemodynamically stable patients is nonoperative management (NOM), involving close observation of the patient, with success rates up to 90% [12]. A recent study among adults with blunt splenic injury suggests that there are prognostic factors for failure of NOM. Strong evidence exists for prognostic patient factors such as age of 40 years or above, Injury Severity Score of at least 25, and American Association for the Surgery of Trauma splenic injury grade of 3 or higher [15]. Failure of therapy leads to more interventions, longer hospital stay, and higher mortality rates, resulting in increased costs and presumably decreased QOL. When NOM fails, angiography and splenic artery embolization (SAE) can be used as a supplement to NOM. The success rate of SAE ranges between 73% and 100%, with an overall success rate of NOM combined with SAE ranging between 86% and 100% (most studies reporting success rates greater than 90%) [16]. Despite this success rate, much remains unknown about splenic function after SAE, although it is speculated that there is a relationship between splenic volume and the immunologic status of the patient [17-20]. Preservation of splenic function might be one of the biggest advantages of NOM and SAE. However, patients treated with SAE have a risk of developing splenic infarction, abscesses, or cysts, with distal embolization having a significantly higher association with major complications compared with proximal embolization [21,22]. Surgery is indicated when a patient is hemodynamically unstable and does not respond to transfusion or when associated intraabdominal injuries require surgical management. Possible disadvantages of surgery are postsplenectomy complications, such as sepsis, thrombocytosis, and a lifetime risk of invasive infections (overwhelming postsplenectomy infection) [23,24]. All complications may have a major impact on patients' QOL.

Cost-effectiveness is important in the choice of treatment. Published study results on hemodynamically stable patients with splenic injury favored nonsurgical management over surgery with better clinical and cost results. SAE as a supplement to NOM trended toward being more cost-effective with a shorter hospital stay, despite comparable failure rates. Procedure-related costs were higher for surgery than for SAE, but total hospital costs were not significantly different [25].

Objectives

There is a growing demand for a (national) guideline or protocol for clinical decision making in traumatic spleen injury. Therefore, it is important to determine the optimal selection criteria for the appropriate management strategy. To achieve this, the entire process surrounding a patient with splenic injury must be considered. Even today, it remains unclear what the impact of QOL is on the entire process. Therefore, the primary objective of this project is to examine the QOL of patients after therapy (NOM, SAE, and surgery) for traumatic spleen injury using a retrospective and prospective group of patients. The secondary objective is to investigate the clinical outcome (eg, complications, reinterventions, and additional therapy), imaging outcome (diagnosis and magnetic resonance spleen imaging after SAE), and cost outcome (cost-effectiveness) and their relation to QOL.

Finally, the data and results acquired from this study may result in the development of a patient-oriented protocol for the management of traumatic spleen injury.

Methods

Study Design

A combination of a retrospective single-center and a prospective multicenter observational cohort study will be conducted, assessing the effects of NOM, SAE, and splenectomy in patients with splenic injury after blunt abdominal trauma.

Participants and Centers of Recruitment

Retrospective Study

The study population comprises patients who had a splenic injury after blunt abdominal trauma and were admitted for treatment in the ETZ Hospital (Elisabeth-TweeSteden Ziekenhuis) in Tilburg, the Netherlands, between January 2005 and February 2017. It concerns both men and women who were 18 years or older at the time of screening (February 2017).

Prospective Study

The prospective study will be performed in 10 selected Dutch hospitals containing experienced interventional radiologists and trauma surgeons qualified to perform SAE and splenectomy, respectively. The group includes 7 level-1 trauma centers (Erasmus MC, ETZ Hospital, Leiden University Medical Center, Amsterdam University Medical Center, Radboud University Medical Center, Hospital Medisch Spectrum Twente, and Isala Hospital) and 3 level-2 trauma centers (Maasstad Hospital Rotterdam, Albert Schweitzer Hospital Dordrecht, and Amphia Hospital Breda). The study will be organized in a network infrastructure in which radiologists and trauma surgeons collaborate.

All patients with a splenic injury after abdominal trauma confirmed by ultrasound/focused assessment with sonography in trauma (US/FAST) and/or computed tomography (CT) at the primary trauma screening at the 10 participating hospitals between March 2017 and December 2018 will be asked to participate.

Inclusion and Exclusion Criteria

Retrospective Study

The inclusion criteria are patients (1) diagnosed with splenic injury after trauma; (2) who underwent NOM, SAE, or surgery at the ETZ Hospital; (3) diagnosed between January 2005 and February 2017; and (4) who were aged 18 years or older at the time of screening (February 2017). Patients will be excluded from questionnaires in case of (1) insufficient knowledge of the Dutch language (verbal and writing) or (2) obviously, death. Patients treated with SAE will be excluded for magnetic resonance imaging (MRI) when they do not want to or are not able to undergo an MRI abdomen (eg, pregnancy or other contraindications).

Prospective Study

To be eligible to participate, patients (1) must be aged 18 years or older, (2) have splenic injury after abdominal trauma (confirmed by US/FAST and/or CT), and (3) must be treated in 1 of the 10 participating hospitals between March 2017 and December 2018. Exclusion criteria are identical to those of the retrospective study.

Sample Size Calculation

As the World Health Organization QOL assessment instrument-Bref (WHOQOL-Bref) comprises multiple facets, the *retrospective data* will be analyzed using a multivariate analysis of covariance (MANCOVA). This technique is more powerful than a univariate analysis of variance, resulting in a required sample size of 135 patients giving a medium effect and power of 0.80. Previous research [26] indicates that the minimal clinically important difference (MCID) of the WHOQOL-Bref slightly varies across its domains. Table 1 shows for each domain the MCID and standard deviation. On the basis of these statistics, the Cohen d effect sizes were calculated and transformed via Cohen f to the f^2 effect sizes required in the MANCOVA power analysis. It turns out that 138 participants are required to detect the average f^2 of 0.061 with a power of 0.80, given a significance level of .05.

With regard to the *prospective data*, mixed linear modeling will be used. Power analysis is performed for a repeated measures design investigating the interaction between treatment and time. We assumed a significance level of .05 and a medium effect size of partial eta squared of 0.05 and an average correlation of 0.50 among the 5 repeated measurements. On the basis of this, we require 33 participants to test with a power of 0.80, whether the 3 treatment groups differ in their change in QOL over the follow-up time. When assuming an effect size similar to the MCID's reported in the table above, 27 participants are required to detect the medium effect of f=0.248 with a power of 0.80.

Study Procedures

Retrospective Study

The clinical and imaging data of all splenic injury patients will be collected from the electronic patient files and registered anonymously. To measure QOL, all patients who are still alive will receive a written letter explaining the study, an informed consent form, the questionnaires, and a prestamped return envelope by mail. When a patient is willing to participate, he/she will sign the consent form and send it back in the return envelope, together with the completed questionnaires, assessing QOL and health status. When patients do not complete or return the questionnaires within 2 weeks, they receive a phone call as a reminder. When patients do not want to participate, the reason will be noted if the patient wants to let it be known. The patients who underwent SAE will be called for a voluntary single MRI abdomen at the ETZ Hospital in Tilburg.

This study has been reviewed and approved by the Medical Ethical Committee Brabant (METC Brabant, protocol number: NL54339.028.15) on January 27, 2016. The study has also been approved by the local ethical committee of the ETZ Hospital on February 9, 2016.

Table 1. Minimal clinically important difference (MCID) per WHOQOL-Bref domain.

Domain	MCID	SD	Cohen d	Cohen f	Effect size f^2
Physical	1.545	3.1	0.498	0.249	0.062
Psychological	1.259	2.5	0.504	0.252	0.063
Social	1.274	2.6	0.490	0.245	0.060
Environmental	1.142	2.3	0.497	0.248	0.062
General	0.876	1.8	0.487	0.243	0.059
Average	1.219	2.460	0.495	0.248	0.061

Prospective Study

In each participating hospital, an interventional radiologist and a trauma surgeon will be designated as principal investigators. Each hospital will also have a research assistant, most likely a radiology technologist or a member of the research team. The daily work will be carried out by the research assistant under the supervision of the principal investigator at the ETZ Hospital (CR). The research assistant will check daily whether potential eligible patients were admitted to the hospital. This will be done by checking the subscription list, verbally checking with the attending (resident) radiologist and trauma surgeon, and verifying the data in the electronic patient record. The subscription list will be placed at the dictation station of the radiologist, where the trauma diagnoses are reported. If a patient is treated for a traumatic spleen injury, baseline characteristics will be collected and he/she will be screened for the inclusion and exclusion criteria. The clinical data of all eligible patients will be collected anonymously in the database. As soon as the patient can talk and is lucid, the assistant will visit the patient to provide a verbal and written explanation about the study. The time for consideration of participation is 1 week. When a patient is willing to participate, he/she will sign a consent form. If not, the reason will be noted if the patient wants to let it be known. Each inclusion will be reported to the principal investigator (CR). Total follow-up time is 1 year after treatment with time points at 1 week and 1, 3, 6, and 12 months after treatment.

Patients will complete the questionnaires at all time points. The questionnaires of time point 1 will most likely be handed out to the patients at the hospital ward (intensive care, medium care, or general ward), usually by the research assistant of that hospital. At the other time points, patients will be sent an email to complete the questionnaires, assessing QOL, quality of care, health care consumption, and return back to work (if applicable). These questionnaires will be completed by the patient using a secure Web-based program: Data Management by Research Manager [27]. If a patient does not have internet, the questionnaires will be provided on paper with a prestamped return envelope. When patients do not complete or return the questionnaires within 2 weeks, they will receive a phone call as reminder. When patients no longer want to participate, the

reason will be noted if they want to let it be known. Patients treated with SAE will receive an MRI abdomen 1 month and 1 year after treatment.

This study has been reviewed and approved by METC Brabant (protocol number: prospective study NL54542.028.16) on May 18, 2016. The study has also been approved by the local ethical committees of the ETZ Hospital (6-13-2016), Maasstad Hospital (7-28-2016), Erasmus MC (9-21-2016), Albert Schweitzer Hospital (10-31-2016), Amphia Hospital (2-23-2017), Leiden University Medical Center (10-17-2017), Amsterdam University Medical Center (VUmc) (1-11-2018), Hospital Medisch Spectrum Twente (1-25-2018), Radboud University Medical Center (5-3-2018), and Isala Hospital (5-17-2018).

Data Collection

Data of the retrospective and prospective study will be recorded in a secure online database (Data Management by Research Manager, Health Solutions Deventer) [27]. In the prospective study, the database will be available to every research assistant. The program complies with the new legislation for collecting and processing personal data in medical scientific research: General Data Protection Regulation dated May 25, 2018.

Clinical Data

For both the retrospective and prospective study, data will be collected from the electronic patient records and trauma registry (Network Emergency Care Brabant). The trauma registry compiles prehospital and hospital data of all trauma patients admitted after presentation to the emergency department. Patients with splenic injury will be identified by the International Statistical Classification of Disease and Related Health Problems (ICD-10) and Abbreviated Injury Scale (AIS) diagnosis codes starting with S36.x and 5442, respectively.

The collected data concern the following: age, sex, systolic blood pressure, hemoglobin, Glasgow coma scale (at arrival), intubation (Yes/No), imaging (US and/or CT), grading spleen injury (American Association for the Surgery of Trauma; see Table 2)[28], type of treatment (NOM/SAE/splenectomy), complications, hospital stay (days), spleen in situ at discharge (Y/N), reinterventions (Y/N), rehospitalization (Y/N), and mortality.

Table 2. American Association for the Surgery of Trauma spleen injury scaling (1994 Revision).

Grade[a]		Injury description
I		
	Hematoma	Subcapsular <10% of surface area
	Laceration	Capsular tear <1 cm parenchymal depth
II		
	Hematoma	Subcapsular 10%-50% of surface area; or intraparenchymal <5 cm in diameter
	Laceration	1-3 cm parenchymal depth, which does not involve a trabecular vessel
III		
	Hematoma	Subcapsular >50% of surface area or expanding; or ruptured subcapsular or parenchymal hematoma; or intraparenchymal hematoma >5 cm or expanding
	Laceration	>3 cm parenchymal depth or involving trabecular vessels
IV		
	Laceration	Involving segmental or hilar vessels producing major devascularization (>25% of the spleen)
V		
	Laceration	Completely shattered spleen
	Vascular	Hilar vascular injury with devascularized spleen

[a]Advance one grade for multiple injuries, up to grade III.

Questionnaires

Patients in the retrospective study will complete the questionnaires once. Patients in the prospective study will complete questionnaires at 1 week and 1, 3, 6, and 12 months after treatment (see Table 3).

World Health Organization Quality of Life Assessment Instrument-Bref

QOL will be assessed with the World Health Organization Quality of Life Assessment Instrument-Bref (WHOQOL-Bref) [29]. This 26-item questionnaire is a short version of the WHOQOL-100, and it assesses 4 domains (physical health, psychological health, social relationships, and environment) as well as 1 general facet, *Overall QOL and General Health*. The questions in the domains are derived from the 24 facets of the WHOQOL-100, with 1 item from each of the facets. Each item is rated on a 5-point rating scale. Higher scores indicate a better QOL [29,30]. The WHOQOL-Bref has good psychometric properties [30-33].

12-Item Short-Form Health Survey

The 12-Item Short-Form Health Survey (SF-12) is a shorter version of the 36-Item Short-Form Health Survey (SF-36), which will be used for evaluating individual patients' health status, researching the cost-effectiveness of a treatment, and monitoring and comparing disease burden. The SF-12 covers 8 domains: physical functioning, role limitations because of physical problems, bodily pain, general health, vitality, social functioning, role limitations because of emotional problems, and mental health [34,35]. From these domains, summary scores for the physical component (PCS) and mental component (MCS) can be computed. The 12 items for the SF-12 were selected such that the SF-12 component scores explain 90% of the variability in PCS and MCS scores of the SF-36 [34]. The SF-36 was used

as a criterion for validation of the SF-12. The SF-12 and the SF-36 components and scales are scored with the algorithms specified by the developer [35]. The minimum possible score is 0 and the maximum possible score is 100. The SF-12 has good reliability and validity [36-42].

Euroqol 5-Dimensional 5-Level Questionnaire

The Euroqol 5-Dimensional 5-level questionnaire (EQ-5D-5L) is a generic health status instrument that measures health-related QOL [43]. The descriptive system of the instrument comprises 5 dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), which can be scored with 5 levels (ranging from no problems to severe problems). For the purpose of cost-effectiveness studies, health status is expressed in utilities, with a scale from 0 (death) to 1 (perfect health). The EQ-5D-5L can be used to derive utilities; the Dutch tariff can be used for this purpose [44]. Moreover, the EQ-5D-5L has good psychometric properties [45-50].

iMTA Productivity Cost Questionnaire

The impact of disease on the ability of a person to perform work should be part of an economic evaluation when a societal perspective is applied. The iMTA Productivity Cost Questionnaire (iPCQ) is a generic, nondisease-specific questionnaire, and it is applied in national and international studies [51]. The questionnaire is currently available in more than 10 languages, including Dutch. Both indirect cost because of absenteeism and the productivity losses because of presenteeism (ie, sick, but working) are taken into account. A manual is available, containing information on the modular structure of the iPCQ and its scoring and valuation methods that are used for cost calculations. By applying productivity costs, the answers of the iPCQ can be monetized and, as such, used in health economic evaluations.

Table 3. Timeline of the prospective study concerning measurements.

Time point	Timing and setting	Demographics	Questionnaires	Imaging
T1	After treatment, during admission	Demographic factors and clinical data	Quality of life (WHOQOL-Bref[a]); health status (SF-12[b]); and health-related QOL[c] (EQ-5D-5L[d])	—[e]
T2	1 month after treatment, at home (MRI[f]: outpatient clinic)	__[g]	Quality of life (WHOQOL-Bref); health status (SF-12); health-related QOL (EQ-5D-5L); productivity costs (iPCQ[h]); and medical consumption (iMCQ[i])	MRI (SAE[j] patients only)
T3	3 months after treatment, at home	__[g]	Quality of life (WHOQOL-Bref); health status (SF-12); health-related QOL (EQ-5D-5L); productivity costs (iPCQ); and medical consumption (iMCQ)	__[e]
T4	6 months after treatment, at home	__[g]	Quality of life (WHOQOL-Bref); health status (SF-12); health-related QOL (EQ-5D-5L); productivity costs (iPCQ); and medical consumption (iMCQ)	__[e]
T5	1 year after treatment, at home (MRI: outpatient clinic)	Clinical data	Quality of life (WHOQOL-Bref); health status (SF-12); health-related QOL (EQ-5D-5L); productivity costs (iPCQ); and medical consumption (iMCQ)	MRI (SAE patients only)

[a]WHOQOL-Bref: World Health Organization Quality of Life assessment instrument-Bref.

[b]SF-12: 12-Item Short-Form Health Survey.

[c]QOL: quality of life.

[d]EQ-5D-5L: Euroqol 5-Dimensional 5-Level questionnaire.

[e]At this time point, no MRIs have been completed.

[f]MRI: magnetic resonance imaging.

[g]At this time point, no demographic or clinical data have been collected.

[h]iPCQ: iMTA Productivity Cost Questionnaire.

[i]iMCQ: iMTA Medical Consumption Questionnaire.

[j]SAE: splenic artery embolization.

iMTA Medical Consumption Questionnaire

The iMTA Medical Consumption Questionnaire (iMCQ) is a generic, nondisease-specific instrument for measuring (direct) medical costs [52]. The instrument is a standardized self-reported questionnaire. The iMCQ includes questions related to frequently occurring contacts with health care providers and can be complemented with extra questions that are relevant for specific study populations. A manual is available for a structured use of the questionnaire. For the valuation of resource use, as obtained from the iMCQ, reference unit prices can be used. These reference prices can be derived from the Dutch manual for costing studies. The manual was commissioned and published by Zorginstituut Nederland and

authored by the Institute for Medical Technology Assessment (iMTA) [53].

MRI Abdomen After Splenic Artery Embolization

Patients in the retrospective study will undergo MRI once at the ETZ Hospital in Tilburg, and patients in the prospective study will undergo MRI 1 month and 1 year after treatment at the hospital where the treatment took place. Not all hospitals have the same MRI scanner. However, the same scan protocol will be used at all locations, leading to comparable images and assessments (see Table 4).

Only patients treated with SAE will receive an MRI of the upper abdomen to evaluate the spleen morphologically (volume, necrosis, splenosis, calcifications, and chronic infarction) and dynamically (diffusion and enhancement).

XSL•FO

RenderX

Table 4. Magnetic resonance imaging scan protocol.

Retrospective study	Prospective study
Axial Dual FFE[a]/GRE[b] (in-out phase); Coronal T2-weighted TSE[c]/FSE[d]; Axial T2-weighted TSE/FSE; Axial BFFE[e]/BGRE[f] Volume (5 mm slices, no gap); Axial DWI[g] with b value=0/400/800	Axial Dual FFE/GRE (in-out phase); Coronal T2-weighted TSE/FSE; Axial T2-weighted TSE/FSE; Axial BFFE/BGRE Volume (5 mm slices, no gap); Axial DWI with b value=0/400/800
3D[h] (noncontrast)	Axial dynamic T1-weighted noncontrast with fat sat; Axial T1-weighted 3D magnetic resonance angiography contrast-enhanced; Axial dynamic T1-weighted contrast-enhanced (2 time points) with fat sat

[a]FFE: Fast Field Echo.

[a]GRE: Gradient Echo.

[c]TSE: Turbo Spin Echo.

[d]FSE: Fast Spin Echo.

[e]BFFE: Balanced Fast Field Echo.

[f]BGRE: Balanced Gradient Echo.

[g]DWI: Diffusion Weighted Images.

[h]3D: three-dimensional.

Statistical Analysis

All analyses will be conducted using SPSS V24.0 (Statistical Package for Social Sciences, Chicago Illinois, USA). Frequencies and descriptive statistics will be calculated to provide an overview of the characteristics of the study population. Statistical test results will be considered significant at a level of $P<.05$.

Concerning the retrospective study, a MANCOVA will be done after correcting theoretically important covariates to assess the differences between treatment groups on the 4 WHOQOL-Bref domains and the general facet. For each type of treatment, a 1 sample t test will be performed for each WHOQOL-Bref scale to compare the QOL scores with reference data.

For both the retrospective and prospective studies, a logistic regression analysis will be performed on the outcome variables' (1) need for reintervention (yes/no) and (2) for each complication (yes/no), assessing the effect of treatment after correcting for theoretically important covariates. With regard to hospital stay in days, a Kruskal-Wallis test will be performed with the group (type of treatment) and days in hospital. Analysis of covariance will be used to compare proximal versus distal SAE, thereby correcting the effect for theoretically important covariates. To confirm/find prognostic factors for failure of NOM, a logistic regression analysis will be performed if NOM is a failure quickly after treatment. Otherwise, a survival analysis will be performed.

Regarding the prospective study, to assess the differences between groups in their change in QOL over time, a linear mixed model analysis will be conducted. To answer our research question, we will focus on the interaction effect between measurement occasion and treatment group, while correcting for theoretically important covariates. For the repeated measures, an unstructured covariance matrix will be used. Item-level missing values will be imputed according to the guidelines of the questionnaires. Scale-level missing values will be handled directly through maximum likelihood estimation, because the mixed model procedure makes use of all available data for each participant over all time points.

To conduct the cost-effectiveness analysis, a cost-effectiveness model will be developed. This model will comprise 2 treatment arms; SAE will be compared with splenectomy. The cost-effectiveness study will be conducted according to the most recent Dutch guidelines for health economic research [54]. As such, the study will be performed from the societal perspective, which means that all costs and benefits should be considered, regardless of by whom the costs are borne or to whom the benefits accrue.

Effects will be expressed in quality-adjusted life years, which constitute a combination of QOL and length of life. QOL will be measured in utilities. Utilities express QOL on a scale from 0 (death) to 1 (perfect health). Utilities can be derived from the EQ-5D-5L [44]. Survival will be derived from international clinical literature. Costs will be estimated according to the Dutch Manual for Costing in Economic Evaluations, update 2015, using a societal perspective [54]. A bottom-up methodology will be used to compute costs; the total number of medical contacts will be multiplied with unit costs. Direct medical costs comprise all costs directly relating to the prevention, diagnostics, therapy, rehabilitation, and care of the intervention. Health care utilization will be derived from the iMCQ [52]. The iPCQ will be used to assess productivity losses [51]. The friction cost method will be used for the calculation of costs because of production losses [55]. This method is the expertise of the iMTA; it is its standard method and has been widely used.

Regulation Statement

The study will be conducted according to the principles of the Declaration of Helsinki (7th revision, 64th World Medical Association General Assembly, Fortaleza, Brazil, October 2013) and in accordance with the Medical Research Involving Human Subjects Act (WMO).

Results

For the retrospective study, the data collection has taken place and the database is complete. Results will be reported in March 2019. Enrollment of participants in the prospective study began in March 2017, and it was completed in December 2018. Data

collection will be completed in December 2019. The first results will be reported in 2019.

Discussion

The SPLENic Injury and Quality of life (SPLENIQ) study is the first study that examines the effect of traumatic spleen injury on patients' QOL in both a retrospective and prospective observational study design. It also examines clinical and imaging outcome as well as cost-effectiveness. For several reasons, these studies add relevant information to the existing literature. First, there is a need for research into QOL after traumatic spleen injury, because this is an important but neglected factor in the trauma care literature. If more knowledge becomes available, this can be taken into account when choosing treatment. Second, the prospective study will be conducted in a multicenter context in 10 hospitals, involving a trauma surgeon and interventional radiologist in each hospital. This creates a strong collaboration between the participating hospitals and medical specialties, which hopefully adds to the inclusion rate. Third, it is still unknown what impact different SAE techniques and materials have on the morphology and volume of the spleen. To investigate this, using MRI is innovative and will provide interesting images containing important and necessary information.

Several factors related to the design and execution must be taken into account. First, patients will be treated with NOM, SAE, or surgery for a specific clinical condition. It may be the clinical situation that determines the long-term outcome, although that outcome is not or partially the result of the treatment. This risk is confounding by indication: the risk that the groups are in fact not easily comparable (ie, selection bias). To keep this to a minimum, the reason for choosing a particular treatment will be registered. In our analysis, we will correct for these confounding factors using propensity score analysis. Correction is only possible when adequate and good-quality information is available about the clinical condition of the patients, which led to the decision. The patient's record will be searched

thoroughly to find this information. Second, trauma patients often have multiple injuries that can affect QOL and clinical outcomes. We are not primarily interested in these additional injuries, but these will be included as a covariate in the analysis. Third, response bias may occur in the questionnaires group. Patients may decline participation because they are not interested or it may be too confronting to think/correspond about their psychological state. Fourth, the severely injured patients may be overrepresented in the nonresponse group concerning the questionnaires. To limit this, these patients will be visited as soon as they are approachable and, if necessary, will be provided assistance to complete the questionnaires. Fifth, the absence of randomization is a (strong) limitation and a potential source of bias, but randomized comparison in managing trauma patients is virtually impossible. Furthermore, it will have strong ethical implications as it is well known that randomizing trauma patients with intraabdominal bleeding, potentially unstable, is something not feasible in clinical practice. Sixth, the sample size calculation is based on the primary objective of the study. For the clinical and cost-effectiveness analysis, this implies that because of the small sample size, the uncertainty about the outcome is large. Seventh, we are aware of response shift bias. Response shift is a change in self-reported QOL that is a result of a change in internal standards (ie, recalibration), values (ie, reprioritization), or meaning of QOL (ie, reconceptualization) [56,57]. Thus, response shift reflects psychological adaptation: we do not consider this as a problem but as a fact of life. In addition, the method to prevent response shift contains recall bias itself. Considering this, and the fact that QOL is a generic outcome measure, ensures that we use the chosen method. Moreover, the retrospective patients will not be included in the prospective study.

In conclusion, the SPLENIQ study responds to the shortage of information about QOL after treatment for traumatic spleen injury. With developing a patient-oriented protocol, a necessary step is taken to customize standard care, which may contribute to a positive effect on QOL and clinical outcome.

XSL•FO

RenderX

Surgeon, Department of Surgery, Amsterdam University Medical Center (VUmc), Amsterdam; C van Dam, Research Assistant, Department of Radiology, Amsterdam University Medical Center (VUmc), Amsterdam; CAH Klazen, Radiologist, Department of Radiology, Hospital Medisch Spectrum Twente, Enschede; RJ de Wit, Trauma Surgeon, Department of Surgery, Hospital Medisch Spectrum Twente, Enschede; A Goolkate, Research Assistant, Department of Radiology, Hospital Medisch Spectrum Twente, Enschede; MJR Edwards, Trauma Surgeon, Department of Surgery, Radboud University Medical Center, Nijmegen; HM Dekker, Radiologist, Department of Radiology, Radboud University Medical Center, Nijmegen; T Tromp, Research Assistant, Department of Surgery, Radboud University Medical Center, Nijmegen; BAAM van Hasselt, Radiologist, Department of Radiology, Isala Hospital, Zwolle; SH van Helden, Trauma Surgeon, Department of Radiology, Isala Hospital, Zwolle; and L Buitenhuis, Research Assistant, Isala Hospital, Zwolle. The participating hospitals are the ETZ Hospital (Elisabeth-TweeSteden Ziekenhuis), Tilburg; Maasstad Hospital, Rotterdam; Erasmus MC, Rotterdam; Albert Schweitzer Hospital, Dordrecht; Amphia Hospital Breda; Leiden University Medical Center, Leiden; Amsterdam University Medical Center (VUmc), Amsterdam; Hospital Medisch Spectrum Twente, Enschede; Radboud University Medical Center, Nijmegen; and Isala Hospital, Zwolle. All hospitals are located in the Netherlands.

Authors' Contributions

CPAMR, JdV, PL, and PNML have contributed to the study conception, design, and funding. CPAMR coordinates the study, manages the study and data collection, and wrote the first drafts of the manuscript. JDV and PNML perform the general supervision of this study. All authors have read, commented on, and approved the final draft of this manuscript. All authors have agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Peer-reviewer report 1 from ZonMw.

[PDF File (Adobe PDF File), 59KB - resprot_v8i5e12391_app1.pdf]

Multimedia Appendix 2

Peer-reviewer report 2 from ZonMw.

[PDF File (Adobe PDF File), 59KB - resprot_v8i5e12391_app2.pdf]

Multimedia Appendix 3

Peer-reviewer report 3 from ZonMw.

[PDF File (Adobe PDF File), 58KB - resprot_v8i5e12391_app3.pdf]

References

1. Ringburg AN, Polinder S, van Ierland MC, Steyerberg EW, van Lieshout EM, Patka P, et al. Prevalence and prognostic factors of disability after major trauma. J Trauma 2011 Apr;70(4):916-922. [doi: 10.1097/TA.0b013e3181f6bce8] [Medline: 21045741]
2. Sluys K, Häggmark T, Iselius L. Outcome and quality of life 5 years after major trauma. J Trauma 2005 Jul;59(1):223-232. [doi: 10.1097/01.TA.0000173832.01547.CB] [Medline: 16096568]
3. van Delft-Schreurs CC, van Bergen JJ, van de Sande P, Verhofstad MH, de Vries J, de Jongh MA. A cross-sectional study of psychological complaints and quality of life in severely injured patients. Qual Life Res 2014 May;23(4):1353-1362. [doi: 10.1007/s11136-013-0546-y] [Medline: 24202637]
4. Polinder S, Meerding WJ, Mulder S, Petridou E, van Beeck E, EUROCOST Reference Group. Assessing the burden of injury in six European countries. Bull World Health Organ 2007 Jan;85(1):27-34 [FREE Full text] [Medline: 17242755]
5. Vles WJ, Steyerberg EW, Essink-Bot ML, van Beeck EF, Meeuwis JD, Leenen LP. Prevalence and determinants of disabilities and return to work after major trauma. J Trauma 2005 Jan;58(1):126-135. [Medline: 15674163]
6. Gabbe BJ, Simpson PM, Harrison JE, Lyons RA, Ameratunga S, Ponsford J, et al. Return to work and functional outcomes after major trauma: who recovers, when, and how well? Ann Surg 2016 Apr;263(4):623-632. [doi: 10.1097/SLA.0000000000001564] [Medline: 26779977]
7. Holtslag HR, Post MW, van der Werken W, Lindeman E. Return to work after major trauma. Clin Rehabil 2007 Apr;21(4):373-383. [doi: 10.1177/0269215507072084] [Medline: 17613579]

8. Holtslag HR, Post MW, Lindeman E, Van der Werken C. Long-term functional health status of severely injured patients. Injury 2007 Mar;38(3):280-289. [doi: 10.1016/j.injury.2006.10.026] [Medline: 17250834]

9. Alexandrescu R, O'Brien SJ, Lecky FE. A review of injury epidemiology in the UK and Europe: some methodological considerations in constructing rates. BMC Public Health 2009 Jul 10;9:226 [FREE Full text] [doi: 10.1186/1471-2458-9-226] [Medline: 19591670]

10. Gaines BA. Intra-abdominal solid organ injury in children: diagnosis and treatment. J Trauma 2009 Aug;67(2 Suppl):S135-S139. [doi: 10.1097/TA.0b013e3181adc17a] [Medline: 19667846]

11. Polinder S, Meerding WJ, van Baar ME, Toet H, Mulder S, van Beeck EF, EUROCOST Reference Group. Cost estimation of injury-related hospital admissions in 10 European countries. J Trauma 2005 Dec;59(6):1283-1290. [doi: 10.1097/01.ta.0000195998.11304.5b] [Medline: 16394898]

12. Schroeppel TJ, Croce MA. Diagnosis and management of blunt abdominal solid organ injury. Curr Opin Crit Care 2007 Aug;13(4):399-404. [doi: 10.1097/MCC.0b013e32825a6a32] [Medline: 17599009]

13. Velmahos GC, Spaniolas K, Tabbara M, Duggan M, Li Y, De Moya MM, et al. Abdominal insufflation decreases blood loss without worsening the inflammatory response: implications for prehospital control of internal bleeding. Am Surg 2008 Apr;74(4):297-301. [Medline: 18453291]

14. Esposito T, Gamelli R. Injury to the spleen. In: Mattox KL, Feliciano DV, Moore EE, editors. Trauma. 4th edition. New York: McGraw-Hill; 2000:683-713.

15. Olthof DC, Joosse P, van der Vlies CH, de Haan RJ, Goslings JC. Prognostic factors for failure of nonoperative management in adults with blunt splenic injury: a systematic review. J Trauma Acute Care Surg 2013 Feb;74(2):546-557. [doi: 10.1097/TA.0b013e31827d5e3a] [Medline: 23354249]

16. Haan JM, Bochicchio GV, Kramer N, Scalea TM. Nonoperative management of blunt splenic injury: a 5-year experience. J Trauma 2005 Mar;58(3):492-498. [doi: 10.1097/01.TA.0000154575.49388.74] [Medline: 15761342]

17. Preece SR, Schriber SM, Choudhury KR, Suhocki PV, Smith TP, Kim CY. Coil embolization of the splenic artery: impact on splenic volume. J Vasc Interv Radiol 2014 Jun;25(6):859-865. [doi: 10.1016/j.jvir.2013.12.564] [Medline: 24534094]

18. Palada I, Eterovic D, Obad A, Bakovic D, Valic Z, Ivancev V, et al. Spleen and cardiovascular function during short apneas in divers. J Appl Physiol (1985) 2007 Dec;103(6):1958-1963 [FREE Full text] [doi: 10.1152/japplphysiol.00182.2007] [Medline: 17947504]

19. Skattum J, Loekke RJ, Titze TL, Bechensteen AG, Aaberge IS, Osnes LT, et al. Preserved function after angioembolisation of splenic injury in children and adolescents: a case control study. Injury 2014 Jan;45(1):156-159. [doi: 10.1016/j.injury.2012.10.036] [Medline: 23246563]

20. Lammers AJ, de Porto AP, Bennink RJ, van Leeuwen EM, Biemond BJ, Goslings JC, et al. Hyposplenism: comparison of different methods for determining splenic function. Am J Hematol 2012 May;87(5):484-489 [FREE Full text] [doi: 10.1002/ajh.23154] [Medline: 22488175]

21. Rosati C, Ata A, Siskin GP, Megna D, Bonville DJ, Stain SC. Management of splenic trauma: a single institution's 8-year experience. Am J Surg 2015 Feb;209(2):308-314. [doi: 10.1016/j.amjsurg.2014.06.034] [Medline: 25457232]

22. Ekeh AP, Khalaf S, Ilyas S, Kauffman S, Walusimbi M, McCarthy MC. Complications arising from splenic artery embolization: a review of an 11-year experience. Am J Surg 2013 Mar;205(3):250-254. [doi: 10.1016/j.amjsurg.2013.01.003] [Medline: 23375704]

23. Raikhlin A, Baerlocher MO, Asch MR, Myers A. Imaging and transcatheter arterial embolization for traumatic splenic injuries: review of the literature. Can J Surg 2008 Dec;51(6):464-472 [FREE Full text] [Medline: 19057735]

24. Bessoud B, Duchosal MA, Siegrist C, Schlegel S, Doenz F, Calmes J, et al. Proximal splenic artery embolization for blunt splenic injury: clinical, immunologic, and ultrasound-Doppler follow-up. J Trauma 2007 Jun;62(6):1481-1486. [doi: 10.1097/TA.0b013e318047dfb8] [Medline: 17563670]

25. Bruce PJ, Helmer SD, Harrison PB, Sirico T, Haan JM. Nonsurgical management of blunt splenic injury: is it cost effective? Am J Surg 2011 Dec;202(6):810-815. [doi: 10.1016/j.amjsurg.2011.06.041] [Medline: 22137139]

26. de Mol M, Visser S, Aerts JG, Lodder P, de Vries J, den Oudsten BL. Satisfactory results of a psychometric analysis and calculation of minimal clinically important differences of the World Health Organization quality of life-BREF questionnaire in an observational cohort study with lung cancer and mesothelioma patients. BMC Cancer 2018 Nov 26;18(1):1173 [FREE Full text] [doi: 10.1186/s12885-018-4793-8] [Medline: 30477456]

27. Research Manager, Health Solutions. Data Management URL: https://my-datamanagement.com/ [accessed 2018-08-08] [WebCite Cache ID 71WKbbKiC]

28. Moore EE, Cogbill TH, Jurkovich GJ, Shackford SR, Malangoni MA, Champion HR. Organ injury scaling: spleen and liver (1994 revision). J Trauma 1995 Mar;38(3):323-324. [Medline: 7897707]

29. WHOQoL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. The WHOQOL Group. Psychol Med 1998 May;28(3):551-558. [Medline: 9626712]

30. The World Health Organization Quality of Life Group. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. Soc Sci Med 1998 Jun;46(12):1569-1585. [Medline: 9672396]

XSL•FO

RenderX

31. Van Esch L, Den Oudsten BL, De Vries J. International Journal of Clinical and Health Psychology(1). 2011. The world health organization quality of life instrument-short form (WHOQOL-BREF) in women with breast problems URL: http://www.aepc.es/ijchp/articulos_pdf/ijchp-367.pdf [accessed 2018-08-23] [WebCite Cache ID 71soFT31i]

32. Skevington SM, Lotfy M, O'Connell KA. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004 Mar;13(2):299-310. [Medline: 15085902]

33. Taylor WJ, Myers J, Simpson RT, McPherson KM, Weatherall M. Quality of life of people with rheumatoid arthritis as measured by the World Health Organization Quality of Life Instrument, short form (WHOQOL-BREF): score distributions and psychometric properties. Arthritis Rheum 2004 Jun 15;51(3):350-357 [FREE Full text] [doi: 10.1002/art.20398] [Medline: 15188318]

34. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [Medline: 8628042]

35. Ware JE, Kosinski M, Keller SD. SF-12: How to score the SF-12 physical and mental health summary scales. 2nd edition. Boston, MA: The Health Insitute, New England Medical Center; 1995. URL: https://www.researchgate.net/publication/242636950_SF-12_How_to_Score_the_SF-12_Physical_and_Mental_Health_Summary_Scales [accessed 2018-08-23] [WebCite Cache ID 71soLVWQm]

36. Cheak-Zamora NC, Wyrwich KW, McBride TD. Reliability and validity of the SF-12v2 in the medical expenditure panel survey. Qual Life Res 2009 Aug;18(6):727-735. [doi: 10.1007/s11136-009-9483-1] [Medline: 19424821]

37. Gandhi SK, Salmon JW, Zhao SZ, Lambert BL, Gore PR, Conrad K. Psychometric evaluation of the 12-item short-form health survey (SF-12) in osteoarthritis and rheumatoid arthritis clinical trials. Clin Ther 2001 Jul;23(7):1080-1098. [Medline: 11519772]

38. Hayes CJ, Bhandari NR, Kathe N, Payakachat N. Reliability and validity of the medical outcomes study short form-12 version 2 (SF-12v2) in adults with non-cancer pain. Healthcare (Basel) 2017 Apr 26;5(2) [FREE Full text] [doi: 10.3390/healthcare5020022] [Medline: 28445438]

39. Shou J, Ren L, Wang H, Yan F, Cao X, Wang H, et al. Reliability and validity of 12-item Short-Form health survey (SF-12) for the health status of Chinese community elderly population in Xujiahui district of Shanghai. Aging Clin Exp Res 2016 Apr;28(2):339-346. [doi: 10.1007/s40520-015-0401-9] [Medline: 26142623]

40. Pakpour AH, Nourozi S, Molsted S, Harrison AP, Nourozi K, Fridlund B. Validity and reliability of short form-12 questionnaire in Iranian hemodialysis patients. Iran J Kidney Dis 2011 Jul;5(3):175-181 [FREE Full text] [Medline: 21525577]

41. King JT, Horowitz MB, Kassam AB, Yonas H, Roberts MS. The short form-12 and the measurement of health status in patients with cerebral aneurysms: performance, validity, and reliability. J Neurosurg 2005 Mar;102(3):489-494. [doi: 10.3171/jns.2005.102.3.0489] [Medline: 15796384]

42. Luo X, George ML, Kakouras I, Edwards CL, Pietrobon R, Richardson W, et al. Reliability, validity, and responsiveness of the short form 12-item survey (SF-12) in patients with back pain. Spine (Phila Pa 1976) 2003 Aug 01;28(15):1739-1745. [doi: 10.1097/01.BRS.0000083169.58671.96] [Medline: 12897502]

43. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]

44. M Versteegh M, M Vermeulen K, M A A Evers S, de Wit GA, Prenger R, A Stolk E. Dutch tariff for the five-level version of EQ-5D. Value Health 2016 Dec;19(4):343-352 [FREE Full text] [doi: 10.1016/j.jval.2016.01.003] [Medline: 27325326]

45. Bilbao A, García-Pérez L, Arenaza JC, García I, Ariza-Cardiel G, Trujillo-Martín E, et al. Psychometric properties of the EQ-5D-5L in patients with hip or knee osteoarthritis: reliability, validity and responsiveness. Qual Life Res 2018 Jul 05;27(11):2897-2908. [doi: 10.1007/s11136-018-1929-x] [Medline: 29978346]

46. Cheung PW, Wong CK, Samartzis D, Luk KD, Lam CL, Cheung KM, et al. Psychometric validation of the EuroQoL 5-Dimension 5-Level (EQ-5D-5L) in Chinese patients with adolescent idiopathic scoliosis. Scoliosis Spinal Disord 2016;11:19 [FREE Full text] [doi: 10.1186/s13013-016-0083-x] [Medline: 27525314]

47. Marti C, Hensler S, Herren DB, Niedermann K, Marks M. Measurement properties of the EuroQoL EQ-5D-5L to assess quality of life in patients undergoing carpal tunnel release. J Hand Surg Eur Vol 2016 Nov;41(9):957-962. [doi: 10.1177/1753193416659404] [Medline: 27435748]

48. Nolan CM, Longworth L, Lord J, Canavan JL, Jones SE, Kon SSC, et al. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. Thorax 2016 Dec;71(6):493-500 [FREE Full text] [doi: 10.1136/thoraxjnl-2015-207782] [Medline: 27030578]

49. Alvarado-Bolaños A, Cervantes-Arriaga A, Rodríguez-Violante M, Llorens-Arenas R, Calderón-Fajardo H, Millán-Cepeda R, et al. Convergent validation of EQ-5D-5L in patients with Parkinson's disease. J Neurol Sci 2015 Nov 15;358(1-2):53-57. [doi: 10.1016/j.jns.2015.08.010] [Medline: 26276513]

50. Sakthong P, Sonsa-Ardjit N, Sukarnjanaset P, Munpan W. Psychometric properties of the EQ-5D-5L in Thai patients with chronic diseases. Qual Life Res 2015 Dec;24(12):3015-3022. [doi: 10.1007/s11136-015-1038-z] [Medline: 26048348]

51. Bouwmans C, Krol M, Severens H, Koopmanschap M, Brouwer W, Hakkaart-van Roijen L. The iMTA Productivity Cost Questionnaire: a standardized instrument for measuring and valuing health-related productivity losses. Value Health 2015 Sep;18(6):753-758 [FREE Full text] [doi: 10.1016/j.jval.2015.05.009] [Medline: 26409601]

52. Bouwmans C, Hakkaart-van Roijen L, Koopmanschap M, Krol M, Severens H, Brouwer W. User manual iMTA medical consumption questionnaire. Rotterdam: Institute for Medical Technology Assessment, Erasmus University Rotterdam, The Netherlands; 2013.

53. Tan SS, Bouwmans CA, Rutten FF, Hakkaart-van Roijen L. Update of the Dutch Manual for Costing in Economic Evaluations. Int J Technol Assess Health Care 2012 Apr;28(2):152-158. [doi: 10.1017/S0266462312000062] [Medline: 22559757]

54. [Netherlands Healthcare Institute]. Diemen, The Netherlands: Zorginstituut Nederland; 2015. Guideline for conducting economic evaluations in health care URL: https://english.zorginstituutnederland.nl/publications/reports/2016/06/16/guideline-for-economic-evaluations-in-healthcare [accessed 2019-04-23] [WebCite Cache ID 77qVDy6xL]

55. Koopmanschap MA, Rutten FF, van Ineveld IBM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ 1995 Jun;14(2):171-189. [Medline: 10154656]

56. Schwartz CE, Sprangers MA. Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. Soc Sci Med 1999 Jun;48(11):1531-1548. [doi: 10.1016/S0277-9536(99)00047-7] [Medline: 10400255]

57. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med 1999 Jun;48(11):1507-1515. [doi: 10.1016/S0277-9536(99)00045-3] [Medline: 10400253]

Abbreviations

CT: computed tomography
EQ-5D-5L: Euroqol 5-Dimensional 5-Level questionnaire
ETZ: Elisabeth-TweeSteden Ziekenhuis
FAST: focused assessment with sonography in trauma
iMCQ: iMTA Medical Consumption Questionnaire
iMTA: Institute for Medical Technology Assessment
iPCQ: iMTA Productivity Cost Questionnaire
MANCOVA: multivariate analysis of covariance
MCID: minimal clinically important difference
MCS: summary scores for the mental component (SF-12)
METC: Medical Ethical Committee
MRI: magnetic resonance imaging
NOM: nonoperative management
PCS: summary scores for the physical component (SF-12)
QOL: quality of life
SAE: splenic artery embolization
SF-12: 12-Item Short-Form Health Survey
SF-36: 36-Item Short-Form Health Survey
SPLENIQ: SPLENic Injury and Quality of life
US: ultrasound
WHOQOL-Bref: World Health Organization Quality of Life assessment instrument-Bref

Protocol

The Adverse Drug Reactions From Patient Reports in Social Media Project: Protocol for an Evaluation Against a Gold Standard

Armelle Arnoux-Guenegou[1], MSc, PhD; Yannick Girardeau[1,2], MSc, MD; Xiaoyi Chen[1], MSc, PhD; Myrtille Deldossi[3], MSc; Rim Aboukhamis[4], M Pharm; Carole Faviez[5], MSc; Badisse Dahamna[6], MSc; Pierre Karapetiantz[1], MSc; Sylvie Guillemin-Lanne[3], MSc; Agnès Lillo-Le Louët[4], MD, Prof Dr; Nathalie Texier[5], M Pharm; Anita Burgun[1,2,7], MD, PhD; Sandrine Katsahian[1,7,8,9], MD, PhD, Prof Dr

[1]INSERM U1138 - Team 22, Information Sciences to Support Personalized Medicine, Centre de Recherche des Cordeliers, Paris, France

[2]Département d'Informatique Médicale, Hôpital Européen Georges-Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

[3]Innovative Projects - Text Mining, Expert System, Paris, France

[4]Centre Régional de Pharmacovigilance, Hôpital Européen Georges-Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

[5]Kappa Santé, Paris, France

[6]Service d'Informatique Biomédicale, D2IM, Centre Hospitalier Universitaire de Rouen, Rouen, France

[7]INSERM U1138 - Team 22, Information Sciences to Support Personalized Medicine, Paris Descartes University, Sorbonne Paris Cité, Paris, France

[8]Clinical Research Unit Hôpitaux Universitaires Paris Ouest, Hôpital Européen Georges-Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

[9]INSERM CIC1418, Clinical Epidemiology, Hôpital Européen Georges-Pompidou, Paris, France

Corresponding Author:
Xiaoyi Chen, MSc, PhD
INSERM U1138 - Team 22
Information Sciences to Support Personalized Medicine
Centre de Recherche des Cordeliers
15 Rue de l'école de Médecine
Esc D, 1e étage
Paris, 75006
France
Phone: 33 71196369
Email: chenxiaoyi619@gmail.com

Abstract

Background: Social media is a potential source of information on postmarketing drug safety surveillance that still remains unexploited nowadays. Information technology solutions aiming at extracting adverse reactions (ADRs) from posts on health forums require a rigorous evaluation methodology if their results are to be used to make decisions. First, a gold standard, consisting of manual annotations of the ADR by human experts from the corpus extracted from social media, must be implemented and its quality must be assessed. Second, as for clinical research protocols, the sample size must rely on statistical arguments. Finally, the extraction methods must target the relation between the drug and the disease (which might be either treated or caused by the drug) rather than simple co-occurrences in the posts.

Objective: We propose a standardized protocol for the evaluation of a software extracting ADRs from the messages on health forums. The study is conducted as part of the Adverse Drug Reactions from Patient Reports in Social Media project.

Methods: Messages from French health forums were extracted. Entity recognition was based on *Racine Pharma* lexicon for drugs and Medical Dictionary for Regulatory Activities terminology for potential adverse events (AEs). Natural language processing–based techniques automated the ADR information extraction (relation between the drug and AE entities). The corpus of evaluation was a random sample of the messages containing drugs and/or AE concepts corresponding to recent pharmacovigilance alerts. A total of 2 persons experienced in medical terminology manually annotated the corpus, thus creating the gold standard, according to an annotator guideline. We will evaluate our tool against the gold standard with recall, precision, and f-measure. Interannotator agreement, reflecting gold standard quality, will be evaluated with hierarchical kappa. Granularities in the terminologies will be further explored.

XSL·FO
RenderX

Results: Necessary and sufficient sample size was calculated to ensure statistical confidence in the assessed results. As we expected a global recall of 0.5, we needed at least 384 identified ADR concepts to obtain a 95% CI with a total width of 0.10 around 0.5. The automated ADR information extraction in the corpus for evaluation is already finished. The 2 annotators already completed the annotation process. The analysis of the performance of the ADR information extraction module as compared with gold standard is ongoing.

Conclusions: This protocol is based on the standardized statistical methods from clinical research to create the corpus, thus ensuring the necessary statistical power of the assessed results. Such evaluation methodology is required to make the ADR information extraction software useful for postmarketing drug safety surveillance.

International Registered Report Identifier (IRRID): RR1-10.2196/11448

Introduction

Background

The detection of new adverse drug reactions (ADRs) has been based on postmarketing surveillance by government agencies (national authorities) derived from spontaneous reporting by health care professionals and patients. The US Food and Drug Administration (FDA) and the European Medicines Agency collect ADR case reports through the FDA's Adverse Event Reporting System (FAERS) [1-4] and the EudraVigilance system, respectively. These systems are useful tools for drug agencies, which mine these huge amounts of structured data to look for new safety concerns that might be related to a marketed product [5-8].

In the last 20 years, internet and social media have become an integral part of people's daily life. Social media is now often used to communicate with other persons having the same health concerns and share information regarding their health conditions, feelings, medications, and many other aspects [9]. Social media is, therefore, a potential provider of information on ADRs. In 2005, the International Society of Drug Bulletins already recognized such a use: *Patient reporting systems should periodically sample and evaluate the scattered drug experiences patients report on the internet* [10]. This new source of knowledge captured the interest from health informatics, statistics, and public health researchers. Although in its infancy, related scientific literature increased in the last decade [4,11-13].

Objectives

In this context, the French Ministry of Industry funded and launched the Adverse Drug Reactions from Patient Reports in Social Media (ADR-PRISM) project. The objective of ADR-PRISM was to make available the contents about ADR, informal and embedded in forums and discussions on the Web, to the actors involved in drug safety (drug companies, agencies,

and pharmacovigilance experts). In the end, the tool developed in the ADR-PRISM project should generate hypotheses about new or poorly documented adverse events (AEs). To reach its goals, the ADR-PRISM consortium gathers a company developing text mining softwares (Expert System), a company specialized in pharmaco-epidemiology (Kappa Santé), 3 academic research groups providing expertise in medical informatics and statistics (National Institute of Health and Medical Research & Cordeliers Research Centre: umrs 1138 team 22 dedicated to Information Sciences to support Personalized Medicine and Laboratory in Medical Informatics and Knowledge Engineering in e-Health, and Biomedicine informatics, Service Catalogue and Index of French Language medical websites SIBM-CISMeF), 2 experts in pharmacovigilance (regional center of pharmacovigilance), as well as Vidal group that supplies the drug database used in most drug prescription systems in France.

From a natural language processing (NLP) perspective, we considered ADR as relationships between drug and AE concepts. On the basis of that, the NLP-based Skill Cartridge for pharmacovigilance developed by Expert System for ADR-PRISM includes a relation extraction module based on (named) entity recognition combined with rules and regular expressions. Before applying it to large collections of forum discussions, we designed a protocol to assess the performance of this ADR information extraction module. The objective of this study is to present this protocol.

Methods

Synopsis

With the objective of sharing a methodology that guarantees the confidence in the results in ADR-PRISM, we followed a way of reasoning inspired by the standards widely adopted in clinical research such as the Standards for Reporting Diagnostic Accuracy studies [14]. These items are displayed in Textbox 1.

Textbox 1. Synopsis of the protocol.

Rationale

- Nowadays, patients extensively use social media. They report on the adverse events they feel because of their medications (further called adverse drug reaction [ADR]) on health forums. Promising studies exist on the extraction of ADR information from social media with natural language processing (NLP) or machine learning tools.

- The consortium Adverse Drug Reaction from Patient Reports in Social Media (ADR-PRISM) has been constituted to create a tool extracting the ADR information from social media. Teams specialized in text mining NLP and pharmacovigilance participate in the consortium.

- Before applying the ADR-PRISM's tool on a larger scale (millions of posts) to draw conclusions on ADRs, our goal was to adapt the principles adopted in clinical research to assess the ADR-PRISM. For example, evaluation was done against a gold standard based on manual annotation of the posts.

Primary objective

- To estimate robustly the performance of the ADR information extraction tool against gold standard.

Primary expected results

- Precision, recall, and f-measure for ADR extraction.

Secondary objectives

- To verify the quality of the gold standard and to evaluate the impact of various conditions (eg, different granularities and sentence constraint of the tool) on the performance of the ADR information extraction tool.

Secondary expected results

- Kappa metrics for interannotator agreement; and precision, recall, and f-measure for ADR extraction in various conditions.

Database

- Posts from the Kappa Santé Detec't database published between January 1, 2007, and October 28, 2016.

Eligibility criteria

- Posts from the Kappa Santé Detec't database that contain at least one drug's or molecule's (active substance) name, and posts randomly selected from the rest of the Kappa Santé Detec't database; a list of drugs and adverse events (AEs) of interest is established by drug safety experts; and ADR: any explicit and positive relationship between a drug and a potential AE where either the drug or the AE or both belong to the list.

Index test method

- ADR information extraction tool: this tool classifies each co-occurrence of drug and AE in a post as positive ADR or negative ADR or no ADR, and maps the drug and the AE expression to *Racine Pharma* and Medical Dictionary for Regulatory Activities (MedDRA), respectively.

Reference method

- Gold standard: manual annotations of the co-occurrence of drug and AE as positive ADR or negative ADR or no ADR including the mapping of the drug and the AE expression with a *Racine Pharma* and a MedDRA entity, respectively. Manual annotations will be provided by 2 annotators with experience in medical terminology.

Sample size

- A 95% CI, with a total width of 0.1, is used to determine the sample size.

Statistical analyses

- Recall, precision, and f-measure calculations for evaluation of the performances of the ADR extraction information tool and interannotator agreement for the evaluation of the gold standard with a Cohen kappa.

Design and Ethics

Design

This project is based on retrospective data collected among threads of discussion accessible on social media (messages from French health forums). A total of 2 experts in pharmacovigilance helped delineating the project. The objective is twofold: (1) focus on certain pharmaceutical products of interest and identify the related ADRs and (2) detect the emergence of general potential problem in public health. On this basis, the extraction of ADR information is expected to (1) perform well on specific concepts for use case study and (2) be able to extract all potential concepts correctly for screening purpose.

To evaluate the ADR information extraction, we therefore build up the global process in 2 phases (Figure 1). In phase 1, we implemented an iterative process of validation and improvement

XSL•FO

RenderX

of the Skill Cartridge for pharmacovigilance. The objectives were (1) to correct the most frequent errors and to train the ADR information extraction module and (2) to estimate the performance indicators and the time required for manual annotation for phase 2 sample size determination. Then, in phase 2, we will conduct a definite assessment of the performance of the ADR information extraction module.

In the rest of the paper, we use the term *AE* for medical events that are present in text for a potential ADR, whereas the term *ADR* is used when a relation between a drug and an AE is established.

Figure 1. Study overview. ADR: adverse drug reaction.
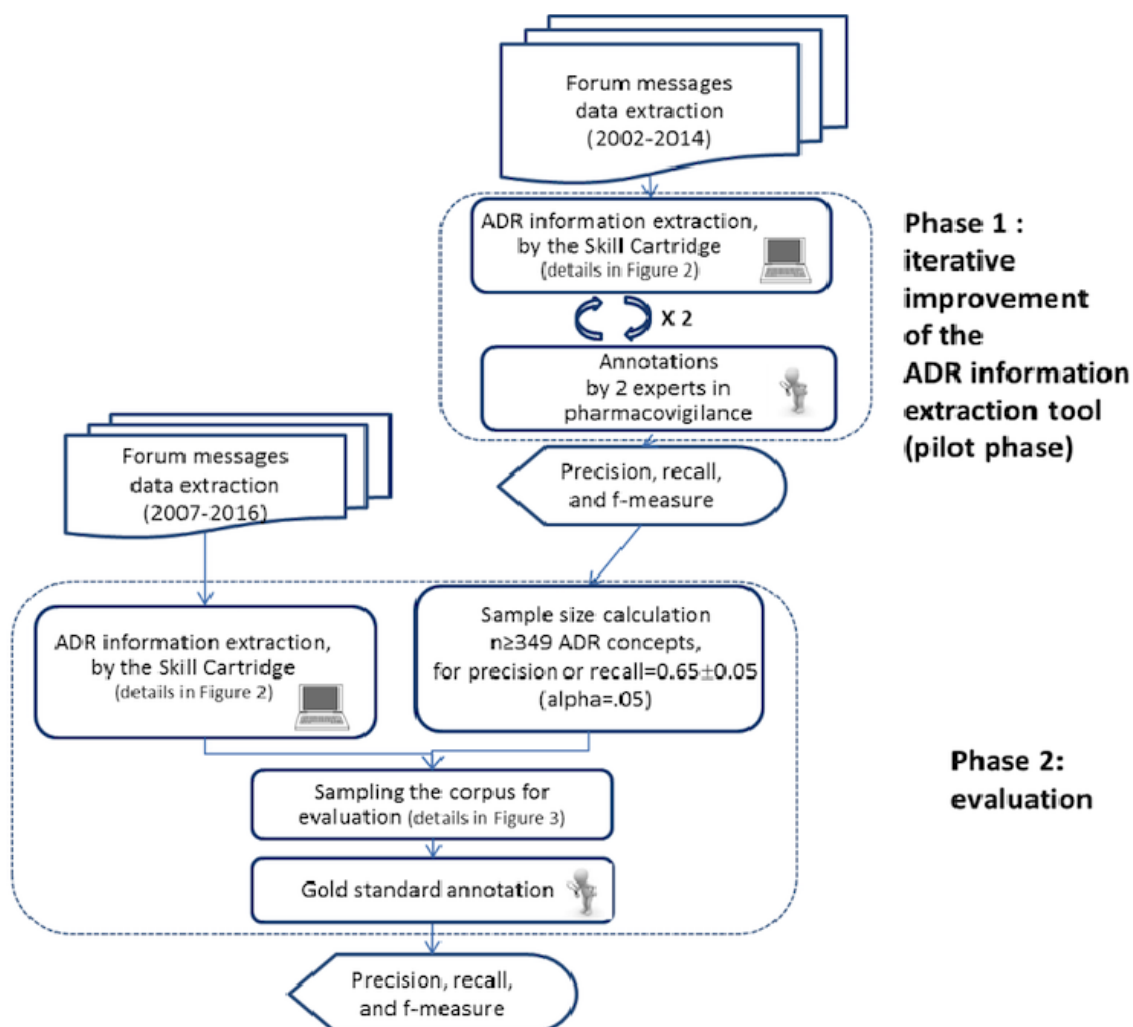
Ethics

This research did not involve experiment on either humans or animals. Ethics and guarantee of data privacy constituted an integral and dedicated working group set up for the ADR-PRISM project. To comply with national regulations, we first registered data collection to the French Data Protection Agency (CNIL for Commission Nationale Informatique et Liberté), which is known as a *normal notification* in technical terminology, on December 23, 2015. We later submitted an authorization request on March 30, 2016, regarding data analysis and validation of the approach adopted about ethics and confidentiality to the same agency. ADR-PRISM's consortium detailed approach is explained in the study by Bousquet et al [15].

The ADR-PRISM project was supported by an ethics advisory board, which was composed of scientists with different scientific expertise: drug safety, public health, and ethics. Their role was to give independent advice regarding ethical issues to the project consortium. The board approved the whole study design, including the protocol presented in the paper.

Adverse Drug Reactions From Patient Reports in Social Media Adverse Drug Reaction Extraction

Resources

We used 3 lexicons to represent drug information and AEs. We codified AEs with the Medical Dictionary for Regulatory Activities (MedDRA) v15.1 [16] classification. *Racine Pharma*, maintained by the SIBM-CISMeF [17], provided an extensive source of drug names that covered all medications available on the French market, including brand names and active ingredients. *Racine Pharma* entries were mapped to the Anatomical Therapeutic Chemical (ATC) system [18], which was used as a classification system for drugs.

The corpus of messages was extracted from the Detec't database [19], a database developed by Kappa Santé´ that collects messages from several French forums using a Web crawler. We limited ourselves to a list of 5 major health forums in French obtained using netscoring [9]. Message extraction was based on a named entity recognition module using a drug lexicon designed by Kappa Santé´ and a fuzzy matching algorithm. The lexicon was based on *Racine Pharma*, the ATC classification,

and a list of medications extracted from the French National Health Insurance database.

Drugs, Adverse Events, and Adverse Drug Reaction Extraction

ADR extraction was performed in 2 steps: first, named entity recognition modules were used to identify drug names and AEs in posts, and then, a relation extraction algorithm was applied to these entities (Figure 2).

Figure 2. Adverse drug reaction pipeline. ADR: adverse drug reaction; ATC: Anatomical Therapeutic Chemical (in ATC classification); MedDRA: Medical Dictionary for Regulatory Activities.
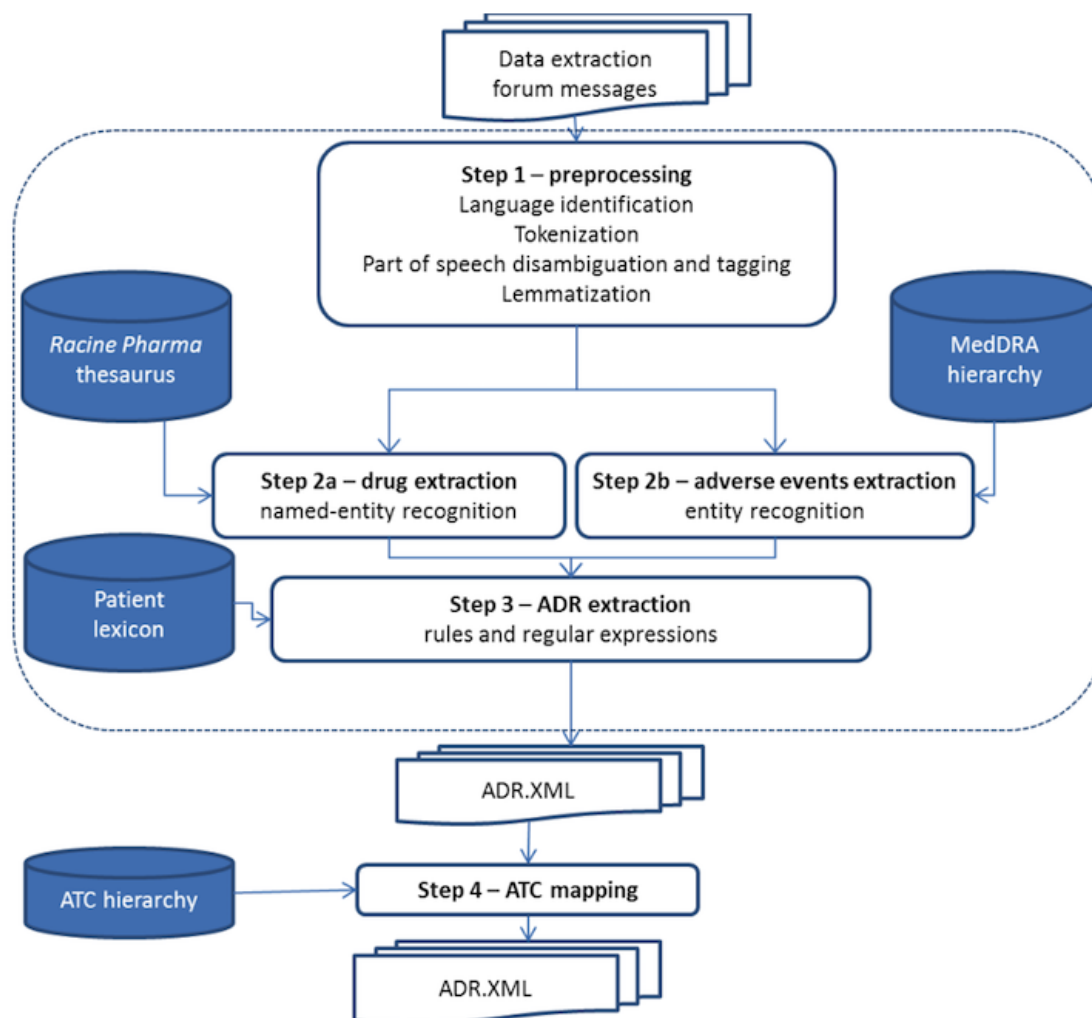
Drugs and Adverse Events Extraction

Regarding drugs, Expert System has developed a named entity recognition module capable of identifying words or tokens (occurrences) listed in *Racine Pharma* and extracting their positions from forum posts.

We mapped the extracted expressions from *Racine Pharma* to the chemical substance level (5th level) of the ATC (last updated version: December 19, 2016) [18]. Considering that the same active ingredient could be found under different trade names, we pooled all mapped expressions within the same ATC chemical subgroup (4th level) to define the drug concepts.

Expert System has developed another module based on MedDRA to identify the AEs and extract their position in the

posts. Fuzzy-matching and enrichment of thesaurus with colloquial terms enabled to take into account the characteristics of the posts on health forums.

Drug and Adverse Event Relationship Extraction (Skill Cartridge for Pharmacovigilance)

An NLP module has been developed by Expert System to capture the specific information regarding the relationship established between a drug and an AE by the post's author. The module combined a set of rules and regular expressions, with a *Patient* lexicon constituted to ensure that the post's author set out a situation of a person taking the drug and experiencing the symptom and excludes general information regarding a drug or an AE. This lexicon contains terms such as *I*, *me*, *my*, *cause*,

test, *take*, *feeling*, *because of*, *provoke*, *intolerance*, and *allergies*. This NLP module also included negation detection.

The algorithm can be summarized as follows:

1. Text was split into sentences based on the punctuation mark.
2. For each sentence, (named) entity recognition modules extracted drugs and AEs.
3. For each pair of drug and AE co-occurring in a sentence, if the NLP module extracted specific information regarding a relation, then the co-occurrence was classified as follows:
   - Explicit and positive ADR (eg, *Abilify causes me such fears that I cannot concentrate to read, work, etc...*);
   - Explicit and negative ADR (eg, *I took some Doliprane and I didn't feel any nausea);*
   - If no specific information was identified, the co-occurrence was classified as no ADR (eg, *Usually, Focalin's adverse effects are loss of appetite, insomnia, and naso-pharyngitis*).

Phase 1: Iterative Improvement of the Adverse Drug Reaction Information Extraction Module

Phase 1 consisted of an iterative process of validations and improvements of the tool. The review on the corpus was constituted by manual annotation of the messages by 2 experts in pharmacovigilance. The manual annotation process used a tool developed by Expert System for in-house testing purposes [20]. This tool did not integrate functionalities for blind manual annotation. No gold standard was established at this step. All along the process, both annotators could see the drugs, AEs, and ADRs extracted by the tool, and their task consisted of validating or invalidating these extractions and, if needed, complementing the annotations performed by the system. The annotators were asked to (1) check all drugs and all AEs in the post, irrespective of whether they have been involved in an ADR relationship, and (2) identify all drug-AE relationships in the same sentence. The output of phase 1 was a first estimation of precision, recall, and f-measure, assessing the capacity to extract information regarding drugs, AEs, and ADRs. More details about the corpus construction, evaluating concept selection, sample size determination, and the preliminary results of this phase were reported in the study by Chen et al [21].

Phase 2: Evaluation

Corpus for Evaluation

Evaluation Dataset

For phase 2, the selection of datasets is given in Figure 3.

On October 28, 2016, Kappa Santé Detec't database contained about 23 millions of posts, and approximately 21 million posts were published after January 1, 2007. Messages about drugs, for which marketing authorization has been withdrawn before the cut-off date, were considered as some AEs might appear at long term or patients might discuss about old drugs and their safety.

The software developed by Kappa Santé for purpose of Web discussions collection offered the possibility of preidentifying the posts containing at least one pharmaceutical or molecule name. We considered the dataset of all posts containing at least one preidentified compound or brand name and combined it with a complementary set of posts randomly sampled from the remainder of the Kappa Santé Detec't database.

The drug, AE, and ADR extraction modules developed for ADR-PRISM were executed on both sets of posts, and the output of the extraction module was considered as the evaluation dataset.

Figure 3. Phase 2 messages selection flowchart. ADR: adverse drug reaction; ATC: Anatomical Therapeutic Chemical (in ATC classification); MedDRA: Medical Dictionary for Regulatory Activities.
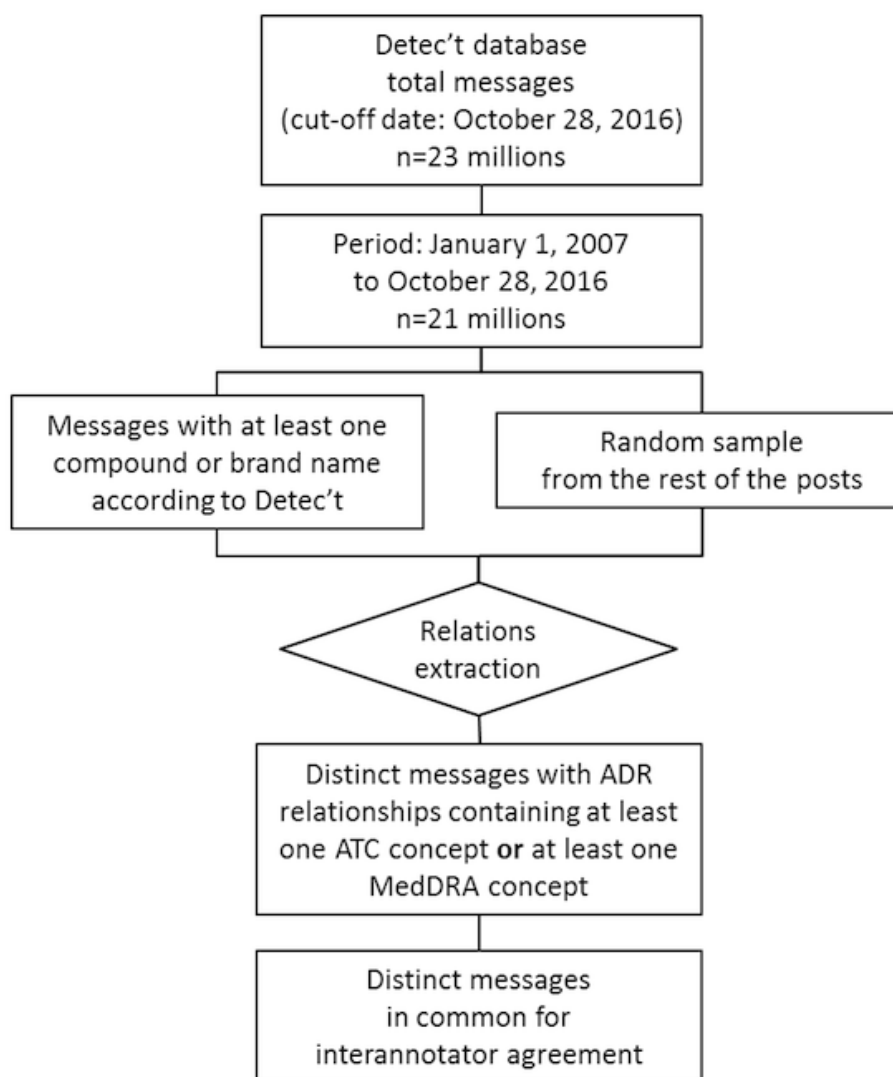
Drug Selection

The evaluation focused on 3 categories of drugs: (1) the most frequently extracted drugs in the corpus, (2) the most sold drugs in France, and (3) the drugs of interest of pharmacovigilance because of recent safety alerts. For the last 2 categories, the frequencies of the recognized named entities were also taken into account to cope with the necessary sample size.

The most sold drugs in 2013 in France are listed below, in their French commercial names, including the 10 most sold drugs with *mandatory* medical prescription for dispensation and the 10 most sold drugs with *optional* medical prescription for dispensation [22]:

- *Mandatory medical prescription for dispensation:* Levothyrox, Uvedose, Lamaline, Dafalgan codéiné, Méthadone, Crestor, Pivalone, Seresta, Emla patch, and Seroplex
- *Optional medical prescription for dispensation:* Doliprane, Dafalgan, Efferalgan, Kardégic, Spasfon, Gaviscon, Dexéryl, Météospasmyl, Biseptine, and Eludril.

The 2 experts additionally worked out a list of pairs of 1 drug concept and 1 AE concept corresponding to alerts that emerged in the last years (Table 1). This list of use cases has been further employed as a basis for the concept selection for evaluation. The concept selection in phase 1 (iterative improvement) [21] was based on the same consideration; therefore, those concepts on which the extraction tool performed well should be excluded from the selection for evaluation.

Table 1. List of drug and adverse event concepts selected as pharmacovigilance use cases.

Active drug ingredient: French names (date of marketing in cases where drug has been withdrawn)	Adverse events of interest	Corresponding Medical Dictionary for Regulatory Activities term	Media coverage or alert date
Gardasil	Autoimmune disease, complex regional pain syndrome, and postural orthostatic tachycardia syndrome	Autoimmune disorders (HLGT[a]) + complex regional pain syndrome (PT[b]) + postural orthostatic tachycardia syndrome (PT)	2013
Meningitec	Quality defect and any adverse event	All terms	2015
Methylphénidate: Ritaline, Concerta, Medikinet, Quasym, and Ritaline	Neurological and cardiac effects	Cardiac disorders (SOC[c]) + nervous system disorders (SOC)	2009 (recommendation)
Acétate de cyprotérone: Diane 35, Cypropharm (2010-2013), Elleacnelle (2010-2013), Evepar, Holgyme (2002-2015), Lumalia (2003-2013), Minerva, and Arielle (1997-2011)	Thromboembolic risk	Embolism and thrombosis (HLGT)	2012 (alert)
Fluoxétine: Prozac, Elindra (2001-2008), Fluctine (1998-2008), and Fontex (2001-2008)	All adverse events	All terms	Mostly 2013
Méthadone	Prolonged time from the start of the Q wave to the end of the T wave during electrocardiogram (approximates the time taken from when the cardiac ventricles start to contract to when they finish relaxing)	QT interval prolongation	2007
Sofosbuvir: Harvoni, Epclusa, and Sovaldi	Bradycardia	Bradycardia (PT)	2013
Codeine: Codenfan, Codoliporane, Migralgine, and Néocodion et Prontalgine	Respiratory disorders	Respiratory disorders (HLGT)	2012 re-evaluation
Hydroxyzine: atarax	Rhythm disorders	Cardiac arrhythmias (HLGT)	2014
Nicorandil: Adancor and Ikorel	Skin ulceration	Skin ulcer (HLT[d])	2012
Midodrine: Gutron	Hypertension	Blood pressure increased (PT)	2013
Galantamine: Reminyl, Galanthen (2012-2015), and Galema (2011-2015)	Skin reaction	Skin and subcutaneous tissue disorders (SOC)	2014
Crizotinib: Xalkori	Heart failure	Heart failures HLGT	2014
Valproate de sodium: Depakine, Imaslav, and Micropakine	Teratogenic effects	Congenital, familial, and genetic disorders (SOC)	2012?
Isotrétinoine: Curacné, Acnetrait, Contracné, Procuta, and Roaccutane	Teratogenic effects + psychiatric disorders	Congenital, familial, and genetic disorders (SOC) and psychiatric disorders (SOC)	2002
Fingolimod: Gilenya	Leukoencephalopathy	Toxic leukoencephalopathy (PT)	2014
Aripiprazole: Abilify	Suicidal behavior	Suicidal and self-injurious behavior (HLGT)	2013 (words of warning: 2016)

[a]HLGT: high-level group term.

[b]PT: preferred term.

[c]SOC: system organ class.

[d]HLT: high-level term.

Adverse Event Concepts Selection

The AE concepts for evaluation corresponded to preferred term (PT) level in the MedDRA hierarchy, and we focused similarly on 2 categories: (1) the most frequent extracted PTs and (2) the PTs of interest of pharmacovigilance (Table 1), guided also by the frequencies of the recognized entities in the corpus to cope with the necessary sample size.

Adverse Drug Reaction Definition

We defined an ADR as any explicit and positive relationship between a drug and an AE where the drug (respectively the AE) belonged to the list of concepts previously selected.

Corpus for the Evaluation (Random Sample From Evaluation Dataset)

Finally, a random sample from the evaluation dataset (ie, among all the ADR extracted by the ADR information extraction module) constituted our corpus for evaluation.
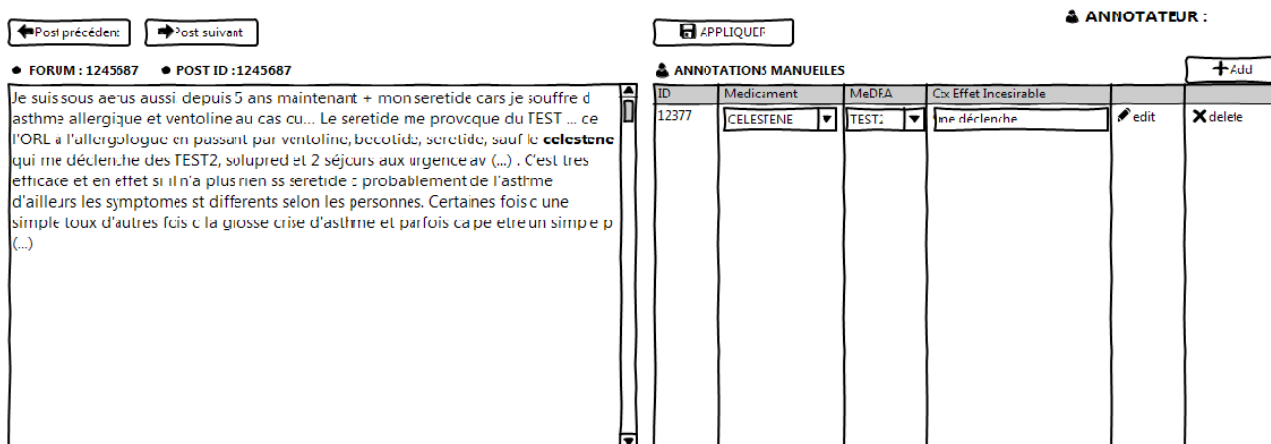
Gold Standard

Adverse Drug Reactions From Patient Reports in Social Media Manual Annotation Platform

We developed a Web application dedicated to manual annotations valuable as gold standard, for the project purposes in phase 2.

This application was based on Java Servlets and JavaScript libraries and connected in Java Database Connectivity to the dataset of posts selected for the gold standard annotation. We used a self-completion mechanism to attach portions of message to *Racine Pharma* and MedDRA terms. We used drag-and-drop operations to fill in a table containing the manual review of the co-occurrences, where each line was dedicated to 1 co-occurrence. For each co-occurrence, the drug and the AE were dragged-and-dropped in the first and the second column, respectively. In the third column, the manual annotator was given a drop-down menu that presented 3 possibilities for defining the co-occurrence: explicit positive ADR, explicit negative ADR, or no ADR. We finally offered the possibility to export this table containing the manual annotations obtained via the application. The interface of this application is shown in Figure 4.

Figure 4. Interface of the manual annotation platform. ID: identifier; MedDRA: Medical Dictionary for Regulatory Activities.

Gold Standard Construction

An annotator guideline has been established to standardize manual annotations. A total of 2 experts in medical terminologies have annotated all ADR relationships for each post in the evaluation corpus, and then, the manual annotation was considered as gold standard. Different from phase 1, here we expected that the annotators annotated all causal relationships between a drug and an AE, even if the drug and AE were not in the same sentence, which will allow us further evaluate the impact of the sentence boundary constraint of the extraction tool.

Each annotator annotated 55.1% (261/474) of the posts in the evaluation corpus; thus, a subset of 10.1% (48/474) of the posts was annotated by both of them. The interannotator agreement could be assessed on these double-annotated posts. In case of disagreements, the 2 annotators discussed to achieve a consensus. If a lot of disagreements had occurred, the annotators would have been asked to learn again the guideline and revise their annotations. All along the process, both annotators were double-blinded: first, from the ADRs identified by the ADR information extraction module, and second, from the other annotator's annotations.

Statistical Analyses

Primary Analysis

To assess the efficacy of the ADR information extraction module as compared with the gold standard, we will globally calculate the recall, precision, and f-measure.

In a post, if the ADR information extraction module identifies the same expression from *Racine Pharma* at the same position; *and* the same AE expression at PT level from MedDRA hierarchy at the same position; *and* the same type of relationship, as did the gold standard, then we will count the extracted *ADR* as a true positive (Figure 5).

Figure 5. Recall, precision, and f-measure definitions. ADR: adverse drug reaction; ADR-PRISM: Adverse Drug Reaction from Patient Reports in Social Media; FN: false negative, manually annotated ADR by the gold standard that is not extracted by the ADR information extraction tool; FP: false positive, positive explicit relationship extracted by the ADR information extraction tool that is not manually annotated by gold standard; TN: true negative; TP: true positive.
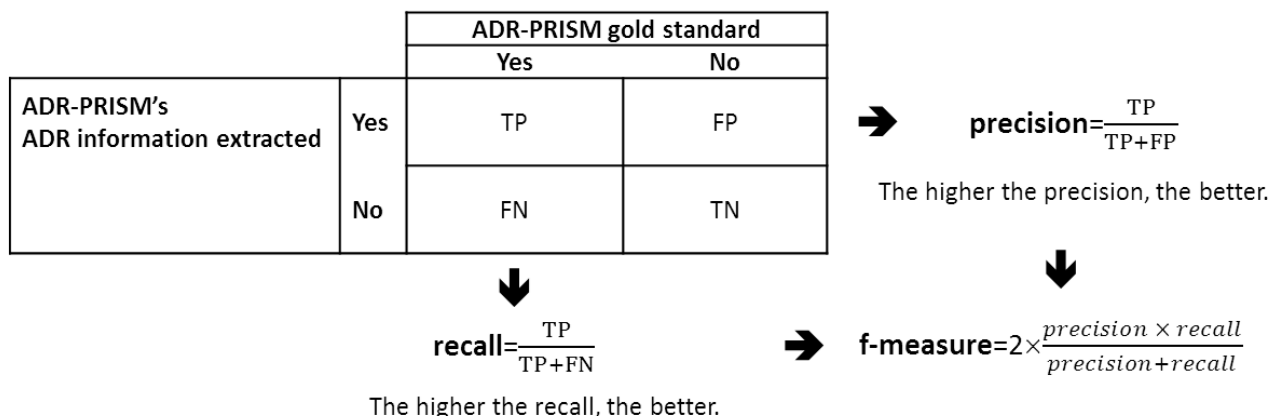
F-Measure Gold Standard's Evaluation

The interannotator agreement will be assessed by a hierarchical Kappa [23]. A hierarchical Kappa will enable to take into account the situation where the 2 annotators will disagree in either the level or the expression in ATC or MedDRA terminologies but will agree on higher levels than the annotated ones. Separate calculations for interannotator agreements by hierarchical Kappa on drug and AE expressions will be provided as well.

Secondary Analyses

We will complete these principal results with the following analyses.

We will take into account MedDRA and ATC granularities by reproducing precision, recall, and f-measure at each level of the hierarchies (system organ class [SOC], high-level group term [HLGT], high-level term [HLT], and preferred term [PT] of MedDRA and anatomical main group, therapeutic subgroup, pharmacological subgroup, and chemical subgroup of ATC).

We will provide a relaxed definition of true-positive ADRs combining the following 3 conditions: (1) the positions in the post of the extracted expressions for drug and AE both match the positions of the drug and AE expressions manually annotated, (2) the extracted and the manually annotated expressions for AE from MedDRA hierarchy are found in the same SOC levels, and (3) the classification of the identified relationship match the classification of the manually annotated relationship. We will calculate recall, precision, and f-measure with this definition for ADRs, drug, and AE identification separately.

We will provide a global estimation of the performance of ADR information extraction module by taking into account other types of messages. These messages are those without information on at least one ADR concept, that is, messages in which (1) ADR information is extracted but is not concerned by the drug- or AE-selected concepts, (2) no ADR information is extracted but only co-occurrences of information about 1 drug and 1 AE, (3) only AE information is extracted, (4) only drug information is extracted, and (5) neither drug nor AE information is extracted. After having sampled each type of message described above, we will then calculate global recall or precision or f-measure on all types of messages by marginal calibration.

All analyses will be performed in *R* software [24].

Sample Size Calculation

Recall can be estimated from a random sample of ADR concepts annotated by the gold standard, whereas precision can be estimated from a random sample of ADR concepts extracted by the ADR information extraction module. However, creating the gold standard on the whole evaluation dataset before sampling in it for the recall was an overwhelming task. At the same time, a vast majority of messages did not contain any information on the drug concept. Thus, we expected a very low proportion of ADR concepts identified by the ADR information extraction module in the evaluation dataset. Therefore, the recall estimated from a random sample of ADR concepts annotated by the gold standard was mathematically approximated by the recall estimated on ADR concepts sampled for the precision.

The sample size required for different expected precision or recall is provided in Table 2. By hypothesizing a global recall of 0.5, we needed at least 384 identified ADR concepts to obtain a 95% CI, having a total width of 0.1. The final determination of sample size was based on the previously estimated precision and recall.

Table 2. Sample size required according to the expected precision and recall and total width of confidence intervals (CIs).

Range of CI	Expected precision or recall values								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
90% CI	35	61	81	92	96	92	81	61	35
95% CI	138	246	323	369	384	369	323	246	138

Differences Between Phase 1 and Phase 2

The main differences between phase 1 and phase 2 are threefold (Table 3). First, in phase 1, the annotations targeted all drugs and AEs even if not involved in an ADR relation, whereas in phase 2, annotation focuses on the ADR relations. Second, in phase 2 (and only in phase 2) a gold standard based on blind manual annotation by 2 experts in medical terminologies was established and the interannotator agreement was measured. In contrast, the manual annotations of phase 1 were not blind, and

most of the work consisted of validating the automatic annotations; thus, the interannotator agreement was not addressed. Finally, in phase 2, additional parameters such as the granularity and the segmentation of the messages (sentence boundaries) will be analyzed. The precision, recall, and f-measure will be calculated with a gold standard at the message level and might be compared with precisions, recalls and f-measures calculated at the sentence level and for all MedDRA and ATC granularities.

Table 3. Common features and differences between phase 1 and phase 2.

Issues	Phase 1	Phase 2
Evaluation focus	Drugs and adverse events entity recognition	Adverse drug reaction relationships recognition
Gold standard (manual annotations)	Not blind	Blind
Gold standard annotators	Pharmacovigilance experts	Experts in medical terminologies
Interannotator agreement	No	Yes
Granularity issue	No	Yes
Sentence boundary issue	No	Yes

Results

At the time of publishing this evaluation protocol, several steps of this project would have been completed. The corpus for evaluation has been constituted, and the NLP tools have identified and extracted the information about ADR inside this corpus. We selected the ADR concepts and constituted the samples of the entities necessary to set up the gold standard. The 2 annotators completed the annotations process.

Data analyses for assessing the interannotator agreement and the performance of the ADR information extraction module as compared with the gold standard are ongoing, and the study results are expected internally before the end of 2018.

Discussion

Summary

With the objective of using messages on health forums as a new source for drug safety, the systems developed to mine the messages must follow strict evaluation rules. This is even more important as these systems might be used to support decision making. The protocol presented in this paper has been designed to evaluate the ADR information extraction Skill Cartridge developed in ADR-PRISM in a pharmacovigilance perspective.

Study Strengths

The protocol presented in this paper is an attempt to apply clinical research–level guidelines [14] to the assessment of such systems.

First, we paid particular attention to the establishment and the validation of the gold standard. The gold standard was established by 2 trained specialists of medical terminologies who annotated selected messages. Manual annotation was performed in a *double-blind* manner, namely, from both the other annotator and the ADR identified by the ADR information extraction module. The annotated corpus, therefore, constituted

a valuable gold standard. On the basis of the study by Sarker et al [4] (see Multimedia Appendix 1), we could find 13 studies where a gold standard, that is, manual annotations used for evaluation, was implemented [25-37]. Conversely, in 8 studies, there was no gold standard [38-45], and the extracted ADRs were compared with already known AEs from FAERS [39-41,43,44] or drug label declared to FDA [38,45] or even AE described in websites [42]. Moreover, in the ADR-PRISM protocol, a common subset of randomly selected messages was annotated in a blind manner by the 2 experts; interannotator agreement evaluation will also ensure ourselves about the quality of this gold standard. In most of studies, the authors gauged the interannotators agreement [25,27,30-32,34,35]. However, it is not systematic [26,28,29,33,36,37].

Second, by calculating a sample size of messages collected from social media, to assess recall, precision, and f-measure, we guaranteed the statistical power to place reliance on our study results.

The chosen terminologies are another crucial aspect of this work. On the one hand, the *Racine Pharma* thesaurus, with 5164 entries, exhaustively covers a large range of drug names and active ingredients. On the other hand, the MedDRA hierarchy is used daily by drug safety experts and considered expressive for this task. By using these terminologies, we expect to increase the sensitivity of the ADR-PRISM Skill Cartridge for pharmacovigilance. Our choice to map *Racine Pharma* to ATC was guided by 2 aspects. First, ATC like MedDRA has a hierarchical structure. Thus, we will be able to evaluate the ADR information extraction tool based on different hierarchical levels of these terminologies. Second, ATC and MedDRA are widely used and internationally agreed reference terminologies. Hence, we expect to provide strong and reproducible results.

The ADR information extraction module was not only based on drug and AE information identification but also on rules and regular expressions. As such, we expect to discard noninformative sentences, addressing general information about

the drug or drug indications. We would also be able to identify unexpected *positive* effects, as for example, headaches that would be reduced by a drug without indication for the treatment of this kind of pain. Only few studies have been able to take this aspect into account [27,29,30,32,34,46].

Study Limitations

Despite its positive aspects, the study exhibits several limitations.

The mapping between the terms listed in the drug thesaurus *Racine Pharma* and the terms referenced in the chemical substance in the ATC hierarchy was incomplete. Among the 5164 terms in *Racine Pharma*, 852 (16.5%, 852/5164) could not be mapped to ATC, for example, some phytotherapies (St John's wort herbal tea, Silver birch juice, extract of licorice root, arum triphyllum compound, arnica, etc). This could have a negative impact on the recall scores calculated according to the ATC levels.

In social media' posts, slangs and colloquial languages are frequent; likewise, syntactic rules are approximate. We chose to use fuzzy-matching and thesaurus enrichment to take into account this bias. However, this approach is inherently nonexhaustive with negative impact on the recall of the ADR information extraction module.

Regarding the gold standard, 2 experts in medical terminologies trained in drug safety performed manual annotations. However, contrary to the phase 1 manual review, none of them could be considered as a pharmacovigilance expert.

The medical informatics community needs shared open corpora to evaluate their methods and tools. Recent efforts have led to making several datasets more accessible and the evaluation of the methods more standardized, for example, the Multiparameter Intelligent Monitoring in Intensive Care corpus [47] and the Informatics for Integrating Biology and the Bedside challenges [48]. Drug safety could benefit from open-access validation datasets made available for research purposes [25,35]. However, several obstacles remain. For example, the ADR-PRISM consortium got approval to reuse posts for research purposes, but the approval is restricted to the project.

Conclusions

The objective of this study is to present the scientific approach developed in the first stage of the ADR-PRISM project. At this stage, our principal objective was to evaluate the performance of the ADR information extraction Skill Cartridge against a gold standard constituted by human annotations. To address this question, our goal was to adapt the principles adopted in clinical research to assess sound and trustworthy measurements of precision, recall, and f-measure.

With the statistical theory, we calculated a sample size. This guaranteed enough ADR information and sufficient narrowness of the confidence interval, to scientifically conclude on our principal objective. We also avoided unnecessarily large extractions for which the process of manual annotations is nothing but both wasteful and time-consuming.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Designs of articles related to ADR information extraction on social media (replicated from Sarker et al).

[PDF File (Adobe PDF File), 94KB - resprot_v8i5e11448_app1.pdf]

References

1. Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G. Detecting adverse events using information technology. J Am Med Inform Assoc 2003;10(2):115-128 [FREE Full text] [Medline: 12595401]
2. Golomb BA, McGraw JJ, Evans MA, Dimsdale JE. Physician response to patient reports of adverse drug effects: implications for patient-targeted adverse effect surveillance. Drug Saf 2007;30(8):669-675. [doi: 10.2165/00002018-200730080-00003] [Medline: 17696579]
3. Hazell L, Shakir SA. Under-reporting of adverse drug reactions : a systematic review. Drug Saf 2006;29(5):385-396. [doi: 10.2165/00002018-200629050-00003] [Medline: 16689555]
4. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. J Biomed Inform 2015 Apr;54:202-212 [FREE Full text] [doi: 10.1016/j.jbi.2015.02.004] [Medline: 25720841]
5. Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, et al. A Bayesian neural network method for adverse drug reaction signal generation. Eur J Clin Pharmacol 1998 Jun;54(4):315-321. [Medline: 9696956]

6.   DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. The American Statistician 1999 Aug;53(3):177. [doi: 10.2307/2686093]

7.   Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. Drug Saf 2002;25(6):381-392. [doi: 10.2165/00002018-200225060-00001] [Medline: 12071774]

8.   van Puijenbroek EP, Bate A, Leufkens HG, Lindquist M, Orre R, Egberts AC. A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions. Pharmacoepidemiol Drug Saf 2002;11(1):3-10. [doi: 10.1002/pds.668] [Medline: 11998548]

9.   Katsahian S, Simond ME, Leprovost D, Lardon J, Bousquet C, Kerdelhué G, et al. Evaluation of internet social networks using net scoring tool: a case study in adverse drug reaction mining. Stud Health Technol Inform 2015;210:526-530. [doi: 10.3233/978-1-61499-512-8-526] [Medline: 25991203]

10.   ISDB. International Society of Drug Bulletins.: ISDB EU; 2005 Jan. ISDB EU: Berlin Declaration on Pharmacovigilance URL: https://www.akdae.de/Arzneimittelsicherheit/Weitere/Archiv/2005/85_200501252.pdf [accessed 2018-06-29] [WebCite Cache ID 70UZb66y8]

11.   Ho T, Le L, Thai DT, Taewijit S. Data-driven approach to detect and predict adverse drug reactions. Curr Pharm Des 2016;22(23):3498-3526. [Medline: 27157416]

12.   Wong A, Plasek JM, Montecalvo SP, Zhou L. Natural language processing and its implications for the future of medication safety: a narrative review of recent advances and challenges. Pharmacotherapy 2018 Jun 09;38(8):822-841. [doi: 10.1002/phar.2151] [Medline: 29884988]

13.   Tricco AC, Zarin W, Lillie E, Pham B, Straus SE. Utility of social media and crowd-sourced data for pharmacovigilance: a scoping review protocol. BMJ Open 2017 Dec 19;7(1):e013474 [FREE Full text] [doi: 10.1136/bmjopen-2016-013474] [Medline: 28104709]

14.   Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 2015 Oct 28;351:h5527 [FREE Full text] [doi: 10.1136/bmj.h5527] [Medline: 26511519]

15.   Bousquet C, Dahamna B, Guillemin-Lanne S, Darmoni SJ, Faviez C, Huot C, et al. The adverse drug reactions from patient reports in social media project: five major challenges to overcome to operationalize analysis and efficiently support pharmacovigilance process. JMIR Res Protoc 2017 Sep 21;6(9):e179 [FREE Full text] [doi: 10.2196/resprot.6463] [Medline: 28935617]

16.   Board I. Medical Dictionary for Regulatory Activities. URL: https://www.meddra.org/ [accessed 2018-06-27] [WebCite Cache ID 70Ua4UBNo]

17.   Catalog and Index of Medical Sites French-language. URL: http://www.chu-rouen.fr/cismef/outils/outils-et-services/ [accessed 2018-06-27] [WebCite Cache ID 70UYyaz1Q]

18.   World Health Organization. ATC/DDD Index 2017 URL: https://www.whocc.no/atc_ddd_index/ [accessed 2018-06-27] [WebCite Cache ID 70UXT3c84]

19.   Kappa Santé. Pharmaco-epidemiology Interventions in public and numeric health URL: https://www.kappasante.com/ [accessed 2018-06-27] [WebCite Cache ID 70UZvaGf0]

20.   Kostadinov F. GitHub. TEMIS LUXID 7.0.1 SKILL CARTRIDGE DEVELOPMENT CYCLE URL: http://fabian-kostadinov.github.io/2015/10/04/temis-luxid701-skill-cartridge-development-cycle/ [accessed 2018-06-27] [WebCite Cache ID 70UXfPs4L]

21.   Chen X, Deldossi M, Aboukhamis R, Faviez C, Dahamna B, Karapetiantz P, et al. Mining adverse drug reactions in social media with named entity recognition and semantic methods. Stud Health Technol Inform 2017;245:322-326. [Medline: 29295108]

22.   Cavalié P, Djeraba A. ANSM [National Agency for the Safety of Medicines and Health Products]. 2014. [Analysis of drug sales in France in 2013] URL: https://ansm.sante.fr/var/ansm_site/storage/original/application/3df7b99f8f4c9ee634a6a9b094624341.pdf [accessed 2018-06-27] [WebCite Cache ID 70UXyC8tD]

23.   Kiritchenko S, Matwin S, Nock R, Famili A. Learning Evaluation in the Presence of Class Hierarchies: Application to Text Categorization. Berlin, Heidelberg: Springer; 2006 Presented at: 19th Conference of the Canadian Society for Computational Studies of Intelligence; June 7-9, 2006; Québec City, Québec, Canada p. 395-406 URL: https://www.svkir.com/papers/Kiritchenko-et-al-hierarchical-AI-2006.pdf

24.   R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016. URL: https://www.r-project.org/ [WebCite Cache ID 77aU4dSIp]

25.   Ginn R, Pimpalkhute P, Nikfarjam A, Patki A, O'Connor K, Sarker A, et al. Arizona State University. 2014. Mining Twitter for adverse drug reaction mentions: a corpus and classification benchmark URL: http://www.nactem.ac.uk/biotxtm2014/papers/Ginnetal.pdf [accessed 2019-04-01] [WebCite Cache ID 77K671ICh]

26.   Hadzi-Puric J, Grmusa J. Automatic Drug Adverse Reaction Discovery from Parenting Websites Using Disproportionality Methods. 2012 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining; August 26-29, 2012; Istanbul, Turkey p. 792-797. [doi: 10.1109/ASONAM.2012.144]

XSL•FO

RenderX

27. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks. 2010 Presented at: Workshop on Biomedical Natural Language Processing (BioNLP 2010) ACL 2010; July 15, 2010; Uppsala, Sweden p. 117-125 URL: https://www.aclweb.org/anthology/W10-1915

28. Liu X, Chen H. AZDrugMiner: an information extraction system for mining patient-reported adverse drug events in online patient forums. 2013 Presented at: International Conference on Smart Health; August 3-4, 2013; Beijing, China p. 134-150. [doi: 10.1007/978-3-642-39844-5_16]

29. Nikfarjam A, Gonzalez GH. Pattern mining for extraction of mentions of adverse drug reactions from user comments. AMIA Annu Symp Proc 2011;2011:1019-1026 [FREE Full text] [Medline: 22195162]

30. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. J Am Med Inform Assoc 2015 May;22(3):671-681 [FREE Full text] [doi: 10.1093/jamia/ocu041] [Medline: 25755127]

31. O'Connor K, Pimpalkhute P, Nikfarjam A, Ginn R, Smith KL, Gonzalez G. Pharmacovigilance on Twitter? Mining tweets for adverse drug reactions. AMIA Annu Symp Proc 2014;2014:924-933 [FREE Full text] [Medline: 25954400]

32. Patki A, Sarker A, Pimpalkhute P, Nikfarjam A, Ginn R, Oconnor K, et al. Arizona State University. 2014. Mining Adverse Drug Reaction Signals from Social Media: Going Beyond Extraction URL: https://tinyurl.com/y4xwmjco [accessed 2019-04-01] [WebCite Cache ID 77K6WFaat]

33. Sampathkumar H, Chen X, Luo B. Mining adverse drug reactions from online healthcare forums using hidden Markov model. BMC Med Inform Decis Mak 2014;14:91 [FREE Full text] [doi: 10.1186/1472-6947-14-91] [Medline: 25341686]

34. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. J Biomed Inform 2015 Feb;53:196-207 [FREE Full text] [doi: 10.1016/j.jbi.2014.11.002] [Medline: 25451103]

35. Segura-Bedmar I, Revert R, Martínez P. Detecting drugs and adverse events from Spanish social media streams. In: Proceedings of the 5th International Workshop on Health Text Mining and Information Analysis (Louhi). 2014 Presented at: 5th International Workshop on Health Text Mining and Information Analysis (Louhi); April 26-30, 2014; Gothenburg, Sweden p. 106-115 URL: https://www.aclweb.org/anthology/W14-1117

36. Yang M, Wang X, Kiang M. Identification of Consumer Adverse Drug Reaction Messages on Social Media. In: PACIS 2013 Proceedings. 2013 Presented at: The Pacific Asia Conference on Information Systems; June 18-22, 2013; Jeju Island, Korea URL: https://aisel.aisnet.org/pacis2013/193

37. Yates A, Goharian N. ADRTrace: detecting expected and unexpected adverse drug reactions from user reviews on social media sites. 2013 Presented at: European Conference on Information Retrieval; March 24-27, 2013; Moscow, Russia URL: https://link.springer.com/chapter/10.1007/978-3-642-36973-5_92 [doi: 10.1007/978-3-642-36973-5_92]

38. Benton A, Ungar L, Hill S, Hennessy S, Mao J, Chung A, et al. Identifying potential adverse effects using the web: a new approach to medical hypothesis generation. J Biomed Inform 2011 Dec;44(6):989-996 [FREE Full text] [doi: 10.1016/j.jbi.2011.07.005] [Medline: 21820083]

39. Bian J, Topaloglu U, Yu F. Towards large-scale Twitter mining for drug-related adverse events. SHB12 (2012) 2012 Oct 29;2012:25-32 [FREE Full text] [doi: 10.1145/2389707.2389713] [Medline: 28967001]

40. Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. AMIA Annu Symp Proc 2011;2011:217-226 [FREE Full text] [Medline: 22195073]

41. Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T, et al. Digital drug safety surveillance: monitoring pharmaceutical products in Twitter. Drug Saf 2014 May;37(5):343-350 [FREE Full text] [doi: 10.1007/s40264-014-0155-x] [Medline: 24777653]

42. Jiang K, Zheng Y. Mining Twitter Data for Potential Drug Effects. 2013 Presented at: The 9th International Conference on Advanced Data Mining and Applications (ADMA); 2013; Hangzhou, China. [doi: 10.1007/978-3-642-53914-5_37]

43. Yang CC, Yang H, Jiang L. Postmarketing drug safety surveillance using publicly available health-consumer-contributed content in social media. ACM Trans Manage Inf Syst 2014 Apr 01;5(1):1-21. [doi: 10.1145/2576233]

44. Yang C, Yang H, Jiang L, Zhang M. Social media mining for drug safety signal detection. 2012 Presented at: Proceedings of the international workshop on Smart healthwellbeing; Octiber 29, 2012; Maui, Hawaii, USA URL: http://www.pages.drexel.edu/~hy95/publications/2012SHB.pdf

45. Yeleswarapu S, Rao A, Joseph T, Saipradeep VG, Srinivasan R. A pipeline to extract drug-adverse event pairs from multiple data sources. BMC Med Inform Decis Mak 2014;14:13 [FREE Full text] [doi: 10.1186/1472-6947-14-13] [Medline: 24559132]

46. Bollegala D, Maskell S, Sloane R, Hajne J, Pirmohamed M. Causality patterns for detecting adverse drug reactions from social media: text mining approach. JMIR Public Health Surveill 2018 May 09;4(2):e51 [FREE Full text] [doi: 10.2196/publichealth.8214] [Medline: 29743155]

47. Saeed M, Villarroel M, Reisner AT, Clifford G, Lehman L, Moody G, et al. Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database. Crit Care Med 2011 May;39(5):952-960 [FREE Full text] [doi: 10.1097/CCM.0b013e31820a92c6] [Medline: 21283005]

48.   Zhu X, Cherry C, Kiritchenko S, Martin J, de Bruijn B. Detecting concept relations in clinical text: insights from a state-of-the-art model. J Biomed Inform 2013 Apr;46(2):275-285 [FREE Full text] [doi: 10.1016/j.jbi.2012.11.006] [Medline: 23380683]

Abbreviations

ADR: adverse drug reaction
ADR-PRISM: Adverse Drug Reaction from Patient Reports in Social Media
AE: adverse event
ATC: Anatomical Therapeutic Chemical (in ATC classification)
FAERS: Food and Drug Administration' adverse event reporting system
FDA: Food and Drug Administration
HLGT: high-level group term
HLT: high-level term
MedDRA: Medical Dictionary for Regulatory Activities
NLP: natural language processing
PT: preferred term
SIBM-CISMeF: Biomedicine informatics service catalogue and index of French language medical websites (French Service d'Informatique Biomédicale-Catalogue et Index des Sites Médicaux de langue Française)
SOC: system organ class

XSL•FO
RenderX

Protocol

Testing a Communication Assessment Tool for Ethically Sensitive Scenarios: Protocol of a Validation Study

Thierry Daboval[1], MSc, MD; Natalie Ward[2], MA, PhD; Jordan R Schoenherr[3], PhD; Gregory P Moore[1], MD; Caitlin Carew[1], MD; Alicia Lambrinakos-Raymond[1], MD; Emanuela Ferretti[1], MSc, MD

[1]Department of Pediatrics, Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, ON, Canada

[2]Performance and Evaluation, Genome Canada, Ottawa, ON, Canada

[3]Department of Psychology, Carleton University, Ottawa, ON, Canada

Corresponding Author:
Thierry Daboval, MSc, MD
Department of Pediatrics
Children's Hospital of Eastern Ontario
University of Ottawa
401 Smyth Road
Ottawa, ON,
Canada
Phone: 1 613 737 7600 ext 3258
Fax: 1 613 737 4847
Email: tdaboval@cheo.on.ca

Abstract

Background: Although well-designed instruments to assess communication during medical interviews and complex encounters exist, assessment tools that differentiate between communication, empathy, decision-making, and moral judgment are needed to assess different aspects of communication during situations defined by ethical conflict. To address this need, we developed an assessment tool that differentiates competencies associated with practice in ethically challenging situations. The competencies are grouped into three distinct categories: communication skills, civility and respectful behavior, clinical and ethical judgment and decision-making.

Objective: The overall objective of this project is to develop an assessment tool for ethically sensitive scenarios that measures the degree of respect for the attitudes and beliefs of patients and family members, the demands of clinical decision-making, and the success in dealing with ethical conflicts in the clinical context. In this article, we describe the research method we will use during the pilot-test study using the neonatal context to provide validity evidence to support the features of the Assessment Communication Tool for Ethics (ACT4Ethics) instrument.

Methods: This study is part of a multiphase project designed according to modern validity principles including content, response process, internal structure, relation to other variables, and social consequences. The design considers threats to validity such as construct underrepresentation and factors exerting nonrandom influence on scores. This study consists of two primary steps: (1) train the raters in the use of the new tool and (2) pilot-test a simulation using an Objective Structured Clinical Examination. We aim to obtain a total of 90 independent assessments based on the performance of 30 trainees rated by 15 trained raters for analysis. A comparison of raters' responses will allow us to compute a measure of interrater reliability. We will additionally compare the results of ACT4Ethics with another existing instrument.

Results: This study will take approximately 18 months to complete and the results should be available by September 2019.

Conclusions: ACT4Ethics should allow clinician-teachers to assess and monitor the development of competency of trainees' judgments and communication skills when facing ethically sensitive clinical situations. The instrument will also guide the provision of meaningful feedback to ensure that trainees develop specific communication, empathy, decision-making, and ethical competencies.

International Registered Report Identifier (IRRID): PRR1-10.2196/12039

XSL•FO
RenderX

Introduction

Importance of Ethics and Communication Education in Medicine

Clinicians face ethically challenging situations such as limiting care, engaging in collective or shared decision-making and surrogate decision-making, and participating in end-of-life discussions. In order to engage in successful clinical decision-making during these situations, specific competencies must be identified, education milestones should be established, and learning needs must be monitored. Given the complexity and ubiquity of these cases requiring tactful communication, high-stakes moral and ethical judgment, and mindfulness, training must start early during undergraduate and continue into postgraduate health care professional education [1-4]. The development of these competencies requires the promotion of skills associated with shared decision-making, increased empathetic accuracy of clinicians, and a focus on decreasing psychological distress for severely ill patients and their families [5-7].

Although all health care professions involve difficult situations, neonatology presents a diverse set of considerations including a wide range of ethically sensitive situations and frequent inter- and intradisciplinary communication. Supporting this, there is a growing recognition that Neonatal-Perinatal Medicine (NPM) trainees require a greater understanding of ethical features of clinical situations and that there is a need to increase the availability and efficacy of ethics and communication programs [8]. Our team implemented a neonatal ethics teaching program that integrates knowledge of ethics acquisition and provides opportunities to practice communication skills in a safe, structured environment [9]. This program helps develop ethical competencies by providing constructive feedback and promoting self-reflection [10]. Our neonatal ethics teaching program and other neonatal ethics published curricula focus on helping trainees refine their knowledge of ethics and learn competencies relevant to professionalism and communication [11,12]. While multiple approaches based on adult learning theory support trainees' education, pedagogical methods best suited to teaching ethics and communication have yet to be identified. Assessment methods that can differentiate specific features of these situations need to be developed. These assessment tools can be used to support three important objectives: monitoring the development of a clinician's knowledge and skills over the course of their career, supporting the evaluation of teaching methods, and guiding the design of effective education programs [13]. We considered these objectives in the development of our assessment tool.

Need for a Communication Assessment Tool for Ethically Sensitive Clinical Situations

Studies focusing on NPM residency training have identified the need to develop tools supported by empirical evidence that assess knowledge and behavioral learning [8,14]. In addition to NPM, many other subspecialties in medicine need such assessment instruments. While knowledge tests in medical ethics competency and attitudes have been described for pediatrics and internal and neonatal medicine [15-18], they are not sufficient in and of themselves; assessment of general competencies such as communication skills must also be considered [9,11,19-21]. Our literature review identified several well-designed tools to assess communication skills during medical interviews. They were typically designed for medical encounters exploring symptoms or providing difficult news to patients and their family members. We failed to identify any that were directly related to medical ethical situations [22-27]. For instance, the Gap-Kalamazoo Communication Skills Assessment Form (GKCAS), which is used to assess communication competencies during complex encounters across medical subspecialties, includes domains such as building relationships, understanding the family perspective, sharing information, demonstrating empathy, and reaching agreement, but we believe it lacks important topics such as moral judgment and ethical conflicts [28].

Foundation of the Communication Assessment Tool for Ethics Framework

Against this background, we created an assessment instrument designed to differentiate competencies related to ethical judgment and decision-making, respect and empathy, and communication skills during ethically sensitive situations. Although developed in the context of NPM, the Assessment of Communication Tool for Ethics (ACT4Ethics) can be applied to a wide range of ethically challenging situations encountered in many fields in the health care professions. This assessment instrument includes key milestones in the roles of medical professional, communicator, and collaborator [11]. It includes competencies aligned with verbal and nonverbal communication skills adapted to medical interviews [23]; communication skills included in delivering bad news [26], demonstrating awareness of ethical features of a situation including virtue ethics; bioethical principles and communicative ethics [29]; ethical judgment and decision-making [30]; and engaging patients in a decision-making process [31]. We used the academic literature to guide the construction of the ACT4Ethics scale. We grouped our assessment criteria into three broad domains reflecting a proposed set of distinct competencies: communication, civility and respectful behavior, and clinical and ethical judgment and decision-making.

The communication skills subscale evaluates basic verbal and nonverbal skills, while also addressing more complex skills such as conversational pragmatics that might arise during a clinical encounter [22,32]. Given the often ambiguous nature of the clinical setting as well as differences in clinicians' and patients' knowledge, this communication subscale also assesses whether a learner effectively closed the loop (ie, sought explicit confirmation that they shared an understanding) during critical periods throughout the encounter as well as at the end of the encounter [33]. Ethically sensitive scenarios regularly include elements such as presenting affectively charged information

that would reasonably disappoint patients and requires the consideration of empathy [26] and respect for their values and beliefs.

The civility and empathy subscale assesses empathy-related skills and overt social cues associated with respect. Civility has become a growing concerning in professional environments [34,35] and clinical settings [36-38]. In contrast to respectful and disrespectful behaviors, incivility reflects actions wherein the intentions of the communicator lack clarity. This lack of clarity can lead to negative consequences over lengthy periods [34]. Similarly, empathy can be defined along two separate dimensions. Whereas cognitive empathy requires understanding another's thoughts and beliefs in order to predict behavior [39], affective empathy involves sharing the emotional response for another's joy and pain [40]. These appear to have a distinctive neurological basis [41,42]. Crucially, in the context of a clinical encounter, patients and family members might not disclose relevant information or consider clinical alternatives if they do not believe that clinicians respect their emotional responses, beliefs, or choices.

The judgment and decision-making subscale assesses clinical and ethical features of judgment and decision-making. Judgment and decision-making require consideration of a clinician's awareness of the situation [43] and whether the clinician gathers and assesses evidence and considers diagnostic and treatment alternatives in an unbiased manner [44] given the constraints of the situation [45]. In the clinical context, this involves considering multiple sources of information such as that provided by other health care professionals, patients, and family members. Even if a clinician is an effective communicator and has demonstrated respect and empathy with a patient, failure to integrate the available information will ultimately undermine the clinical encounter.

The Communication Assessment Tool for Ethics Scale

Figures 1 and 2 present the structure of the ACT4Ethics scale.

Figure 1. Assessment Communication Tool for Ethics (ACT4Ethics), page 1.

Participants	Define the roles of all participants of the encounter (e.g.: caregiver, mother, father, family, friend, trainees and level of training, experienced or novice allied health professionals, subspecialist physician, translator, other …)

Did the learner …		Stage of Competency¶					🚩	Examples of Behaviors to Support Assessment and Corresponding Stages¶¶ *(please note that examples include positive and negative behaviors)*
		1	2	3	4	5		
Communication	… demonstrate verbal communication skills	☐	☐	☐	☐	☐	☐	Red flag — *Interrupts, uses a forceful voice or dominates conversation, causing patient/parents distress* 1-2 — Uses convention: introduces self and explains role with medical team, uses layman's terminology 3-4 — Aligns goals and maintains focus of the meeting with patient/parent wishes 4-5 — Builds and maintains rapport with patient and parents
	… demonstrate non-verbal communication skills	☐	☐	☐	☐	☐	☐	Red flag — *Inappropriately uses nonverbal cues such as closed or aggressive posture, abrupt exit/entrance, restlessness, or inappropriately frown/smile* 1-2 — Makes appropriate amount of eye contact, standing v/s sitting, leaning toward patient 2-3 — Uses adjusted voice intonation, active listening, shows attentiveness 3-4-5 — Responds to nonverbal cues (e.g., body language) from patient/parents and uses own nonverbal communication appropriately
	… demonstrate adaptability in the structure of the encounter	☐	☐	☐	☐	☐	☐	Red flag-1 — *Strictly adheres to checklist without flexibility* 2-3 — Integrates patient/parent expectations into agenda of encounter 4-5 — Adopts dynamic and flexible flow of conversation, following verbal cues from family, allows for moments of silence
	… close loops during and at the end of the encounter	☐	☐	☐	☐	☐	☐	Red flag — *Does not answer patient's or parents' questions, does not formulate a plan for next steps* 1-2 — Leaves an opening for future discussion 3-4 — Directly/clearly addresses patient's or parents' questions 3-4-5 — Periodically summarizes and uses transitions to facilitate understanding of patient/parents 4-5 — Closes encounter leaving family with realistic hope and expectations, but recognizing uncertainty
Civility and Respect	… demonstrate perspective-taking, cognitive empathy	☐	☐	☐	☐	☐	☐	Red flag — *Ignores or misses patient's concerns, queries. Disregards differences in values.* 1-2/ 2-3 — Identifies (1 – 2), articulates (2 – 3) patient's/parents' concerns 2-3-4 — Solicits and explores patient's/parents' values, beliefs, perspectives and preferences 4-5 — Addresses and validates patient's/parents' concerns, emotional features, values and norms
	… demonstrate compassion, affective empathy	☐	☐	☐	☐	☐	☐	Red flag — *Ignores patient's emotional reactions, or becomes inappropriately overcome with positive/negative emotion* 1-2/ 3-4 — Listens (1 – 2), display genuine emotional responses (3 – 4) 4-5 — Validates and supports patient's/parents' feelings and maintains emotional composure
	… demonstrate sensitivity to perspective difference, empathy diversity	☐	☐	☐	☐	☐	☐	Red flag — *Stereotypes patient/parents based on culture, religion or gender* 1-2/ 3-4 — Identifies (1–2), addresses (3–4) differences in perspectives in between patient, parents, physician and other participants 4-5 — In the context of patient's/parents' culture, religion or gender; interprets emotions, gestures and facial expressions appropriately
Clinical and Ethical Judgment and Decision-Making	… incorporate standards of practice	☐	☐	☐	☐	☐	☐	Red flag — *Acts in conflict with professional ethics* 1-2 — Identifies the standards of practice 3-4-5 — Acknowledges and appreciates the standards of practice in the context of the encounter
	… appreciate the complexity of the decision	☐	☐	☐	☐	☐	☐	Red flag — *Fails to incorporate patient's/parent's needs and perspectives* 1-2-3 — Demonstrates an understanding of the complexity, uncertainty, and ambiguity of the situation 2-3-4 — Acknowledges that patient's/parents' choices might not follow guidelines and evidence based medicine 4-5 — Weighs the evidence and conventions in clinical care with patient's/parents' needs
	… engage patient/parents in the decision	☐	☐	☐	☐	☐	☐	Red flag — *Deliberately influences patient or parents toward a decision without justification* 1-2 — Provides medical information that facilitate patient's/parents' participation 3-4-5 — Balances information provided on the specific needs and preferences of the family 3-4 — Incorporates patient's/parents' perspectives, values and religious beliefs into the decision-making process 4-5 — Engages and invites silent participants to verbalize their perceptions when appropriate
	… manage disagreements/ ethical conflicts	☐	☐	☐	☐	☐	☐	Red flag — *Ignores or fails to address disagreements between participants (e.g. healthcare providers, patient, parents)* 1-2/3-4 — Recognizes ethical conflicts (1 – 2) and articulates the basis of the conflict (3 – 4) 5 — While facilitating ethical resolution toward an agreement, incorporates the situation's specific characteristics and the patient's best interest, wishes and autonomy; stay open to the involvement of an interdisciplinary team
	… take steps toward an agreement	☐	☐	☐	☐	☐	☐	Red flag — *Fails to obtain agreement on next steps that will lead to a final decision* 1-2/3-4 — Summarizes the critical features of the encounter (1 -2) and obtains agreement on the next steps (3 -4) 4-5 — Leaves encounter with mutual understanding and by identifying common ground

Please refer to the instructions provided in next page to complete the scale.

XSL•FO
RenderX

Figure 2. Assessment Communication Tool for Ethics (ACT4Ethics), page 2.

Relevant situational factors that interfered with an ethically justified decision-making process	Not a concern	Moderate concern	Major concern
The physical environment of the encounter was not appropriately set-up prior to the meeting	☐	☐	☐
Uncontrollable environment factors (e.g., noise, interruption, lighting, temperature)	☐	☐	☐
Time limitations that were attributable to the clinical context	☐	☐	☐
Physician's physical or mental/cognitive fatigue (e.g.: post call, etc.)	☐	☐	☐
Other: (e.g., patient's emotions, reactions) _____	☐	☐	☐

Comments on other concerns not mentioned in the competency continuum above:

Comments on positive examples during the encounter:

OVERALL ASSESSMENT	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
At what stage of competency did the trainee perform?	☐	☐	☐	☐	☐

	No	Somewhat	Yes
Did the trainee perform at their current stage of training?	☐	☐	☐
Did the trainee perform in a manner that exceeded their expected stage of training?	☐	☐	☐

¶Definition of stages of competency		
Stage 1	*Entry to residency*	*Function at the level of a **pre-graduate trainee**.*
Stage 2	*Transition to discipline*	*This stage emphasizes the orientation and assessment of new trainees arriving from different medical schools and programs including outside of Canada. This stage corresponds to the **junior resident status in their first 2-3 months** of residency.*
Stage 3	*Foundation of discipline*	*This stage covers broad-based competencies that every trainee must acquire before moving on to more advanced, discipline-specific competencies. This stage continues to correspond with **junior resident status**.*
Stage 4	*Core of discipline*	*This stage covers the core competencies that make up the majority of a discipline. This stage corresponds to the early stage of **senior resident status**.*
Stage 5	*Transition to independent practice*	*At this stage the trainee should demonstrate readiness to make the transition to autonomous practice. This stage of residency continues to correspond with the later stage of **senior resident or junior attending** role.*

Instructions to complete the scale provided in previous page

🚩 **Bolded items** flag professional concerns that could be independent of the stage of competency scored: check box when important concerns arose during the encounter that jeopardized the relationship between the patient/parents and the medical team

¶ **Stages of competency evaluated on a scale 1 to 5** where stage "1" correspond to entry to residency, Stage 2 to transition to discipline, stage 3 to foundation of discipline, stage 4 to core of discipline, stage "5" correspond to transition to independent practice (see Table for more details)

¶¶ **In parenthesis**, a range of competency developmental stage a rater should observe is provided. The stage at which the behavior should be observed is **bolded**.

The ACT4Ethics scale includes 12 competencies aligned to the three domains included in the framework. The tool includes a 5-level rating scale for each competency with examples of milestones and stages in accordance with the continuum base by design model [11] to guide the rating. A red flag is added to the rating scale to allow the assessing clinicians to note concerns severe enough that a trainee should not currently be permitted to talk to patients or family members without supervision. The tool allows capturing the clinical context of the interaction and the relevant situational factors that might have justifiably interfered with the ethical judgment and decision-making processes. Such qualifications are necessary in order to account for any variability introduced by atypical features of clinical encounters likely to occur outside the simulation context. In the end, ACT4Ethics includes an overall assessment scale. Ideally, ratings on this overall scale should positively correlate with the averages obtained from the individual subscales.

The pilot study will present the tool without displaying the hypothesized targeted competencies depicted in subscale labels in order to ensure that raters will not be biased in assigning their ratings. For instance, while we have grouped certain items within a specific subscale, raters might assign ratings that do not support organizing ACT4Ethics in this manner.

Aim of the Study

The objective of this study is to pilot-test ACT4Ethics. This study will be divided into two parts—part A: training the raters

to use the tool and part B: obtaining validity evidence to support the scale construction [46].

Methods

Setting and Participants

The study will take place between February and June 2019 at the University of Ottawa Faculty of Medicine adjacent of the Children's Hospital of Eastern Ontario (CHEO) and the Ottawa Hospital General Campus. CHEO is a tertiary pediatric center that houses a 3B neonatal intensive care unit, and the Ottawa Hospital General Campus is equipped with a birthing unit and a level 3A neonatal intensive care unit. The University of Ottawa Faculty of Medicine's NPM training program is a 2- to 3-year residency accredited by the Royal College of Physicians and Surgeons of Canada.

The teaching tool used in the context of an Objective Structured Clinical Examination (OSCE) session is designed to test clinical skill performance that trainees in health care professions are expected to acquire [47]. The one-station OSCE will consist of a case focusing on antenatal findings of multiple congenital anomalies without a specific genetic syndrome identified. It will be divided into two parts for a total of 30 minutes. The first 15 minutes of the session will be videotaped and include one trainee and one standardized patient. One of the two coauthors (TD or EF) who directly observed the interaction will provide

confidential formative feedback in the second 15 minutes of the session, which will not be videotaped.

The pilot group will consist of a purposive sample [48] of former and current trainees from the NPM, Maternal-Fetal Medicine, Obstetrics, Pediatric Palliative Care, Genetics, and General Pediatrics residency programs (n=30). Our recruitment strategy intends to obtain a wide distribution of performances to ensure that raters will use the full range of the scale. We will recruit and train academic staff (n=15) to review the videos of the pilot group. Standardized patients (n=2) will be recruited to play-act the patient in the OSCE.

Ethical Considerations

Academic staff members from a variety of medical fields including collaborators who have participated in the creation of ACT4Ethics will also be piloting the tool. A research assistant or one of the authors (ALR) will approach academic staff, trainees, and standardized patients for consent. We will inform trainee participants that their choice will not affect their residency training assessment. Each academic staff participant will receive an incentive. The standardized patients, from the University of Ottawa Health Science Faculty, will be reimbursed for administration, training, and parking fees. The CHEO and Ottawa Health Science Network research ethics boards have approved this study.

Study Design

This pilot cohort study is part of a multiphase project designed according to modern validity principles—content (phases 1 and 2: creation of a blueprint), response process (phase 3: pilot test), internal structure, relation to other variables, and consequences (phase 4: implementation) [46]—where evidence that supports a particular interpretation of the results is collected [49,50]. The design also considers threats to validity such as construct underrepresentation and factors exerting nonrandom influence on scores.

Part A: Train the Raters

We will invite 15 academic staff members for training on ACT4Ethics and the GKCAS clinician/faculty form. A 45- to 60-minute video-based session developed by members of the project team (TD, EF, JS, and ALR) will be used to train academic staff/future raters on ACT4Ethics and help them achieve and promote consistency in assessment [51]. Future raters will test the tool on three short videos depicting a typical clinical scenario.

The short videos (approximatively 15 minutes in length) will be developed with the participation of one of the coauthors (ALR) acting out three levels of competence: transition to discipline, core of discipline, and advanced expertise [11], while interacting with a parent played by a standardized patient. A half-day session will involve at least three coauthors (TD, EF, ALR) or collaborators defining and standardizing the script for the videos. Each video will correspond to different levels of competence. They will include a number of key decision points where future raters should be able to identify and address specific features of a situation concerning communication (eg, ambiguous statements that require closing the loop),

decision-making (eg, essential diagnostic information not presented), and respect/empathy (eg, family behavior or personal/religious beliefs affecting interaction).

Part B: Obtain Validity Evidence

The pilot-test will be conducted using 30 distinct 15-minute videotaped interactions between a trainee and a standardized patient. We will ask the trainees to provide information on their previous experience in ethical encounters in their practice and ethical training, level of confidence in navigating an ethically charged situation, and basic demographics (eg, gender, age, year, and subspecialty of training). We will also ask them to complete an evaluation survey of the session.

Each trained rater will receive the videos of 6 different encounters, and they will use ACT4Ethics and GKCAS clinician/faculty form [27] to score trainee performance. The scores and debriefing will not be included in current trainees' final subspecialty training evaluation. After providing their ratings, trained raters will provide feedback on a standardized satisfaction survey. Their feedback will allow comparison of rater views between ACT4Ethics and the GKCAS clinician/faculty form.

Data Analysis

The research assistant or coauthor will enter all data into SPSS Statistics (IBM Corp). Descriptive statistics (mean and standard deviation or median and interquartile range) will be used to characterize participant demographics, responses to the survey, and the scores obtained with ACT4Ethics and compare them to the GKCAS clinician/faculty form.

Trainees' prior experience with regard to ethical situations will allow us to measure its effect on their performance and interaction with the standardized patient, as well as on their overall scores from the ACT4Ethics and GKCAS tools.

To compare the scales, we will examine (1) the order of individual participants in terms of their aggregate performance for each scale and (2) the correlation between items from each subscale that should show a high degree of correspondence (eg, items that assess empathy in the civility and respect subscale). We anticipate that the trained raters will use the complete range of response scale and have their ratings correspond to the issues presented in the training videos. Each rater will assess a total of 6 videos to determine the interrater reliability of ACT4Ethics. This process will result in 90 independent evaluations assessed by 15 trained raters with each encounter being evaluated three times—three nested scores within participants. We will conduct a generalizability analysis with rater nested within video and crossed with item. This analysis will allow us to determine the proportion of variance accounted for by each variable as well as allowing us to generate reliability coefficients related to both interrater reliability and the internal consistency of the items on the scale. We will use the conventional level of reliability of .8 or higher for the overall subscale to analyze our results. We will additionally obtain correlations of ratings for items between subscales and within a subscale. The scores from raters using both assessment instruments will allow an evaluation of the correlation between ratings.

Responses from the satisfaction survey completed by the trained raters will be used to assess the acceptance and usability of ACT4Ethics and explore facilitators and barriers for future implementation. We will review feedback on the experience and on the tool to adapt ACT4Ethics as appropriate.

Implementation Phase

The team will seek funding first to support the translation of ACT4Ethics into an electronic format ready for the one45 software platform, which supports the performance assessment of learners [52], and second, to facilitate the implementation of the tool in NPM, Pediatric, Maternal-Fetal Medicine, Medical Genetics, and Pediatric Palliative Care residency training programs. During this phase, the coauthors will continuously examine the construct validity of the assessment instrument.

Results

We anticipate this project will take a total of 18 months to complete and expect the results to be available by September 2019.

Discussion

Assessment of Communication Tool for Ethics to Maximize Learning of Communication Skills

Engaging in conversations with patients and families facing ethically challenging situations requires well-trained staff with advanced communication skills to support them. These skills also facilitate shared decision-making about the provision of care that is in the patient's best interest [53,54]. Physicians need to be taught these skills during their training. Without a well-designed assessment instrument specifically adapted to ethically sensitive situations, assessment of communication relies on subjectivity [55]. Such assessments will not maximize the learning of communication skills adapted to these encounters, leaving future physicians with underdeveloped skills potentially increasing the risk of conflict between patients, families, and the physician [56]. Although we are using an OSCE in the context of neonatology for the validation of our assessment instrument, the principles included in ACT4Ethics are also applicable in many other areas in health care professions. Our tool considers the specific affective, interconnected elements of verbal and nonverbal communication and cognitive features of ethically challenging clinical sets of circumstances encountered in medicine [24].

Assessment Scale Construct Validity

ACT4Ethics was created through a multiphase process to develop a valid and reliable assessment tool [57]. Using an integrated knowledge translation model [58] to facilitate the implementation phase of ACT4Ethics into different medicine subspecialty residency programs, the project engaged key collaborators from different medical fields [49] with expertise in physician-patient communication, clinical ethics, and/or medical education. They participated in focus groups, dyadic interviews, and Web-based surveys to define the goals, use, and content of the assessment tool.

The instrument was developed to be (1) comprehensive, (2) easily applied within the clinical context, and (3) capable of adapting to many subspecialties across the health care professions while providing expedient assessment. The overall goal of ACT4Ethics is to guide specific, meaningful feedback to facilitate trainee identification of the ethical affordances of communication regarding complex clinical situations that arise in NPM and many medical subspecialties.

During scale development, collaborators identified components of unique competencies that would enable clinicians to balance the demands of rigorous clinical decision-making with the need to ensure that they address the ethically sensitive features of these scenarios. While related, we think that each of the 12 competencies used within the scale should allow us to examine how they might independently contribute to a successful clinical encounter. As an important determinant of the construct validity of an assessment instrument [46], each of the three subscales should assess distinct features of a clinical scenario and show minimal overlap in terms of the ratings. For instance, items on ACT4Ethics that assess a learner's civility and respectful behavior should only show a weak relationship with items related to clinical judgment and communication. Ratings for items within individual subscales (eg, all communication items) should correlate with each other. Differentiating responses for items within one subscale from those in a different subscale will validate the internal structure of ACT4Ethics. Ratings should increase following a training session in a pre- and posttest design. Such a pattern would suggest that the scale captures the response processes associated with trainee acquisition of these skills [46].

Our study will attempt to demonstrate how ACT4Ethics is an improvement over other communication assessment scales. For instance, while ACT4Ethics and GKCAS [28] both assess empathy, our scale breaks down empathy in cognitive (eg, perspective-taking) [39] and affective (eg, feelings) [40] components, whereas GKCAS does not make this distinction. These constructs reflect an important distinction that is likely to be relevant in clinical practice: while a clinician might be able to understand a family member's concerns, they might not be adequately emotionally responsive. Similarly, while GKCAS addresses relationship building, this construct is somewhat ambiguous. We instead assume that relationship building should be assessed by looking at the clinician's ability to adapt, demonstrate sensitivity to differences in perspective by different individuals within a clinical encounter, engage family members in the decision, and manage disagreement and ethical conflicts. Our analyses will examine how ACT4Ethics and GKCAS relate to one another.

The logic and thoroughness behind the creation of ACT4Ethics support our belief that it will improve the assessment of trainees and positively contribute to providing objective, meaningful feedback on communication skills during challenging ethical situations. Nonspecific in-training rotation evaluations and other tools lack this type of feedback. By assessing competencies associated with communication skills, civility and empathy, along with clinical and ethical judgment and decision-making, we should be able to identify specific strengths and weaknesses of a particular trainee. Feedback from the ACT4Ethics scale

can then be used by learners to promote self-reflection and strengthen their competencies in specific domains of practice. Overall, trends in ACT4Ethics scores of multiple trainees could allow program directors to readily identify teaching deficiencies and allow them to adjust their curriculum accordingly.

Curriculum and Accreditation Needs

ACT4Ethics will assist with local curriculum and program development. We are currently evaluating the effect of our local neonatal ethics teaching program. Our evaluation follows Guskey's levels of training program evaluation [13]. We have already evaluated reactions (Guskey level 1) and demonstrated that the neonatal ethics teaching program sessions were well-received by trainees and participants, with an overall satisfaction score of 5.8 out of 7 [2]. Currently, we are running our pre- and posttraining knowledge test with our ongoing cohorts to evaluate the acquisition of knowledge related to ethics (Guskey level 2) [18]. The support received from the University of Ottawa to implement the program and develop our tools and the expansion to other residency programs indicates that the organization supports our model (Guskey level 3: organization support and change). ACT4Ethics will allow our research team to evaluate learning and use of skills (Guskey level 4). For example, this tool may help to evaluate the efficacy of neonatal ethics teaching programs and other specific teaching strategies to improve learning and use of both knowledge and skills to navigate ethically challenging clinical situations. We believe that this tool can also provide other education and research teams with a means to evaluate the success of their educational interventions and programs.

The ACT4Ethics scale addresses a need identified by accreditation bodies. With an overarching goal of contributing to and improving the means through which the milestones included in the Royal College of Physicians and Surgeons Canada 2015 CanMEDS roles of communicator, professional, and collaborator are assessed, ACT4Ethics will allow mentors and supervisors to assess and monitor competency levels of trainees. Although the validation evidence will be obtained using the context of neonatology, the creation of videos depicting typical ethically sensitive scenarios encountered in other areas in health care professions can be used to demonstrate its generalizability. We anticipate that ACT4Ethics will guide clinician-teacher supervisors, making explicit which communication skills to assess and how to rate them according to a clear range of competence. The tool can provide trainees with a clear sense of what the expectations are for their communication skills during difficult and ethically sensitive conversations. Consistent and relevant feedback is more likely to affect trainees' learning of skillful behaviors and communication skills [59], potentially improving patients' and parents' satisfaction relationship with the physician-trainee.

Acknowledgments

Authors' Contributions

We thank the following collaborators who participated in the development of the communication assessment tool scale: Nadya Ben Fadel, MD, Neonatology, Children's Hospital of Eastern Ontario; Karen Fung Kee Fung, MD, Maternal-Fetal Medicine, The Ottawa Hospital; Sarah Lawrence, MD, Neonatology, Children's Hospital of Eastern Ontario; Anna-Theresa Lobos, MD, Pediatric Intensive Care, Children's Hospital of Eastern Ontario; Gillian MacLean, MD, Fellow, Neonatology, Children's Hospital of Eastern Ontario; Paul Muirhead, LLM, LLB; Julie Richer, MD, Pediatric Genetics, Children's Hospital of Eastern Ontario; and Chris Vadeboncoeur, MD, Palliative Care Medicine, Children's Hospital of Eastern Ontario.

Conflicts of Interest

None declared.

References

1. Stokes TA, Watson KL, Boss RD. Teaching antenatal counseling skills to neonatal providers. Semin Perinatol 2014 Feb;38(1):47-51. [doi: 10.1053/j.semperi.2013.07.008] [Medline: 24468569]

2. Daboval T, Moore G, Ferretti E. How we teach ethics and communication during a Canadian neonatal perinatal medicine residency: an interactive experience. Med Teach 2013;35(3):194-200. [doi: 10.3109/0142159X.2012.733452] [Medline: 23102158]

3. Boss RD, Urban A, Barnett MD, Arnold RM. Neonatal Critical Care Communication (NC3): training NICU physicians and nurse practitioners. J Perinatol 2013 Aug;33(8):642-646. [doi: 10.1038/jp.2013.22] [Medline: 23448940]

4. Cuttano A, Scaramuzzo RT, Gentile M, Ciantelli M, Sigali E, Boldrini A. Education in neonatology by simulation: between reality and declaration of intent. J Matern Fetal Neonatal Med 2011 Oct;24 Suppl 1:97-98. [doi: 10.3109/14767058.2011.607572] [Medline: 21942603]

5. Tongue JR, Epps HR, Forese LL. Communication skills. Instr Course Lect 2005;54:3-9. [Medline: 15948430]

6. Emanuel LL, Alpert HR, Baldwin DC, Emanuel EJ. What terminally ill patients care about: toward a validated construct of patients' perspectives. J Palliat Med 2000;3(4):419-431. [doi: 10.1089/jpm.2000.3.4.419] [Medline: 15859694]

7. Papageorgiou A. Medical & ethical issues related to extreme prematurity. J Perinat Med 2013;41(s1):121. [doi: https://doi.org/10.1515/jpm-2013-2001]

8. Cummings CL. Teaching and assessing ethics in the newborn ICU. Semin Perinatol 2016 Dec;40(4):261-269. [doi: 10.1053/j.semperi.2015.12.016] [Medline: 26831209]

9. Daboval T, Ferretti E, Moore G. Innovative holistic teaching in a Canadian Neonatal Perinatal Residency Program. Hastings Cent Rep 2014;44(6):21-25. [doi: 10.1002/hast.384] [Medline: 25412973]

10. Ferretti E, Rohde K, Moore G, Daboval T. Catch the moment: the power of turning mistakes into "precious" learning opportunities. Paediatr Child Heal 2018:4. [doi: 10.1093/pch/pxy102/5074051]

11. Frank J, Snell L, Sherbino J. Royal College of Physicians and Surgeons of Canada. 2015. CanMEDs 2015: Physician Competency Framework URL: http://canmeds.royalcollege.ca/uploads/en/framework/CanMEDS%202015%20Framework_EN_Reduced.pdf [accessed 2019-04-18] [WebCite Cache ID 77jl9nWcr]

12. Accreditation Council for Graduate Medical Education. 2016. ACGME program requirements for graduate medical education in the subspecialties of pediatrics URL: https://www.acgme.org/Portals/0/PFAssets/ProgramRequirements/329_neonatal-perinatal_medicine_2017-07-01.pdf [accessed 2019-04-18] [WebCite Cache ID 77jlMW19i]

13. Guskey T. Evaluating Professional Development. Thousand Oaks: Sage Publishing; 2000.

14. Daboval T, Ferretti E, Moussa A, van Manen M, Moore G, Srinivasan G, et al. Needs assessment of ethics and communication teaching for neonatal perinatal medicine programs in Canada. Paediatr Child Health 2018 Aug 28:1-9. [doi: 10.1093/pch/pxy108]

15. Sulmasy DP, Geller G, Levine DM, Faden R. Medical house officers' knowledge, attitudes, and confidence regarding medical ethics. Arch Intern Med 1990 Dec;150(12):2509-2513. [Medline: 2244767]

16. Kesselheim JC, McMahon GT, Joffe S. Development of a Test of Residents' Ethics Knowledge for Pediatrics (TREK-P). J Grad Med Educ 2012 Jun;4(2):242-245 [FREE Full text] [doi: 10.4300/JGME-D-11-00280.1] [Medline: 23730449]

17. Cummings CL, Geis GM, Feldman HA, Berson ER, Kesselheim JC. Assessing ethics knowledge: development of a test of ethics knowledge in neonatology. J Pediatr 2018 Dec;199:57-64. [doi: 10.1016/j.jpeds.2018.04.004] [Medline: 29754867]

18. Moore GP, Ferretti E, Daboval T. Developing a knowledge test for a neonatal ethics teaching program. Cureus 2017 Dec 20;9(12):e1971 [FREE Full text] [doi: 10.7759/cureus.1971] [Medline: 29492360]

19. Browning DM, Solomon MZ. Relational learning in pediatric palliative care: transformative education and the culture of medicine. Child Adolesc Psychiatr Clin N Am 2006 Jul;15(3):795-815. [doi: 10.1016/j.chc.2006.03.002] [Medline: 16797450]

20. Ellman MS, Fortin AH. Benefits of teaching medical students how to communicate with patients having serious illness: comparison of two approaches to experiential, skill-based, and self-reflective learning. Yale J Biol Med 2012 Jun;85(2):261-270 [FREE Full text] [Medline: 22737055]

21. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [Medline: 2400509]

22. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. Acad Med 2001 Apr;76(4):390-393. [Medline: 11299158]

23. Kurtz S, Silverman J, Draper J. Teaching and Learning Communication Skills in Medicine. Oxford: Radcliffe Medical Press; 1998.

24. Singer PA, Robb A, Cohen R, Norman G, Turnbull J. Performance-based assessment of clinical ethics using an objective structured clinical examination. Acad Med 1996 May;71(5):495-498. [Medline: 9114869]

25. Lynch DC, Surdyk PM, Eiser AR. Assessing professionalism: a review of the literature. Med Teach 2004 Jun;26(4):366-373. [doi: 10.1080/01421590410001696434] [Medline: 15203852]

26. Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. SPIKES—a six-step protocol for delivering bad news: application to the patient with cancer. Oncologist 2000;5(4):302-311 [FREE Full text] [Medline: 10964998]

27. Peterson EB, Calhoun AW, Rider EA. The reliability of a modified Kalamazoo Consensus Statement Checklist for assessing the communication skills of multidisciplinary clinicians in the simulated environment. Patient Educ Couns 2014 Sep;96(3):411-418. [doi: 10.1016/j.pec.2014.07.013] [Medline: 25103180]

28. Calhoun AW, Rider EA, Meyer EC, Lamiani G, Truog RD. Assessment of communication skills and self-appraisal in the simulated environment: feasibility of multirater feedback with gap analysis. Simul Healthc 2009;4(1):22-29. [doi: 10.1097/SIH.0b013e318184377a] [Medline: 19212247]

29. Daboval T, Shidler S. Ethical framework for shared decision making in the neonatal intensive care unit: communicative ethics. Paediatr Child Health 2014 Jun;19(6):302-304 [FREE Full text] [Medline: 25332659]

30. Rest J, Thoma S, Edwards L. Designing and validating a measure of moral judgment: stage preference and stage consistency approaches. J Educ Psychol 1997;89(1):5-28. [doi: 10.1037/0022-0663.89.1.5]

31. Daboval T, Shidler S, Thomas D. Shared decision making at the limit of viability: a blueprint for physician action. PLoS One 2016;11(11):e0166151 [FREE Full text] [doi: 10.1371/journal.pone.0166151] [Medline: 27893823]

32. Kurtz S, Silverman J, Benson J, Draper J. Marrying content and process in clinical method teaching: enhancing the Calgary-Cambridge guides. Acad Med 2003 Aug;78(8):802-809. [Medline: 12915371]

33. Brindley PG, Reynolds SF. Improving verbal communication in critical care medicine. J Crit Care 2011 Apr;26(2):155-159. [doi: 10.1016/j.jcrc.2011.03.004] [Medline: 21482347]

34. Andersson L, Pearson C. Tit for tat? The spiraling effect of incivility in the workplace. Acad Manag Rev 1999 Jul;24(3):452-455. [doi: 10.2307/259136]

35. Cortina LM, Magley VJ, Williams JH, Langhout RD. Incivility in the workplace: incidence and impact. J Occup Health Psychol 2001 Jan;6(1):64-80. [Medline: 11199258]

36. Guidroz AM, Burnfield-Geimer JL, Clark O, Schwetschenau HM, Jex SM. The nursing incivility scale: development and validation of an occupation-specific measure. J Nurs Meas 2010;18(3):176-200. [Medline: 21290924]

37. Hutton S, Gates D. Workplace incivility and productivity losses among direct care staff. AAOHN J 2008 Apr;56(4):168-175. [Medline: 18444405]

38. Papastavrou E, Efstathiou G, Tsangari H, Suhonen R, Leino-Kilpi H, Patiraki E, et al. Patients' and nurses' perceptions of respect and human presence through caring behaviours: a comparative study. Nurs Ethics 2012 May;19(3):369-379. [doi: 10.1177/0969733011436027] [Medline: 22581506]

39. Eslinger PJ. Neurological and neuropsychological bases of empathy. Eur Neurol 1998;39(4):193-199. [doi: 10.1159/000007933] [Medline: 9635468]

40. Davis M. Empathy: A Social Psychological Approach. New York: Taylor & Francis; 1994.

41. Shamay-Tsoory SG, Aharon-Peretz J. Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. Neuropsychologia 2007 Oct 01;45(13):3054-3067. [doi: 10.1016/j.neuropsychologia.2007.05.021] [Medline: 17640690]

42. Shamay-Tsoory SG, Aharon-Peretz J, Perry D. Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. Brain 2009 Mar;132(Pt 3):617-627. [doi: 10.1093/brain/awn279] [Medline: 18971202]

43. Endsley MR. Toward a theory of situation awareness in dynamic systems. Hum Factors 1995 Mar 01;37(1):32-64. [doi: 10.1518/001872095779049543]

44. Kahneman D. Thinking, Fast and Slow. New York: Anchor Canada; 2011.

45. Klein G. Sources of Power: How People Make Decisions. Cambridge: The MIT Press; 1999.

46. Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol 1995;50(9):741-749. [doi: 10.1037//0003-066X.50.9.741]

47. Harden RM, Stevenson M, Downie WW, Wilson GM. Assessment of clinical competence using objective structured examination. Br Med J 1975 Feb 22;1(5955):447-451 [FREE Full text] [Medline: 1115966]

48. Creswell J, Clark V. Designing and Conducting Mixed-Methods Research. 2nd Edition. Thousand Oaks: Sage Publications; 2011.

49. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med 2006 Feb;119(2):166.e7-166.e16. [doi: 10.1016/j.amjmed.2005.10.036] [Medline: 16443422]

50. Downing SM. Validity: on meaningful interpretation of assessment data. Med Educ 2003 Sep;37(9):830-837. [Medline: 14506816]

51. Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence: a randomized trial. Ann Intern Med 2004 Jun 01;140(11):874-881. [Medline: 15172901]

52. one45. URL: https://www.one45.com/ [accessed 2019-04-18] [WebCite Cache ID 77jmgfzxW]

53. Stirrat GM. Education in ethics. Clin Perinatol 2003 Mar;30(1):1-15. [Medline: 12696782]

54. Janvier A, Barrington K, Farlow B. Communication with parents concerning withholding or withdrawing of life-sustaining interventions in neonatology. Semin Perinatol 2014 Feb;38(1):38-46. [doi: 10.1053/j.semperi.2013.07.007] [Medline: 24468568]

55. O'Donnell JA, Oakley M, Haney S, O'Neill PN, Taylor D. Rubrics 101: a primer for rubric development in dental education. J Dent Educ 2011 Sep;75(9):1163-1175 [FREE Full text] [Medline: 21890846]

56. Levinson W, Roter DL, Mullooly JP, Dull VT, Frankel RM. Physician-patient communication. The relationship with malpractice claims among primary care physicians and surgeons. JAMA 1997 Feb 19;277(7):553-559. [Medline: 9032162]

57. Hamstra S. Keynote address: the focus on competencies and individual learner assessment as emerging themes in medical education research. Acad Emerg Med 2012 Dec;19(12):1336-1343 [FREE Full text] [doi: 10.1111/acem.12021] [Medline: 23279242]

58. Lavis JN, Robertson D, Woodside JM, McLeod CB, Abelson J, Knowledge Transfer Study Group. How can research organizations more effectively transfer research knowledge to decision makers? Milbank Q 2003;81(2):221-248 [FREE Full text] [Medline: 12841049]

59. Allen D, Tanner K. Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. CBE Life Sci Educ 2006;5(3):197-203 [FREE Full text] [doi: 10.1187/cbe.06-06-0168] [Medline: 17012210]

Abbreviations

NPM: neonatal-perinatal medicine
GKCAS: Gap-Kalamazoo Communication Skills Assessment Form
ACT4Ethics: Assessment of Communication Tool for Ethics

CHEO: Children's Hospital of Eastern Ontario
OSCE: Objective Structured Clinical Examination

XSL·FO
RenderX

Protocol

Household Surveys in the General Population and Web-Based Surveys in IQOS Users Registered at the Philip Morris International IQOS User Database: Protocols on the Use of Tobacco- and Nicotine-Containing Products in Germany, Italy, and the United Kingdom (Greater London), 2018-2020

Zheng Sponsiello-Wang[1], PhD; Peter Langer[1], PhD; Luis Prieto[1], PhD; Mariia Dobrynina[1], PhD; Dimitra Skiada[1], PhD; Nathalie Camille[1], PhD; Rolf Weitkunat[1], PhD; Frank Lüdicke[1], MD

Philip Morris Products SA, Neuchatel, Switzerland

Corresponding Author:
Zheng Sponsiello-Wang, PhD
Philip Morris Products SA
Quai Jeanrenaud 5
2000 Neuchatel,
Switzerland
Phone: 41 582422129
Fax: 41 582422811
Email: zheng.wang@pmi.com

Abstract

Background: Philip Morris International (PMI) has developed a novel heat-not-burn tobacco product, Tobacco Heating System (THS), which is marketed under the brand name of IQOS with HEETS (IQOS). The aerosol generated by THS has substantially fewer toxicants than combustible cigarette smoke, although the extent of the reduction of harmful and potentially harmful constituents reported varies between studies. To evaluate the potential harm reduction associated with IQOS use, the assessment of the uptake and continued use of IQOS in the context of all other tobacco- and nicotine-containing products is crucial. In March 2018, PMI launched cross-sectional surveys in Germany, Italy, and the United Kingdom (Greater London) to estimate the prevalence and use patterns of IQOS and other tobacco- and nicotine-containing products use in these 3 markets following the commercialization of IQOS. This study describes the protocol of the surveys.

Objective: The objectives of these surveys are to estimate the prevalence of tobacco- and nicotine-containing products use; describe past and current patterns of use; and explore their associations with self-reported health, motivation to use, risk perceptions, and perceived aesthetic changes.

Methods: The overall design of the surveys is similar in all 3 countries. Repeated cross-sectional surveys are being conducted annually over 3 consecutive years (2018 to 2020) and in 2 samples: a representative sample of the general population and a sample of IQOS users. A total of 6085 adults per year will be selected from the general population for each survey through multistage stratified sampling, and participants will respond to face-to-face computer-assisted personal interviews. In addition, 1404, 1384, and 1246 IQOS users per year in Germany, Italy, and Greater London, respectively, will be randomly selected from the PMI IQOS user database and will be invited to complete the Web-based survey using computer-assisted self-interviews. The Smoking Questionnaire is used to assess the tobacco use patterns of the participants.

Results: The recruitment of the general population sample began in March 2018 and that of the IQOS user sample began in April 2018. The data collection is ongoing, and the results of the first year data analysis are expected to be available by June 2019.

Conclusions: As the design of the 3 surveys is similar, the results will allow for cross-countries comparison of the prevalence of IQOS and other tobacco- and nicotine-containing products use as well as patterns of use and associated factors.

International Registered Report Identifier (IRRID): DERR1-10.2196/12061

XSL•FO
RenderX

KEYWORDS

cross-sectional survey; prevalence; tobacco products; tobacco heating system (THS); IQOS

Introduction

Background

Tobacco smoking is the leading cause of preventable disease and is responsible for about 6 million deaths across the world each year [1]. Smokers are far more likely than nonsmokers to get heart disease, lung cancer, chronic obstructive pulmonary disease, and other diseases [2]. For smokers, the best way to reduce the harm and risks of smoking-related disease is to quit. However, nicotine is addictive, and smoking cessation has proven difficult for many smokers [2,3]. Although there is little evidence that nicotine itself causes smoking-related diseases when decoupled from smoke [4,5], in getting nicotine from smoking, smokers are exposed to enormous harm because of the inhalation of toxic components resulting from the combustion of cigarettes. When a smoker does not want to stop all nicotine use, harm minimization implies striving for the complete elimination of smoked tobacco exposure by substituting it with the use of less harmful noncombusted forms of nicotine instead of smoking [4]. Tobacco harm minimization and reduction have been increasingly recognized by the scientific society and health authorities as a valuable and promising strategy to decrease smoking-related population harm [4,6]. One of the Food and Drug Administration's (FDA) new tobacco strategy is to recognize and clarify the role that potentially less harmful tobacco products could play in improving public health [4,5].

Toward this end, Philip Morris International (PMI) has developed a novel heat-not-burn product, Tobacco Heating System (THS). THS was designed to generate an aerosol that has substantially fewer toxicants than combustible cigarette smoke by heating the tobacco at a temperature that avoids combustion. To assess the potential health benefit of THS, a multistep assessment plan has been developed by PMI to demonstrate that (1) THS reduces harm and the risk of tobacco-related disease to individual tobacco users and (2) THS benefits the health of the population as a whole. To address the individual risk reduction associated with THS use, premarket assessment studies including aerosol chemistry, in vitro and in vivo toxicity studies, and clinical studies coherently demonstrate that switching from cigarette smoking to THS use consistently and significantly reduces the exposure to harmful and potentially harmful constituents (HPHCs) to levels that approach the reductions associated with smoking abstinence [7-15]. Numerous independent studies [16-19] and government-affiliated labs including the FDA's Southeast Tobacco Laboratory [20], UK Committee on Toxicity [21], German Federal Institute for Risk Assessment (BfR) [22], Japan National Institute of Public Health [23], New Zeeland CRL Energy Ltd [24], and Dutch National Institute for Public Health and Environment (RIVM: Rijksinstituut voor Volksgezondheid en Milieu) [25] have evaluated the aerosol chemistry of THS and confirmed the lower levels of HPHCs in the emissions of THS. Public Health of England [26] has systematically reviewed the existing evidence on THS and concluded that THS is likely to expose users to lower levels of HPHCs. However, higher quantities of several components other than HPHCs have been reported [27]. Auer et al [28] reported a much higher concentration of the polycyclic aromatic hydrocarbon acenaphthene for THS relative to cigarettes. According to the FDA (FDA briefing document, Meeting of the Tobacco Products Scientific Advisory Committee, 2018), the data published by Auer et al [28] are not considered adequate for comparing the levels of HPHCs between THS and combustible cigarettes. Personal communication between FDA reviewers and the paper's authors indicated that Auer et al [28] used a smoking device designed in their facility to capture mainstream aerosol with a modified International Organization for Standardization smoking regimen to perform the analysis. In addition, the identity of some of the compounds, such as acenaphthene, cannot be confirmed as the method used is not selective. PMI premarket consumer perception and behavior evaluations on the intensity of THS use, such as abuse liability, product appeal, and consumer perception, have shown that THS is not only an acceptable alternative to cigarettes for at least part of the adult cigarette smoker population but is also properly understood by adult smokers and nonsmokers [29,30]. To address THS benefit to the health of the population as a whole, a comprehensive postmarket program including cross-sectional surveys has been established to collect relevant data on the prevalence of products use, patterns of use, and product perception. One key determinant of population harm is the assessment of the prevalence, uptake, and continued use of THS and in the context of all other tobacco- and nicotine-containing products [4,31,32]. It has been pointed out that population net exposure to harmful toxicants depends on the actual patterns and prevalence of product use [4]. The potential population health benefit of THS use would be achieved if, in addition to reduced toxicity, THS is widely accepted by smokers, does not attract nontobacco users, and does not negatively influence smokers who intend to quit.

THS, which is commercialized under the brand name of IQOS with HEETS (IQOS), was first launched in Nagoya, Japan, in 2014. Shortly after national expansion and commercialization in Japan in 2016, PMI initiated a 3-year cross-sectional survey to assess tobacco use prevalence and patterns of tobacco product use in the Japanese population [33]. The first-year survey results show that among the Japanese population, the prevalence of current use was 18.5% for any tobacco product, 17.6% for cigarettes, 1.8% for IQOS, and 0.7% for electronic cigarettes (e-cigarettes) [34]. IQOS is commercialized in more than 40 markets worldwide at present. In Italy and Germany, the product was launched in November 2014 and June 2016, respectively. In the United Kingdom, IQOS was launched in Greater London in November 2016 as a test market. To estimate the prevalence of tobacco use including IQOS and assess product use patterns in these 3 markets, PMI launched the repeated cross-sectional surveys in 2018. As in the Japanese survey, these cross-sectional surveys are being conducted in 2 samples: a representative sample of the general population and a sample of IQOS users.

Objectives

The objectives of these surveys are to (1) estimate the prevalence of current tobacco- and nicotine-containing products use including IQOS; (2) describe product use patterns, that is, never use, initiation, product use transition, cessation, reinitiation, and relapse; (3) explore the associations between self-reported health status and use of tobacco- and nicotine-containing products in the general population as well as in a targeted sample of IQOS users in each country; and (4) among the targeted samples of IQOS users to explore the associations between patterns of tobacco- and nicotine-containing products use (including misuse) and motivation to use novel tobacco products, perceived quality attributes of IQOS (eg, risk aesthetic changes), and consumer satisfaction.

Methods

Survey Setting

The surveys are performed in accordance with ethical principles that have their origin in the Declaration of Helsinki [35] and are consistent with Good Epidemiological Practice [36] and International Ethical Guidelines for Epidemiological Studies [37]. Before the start of the surveys, a confirmation that approval is not required according to local laws has been obtained from the ethics committee of each market.

In all 3 markets, the cross-sectional surveys are being conducted annually in 2 samples over 3 consecutive years from 2018 to 2020: a representative sample of the general population and an IQOS user sample drawn from registered users in a PMI IQOS user database. The general population sample serves to estimate the prevalence of current tobacco- and nicotine-containing products use. The latter is chosen to be able to describe the patterns of IQOS use with an acceptable precision that is deemed to be hardly possible, with an anticipated low IQOS use prevalence in the representative general population sample shortly after product launch.

General Population Sample

The general population samples in the 3 markets are adults older than 18 years, living in Germany, Italy, and Greater London. The inclusion criteria are listed in Table 1. In Germany and Greater London, the subjects are randomly sampled from the general population in 3 steps: area sampling (primary sampling point), household selection, and selection of target persons (Table 1).

In Germany, census data of the Work Group of German Market and Social Research Institutes (Arbeitskreis Deutscher Markt- und Sozialforschungsinstitute e. V. ADM-Sampling-System for face-to-face surveys) are used to draw a stratified random sample. Using the ADM-Sampling system, a stratified random sample of 258 sample points is selected, taking into account the region and size of the community. From each sample point, the target households are randomly selected. Within each household, the target respondent is identified through the next birthday method, that is, the person in the household whose birthday is next.

The primary sampling points in Greater London are output areas (OA; the smallest administrative unit). A stratified random sample of 305 OAs is selected with probability proportionate to size among 32 Boroughs. From each OA, a total of 55 addresses are randomly chosen from the Postal Address File, a database that contains all known "Delivery Points" and postcodes in the United Kingdom. The target respondent in the household is selected based on the next birthday method.

In Italy, a stratified random probability sample is drawn according to a set of 3 different subpopulations or strata: (1) municipalities (140 in total), (2) electoral wards within the municipalities, and (3) individuals within the lists of the electoral wards. The sizes of the random samples from each stratum are proportional to the strata population sizes (proportionate stratification).

For each market, 6085 participants per year are to be interviewed. On the basis of initial results from the Japanese cross-sectional survey [33], assuming 1% IQOS uptake in each market, this sample size is sufficient to estimate the prevalence with a 95% confidence and a precision of ±0.25 percent points.

In Germany and Italy, the surveys are conducted as part of a face-to-face computer-assisted personal interviews (CAPI) omnibus. The omnibus is a syndicated survey with multiple participating clients, the questionnaires being divided into sections. Annually, the data are collected over 6 to 7 waves, with over 1000 participants per wave.

In Greater London, face-to-face interviews of the target participants are performed with CAPI. The annual data collection comprises 4 waves with a minimum of 1250 interviews per wave.

Table 1. Inclusion criteria and sample frames of the general population and IQOS users in cross-sectional surveys in Germany, Italy, and Greater London.

Samples	Sample frame	Inclusion criteria	Survey methods
General population	Area sampling (primary sampling point): 258 regions sample points in Germany, 140 municipalities in Italy, and 305 output areas in London; household (Germany and Greater London) or electoral ward selection (Italy) within the primary sampling point; and identify target persons from the household (Germany and Greater London) or from the lists of the electoral wards (Italy)	Aged ≥18 years; currently residing in Germany, Italy, or Greater London; able to read, write, and understand German, Italian, or English; and consent to participate in the survey	Face-to-face omnibus computer-assisted personal interviews (Germany and Italy) and face-to-face computer-assisted personal interviews (Greater London)
IQOS users	Random sample from PMI[a] Germany, Italy, and UK IQOS user database	In addition to the above criteria, has used more than 100 THS[b] tobacco sticks in his or her lifetime, is currently using IQOS, has access to the internet, and is currently not employed by PMI or any of its affiliates	Web-based; computerized assisted self-interviews

[a]PMI: Philip Morris International.

[b]THS: Tobacco Heating System.

IQOS User Sample

In all 3 markets, the PMI IQOS user database consists of adult smokers who purchased the IQOS devices either via a website, by telephone, or from IQOS retail stores; registered their IQOS devices in the database; and are willing to be contacted for market research purposes. IQOS users are randomly sampled from the database among those who did not participate in PMI market research panel. The participants are invited to participate in the study by email with a link to the study questionnaire. The Web-based surveys on the IQOS user sample are conducted with computerized assisted self-interviews. The inclusion criteria are listed in Table 1. As a token of appreciation, the participants receive a gift voucher of EUR 10 or GBP 10.

On the basis of initial results from the Japanese cross-sectional survey [33] that 63.4% of IQOS users fully converted to exclusive IQOS use, 1404, 1384, and 1246 IQOS users per year in Germany, Italy, and Greater London, respectively, will be surveyed. This sample size will be sufficient to estimate the proportion of fully converted IQOS users with a 95% confidence and a precision of ± 2.5 percent points.

The annual data collection of IQOS user samples is conducted in 4 waves coinciding with the waves for the general population survey.

Survey Questionnaire

The survey questionnaire (Multimedia Appendices 1 and 2), consisting of 3 parts, was developed to address the survey objectives.

Part 1 comprises a multidimensional Smoking Questionnaire (SQ) [38,39], which was developed to standardize the assessment of cigarette smoking exposure covering the major dimensions of cigarette smoking. SQ is used to assess cigarette smoking history and behavior. Test-retest reliability of the SQ and concurrent validity with the Behavior Risk Factor Surveillance System 2011 questionnaire have demonstrated that the SQ is reliable and easy to use [39]. Most importantly, the information collected using SQ allows for classification of

smoking behavior and history according to the World Health Organization (WHO) definitions [40]. It can also address dimensions of smoking behavior and history that are not covered by WHO definitions, for example, on the duration of smoking or cessation, on smoking intensity, and on smoking patterns, with regard to occasional smoking. In addition, the SQ is flexible, and it can be easily adapted to assess the exposure to other novel tobacco products such as IQOS, e-cigarettes, PLOOM, and glo. To ensure that the information related to each product is correctly understood, a description together with the pictures of each product is presented to the participants before displaying the questionnaires. Directly after the SQ for cigarette smoking, questions related to IQOS use are asked. Before assessing other novel tobacco- and nicotine-containing products, a specific question is asked to distinguish between e-cigarettes and other products such as PLOOM and glo. Information on the current use behavior regarding smokeless tobacco and other tobacco- or nicotine-containing products such as cigars, cigarillos, pipes, and water-pipes as well as nicotine-replacement therapy (NRT) products is also collected. In addition, first product use, age of initiation, and quit attempt are included in the survey questionnaire. The motivation to use IQOS and/or any other emerging tobacco products is collected according to the German Study on Tobacco Use (DEBRA: "Deutsche Befragung zum Rauchverhalten") questionnaire [41].

The second part assesses self-reported health status, which is included to explore associations with patterns of tobacco product use. Participants are asked to rate the overall state of their health, how their health compares with the average person of their age, and how much they are worried about their health [42]. The type and extent of comorbidities are measured with a validated instrument for epidemiological studies based on self-reported outcomes [43]. In addition, the Self-Reported Changes Questionnaire, which records changes since starting using the product in a number of relevant domains where IQOS may have potential benefits (eg, teeth coloring, breath smell, exercise capacity, and skin appearance), is included for IQOS user survey.

XSL•FO

RenderX

The third part is specific to assess IQOS users' use experience and perceived risk. The degree to which participants experience the reinforcing effects of using IQOS compared with cigarette smoking, such as satisfaction, psychological rewards, aversion, enjoyment of respiratory tract sensations, and craving reduction, is assessed with the Tobacco/Nicotine-containing Product Evaluation Questionnaire (ToNiPEQ). ToNiPEQ is an adapted version of the validated Modified Cigarette Evaluation Questionnaire [44], as proposed for smokeless tobacco [45]. Information on the type and frequency of potential misuse of THS tobacco sticks and frequency of potential misuse of the IQOS electronic device is collected with a misuse questionnaire developed by PMI. The perceived health risk of getting 18 diseases/conditions associated with smoking and using IQOS is assessed with the Perceived Risk Instrument-general version (PRI-G). PRI-G is a self-report instrument with a unidimensional scale that has been developed by PMI based on Rasch Measurement Theory and traditional psychometric methods [46].

Along with the above 3 components, the demographic characteristics of the participants including age, sex, income, education, occupation, and ethnicity are collected. In Greater London, social grade based on occupation, which is widely used for social classification in both official statistics and academic research, is rated by the interviewer for the general population survey.

In all 3 countries, the survey questionnaires are translated into local languages to minimize systematic biases because of differences in linguistic expression.

Outcome Measures

To estimate the prevalence and describe product use patterns, the product use behavior will be categorized based on individual product use or combined product use, changes in product use behavior, and status of product use for each tobacco- and nicotine-containing product.

IQOS use patterns are classified according to the current consumption as either exclusive (no other product than IQOS), dual (IQOS and 1 more product), and poly (IQOS and ≥2 products) use, irrespective of relative consumption levels.

For the changes in product use behavior, the following behaviors will be analyzed: initiation (start using a tobacco product for the first time), relapse (start using a tobacco product again after stopping for less than 12 months), reinitiation (start using a tobacco product after having stopped for more than 12 months), and product use transition (change of consumption from 1 tobacco product to another different tobacco product), in which the transition can be further categorized based on the numbers of products use after the transition, for example, complete transition (exclusive use of switched product) or partial transition (combined use of switched product with other products).

The status of each product use will be categorized as current use, former use, and nonuse based on WHO definitions or adapted definitions. For example, for IQOS use, a current IQOS user is defined as a person who uses HEETS tobacco sticks either every day or occasionally at the time of the survey and

has used 100 or more HEETS tobacco sticks in his/her lifetime. A former IQOS user is defined as a person who was formerly using IQOS either daily or occasionally but does not use IQOS at the time of the survey and has used 100 or more HEETS tobacco sticks in his/her lifetime. A non-IQOS user is a person who, at the time of the survey, does not use IQOS at all. For each status of individual product use, the use pattern will be further subcategorized according to WHO definitions. Current individual product use is subcategorized as daily or occasional use. Occasional use is subcategorized as reducer, experimenter, and continuing occasional user. Nonuse is subcategorized as ex-use, never use, and ex-occasional use. For example, a never IQOS user is a person who has never used IQOS at all or has never used IQOS daily and has used less than 100 HEETS tobacco sticks in his/her lifetime.

In addition, years between quitting cigarette smoking and initiation of IQOS, reasons for using IQOS together with subjects' level of interest to quit, and quit attempt will be analyzed. Although the results might not be able to provide direct information on the differences between IQOS users who would not have quit smoking cigarettes and those who would have had if IQOS had not been on the market, it will allow for assessing the impact of IQOS on cigarette smoking behavior.

Data Analysis

Response rates will be computed in accordance with the American Association for Public Opinion Research guidelines [47]. As the surveys are carried out with all households units that are contacted and are willing to participate, unit nonresponse can be distributed disproportionally and may cause biases in the sample. To account for these distortions, data will be standardized to the 2010 world population [48] and the population of each country, that is, 2011 German population, 2011 Italian population, and census data of 2011 UK population, Greater London population. Three standardizations will be performed: (1) by age, (2) by sex, and (3) by age-sex combinations. Standardized prevalence will be calculated as follows:

$$DSR_k = (\textstyle\sum_{i=1} N_i P_{ki}) / (\textstyle\sum_{i=1} N_i) \ [1]$$

Where DSR_k is the directly standardized prevalence of characteristic k (use of a certain tobacco product), p_{ki} is the observed prevalence of k in subgroup$_i$, and N_i is the size of the population in subgroup$_i$. The direct standardization weights the stratum-specific estimates by the size of the corresponding strata in the population.

Nonstandardized and standardized data will be analyzed and summarized descriptively. For continuous data, summary statistics will include the number and percent of subjects (n), the number of subjects with missing data, the arithmetic mean with 95% CIs, SD, median, first and third quartiles, minimum, and maximum; for log-normal data, the geometric mean and coefficient of variation will be presented. For categorical data, frequency counts, percentages, and 95% CIs will be presented. For the calculation of summary statistics and statistical analysis, unrounded data will be used.

Analyses will be stratified by the following factors: tobacco product used, sex, age, time since product initiation, intensity of use, and socioeconomic parameters (eg, income, education, and occupation). The trends in prevalence, initiation and cessation rates, as well as user behaviors will be presented based on annual data. The associations between patterns of tobacco product use and self-reported health status will be analyzed. In addition, for the population of IQOS users, the association between patterns of use (including misuse) with motivations to use, risk perceptions, perceived aesthetic changes, and consumer's satisfaction will be explored.

Results

In all 3 regions, the first-year data collection began in March 2018 for the general population and in April 2018 for IQOS users. We have planned to complete the first-year data collection by the end of 2018 for Germany and Italy and by the beginning of 2019 for Greater London.

Directly after the completion of the first-year data collection, the second-year data collection will start. In total, the annual data collection will be repeated for 3 consecutive years from 2018 to 2020.

The results of the first-year data analysis are expected to be available by June 2019.

Discussion

Strengths

The cross-sectional surveys in Germany, Italy, and Greater London, following the design of an on-going Japanese cross-sectional survey, are being conducted in representative samples of the general population and in nonrepresentative samples of registered IQOS users. Additional improvements have been implemented, including (1) sampling methods, (2) sample size calculation, and (3) survey questionnaire. For the general population sample, a random probability population-based sampling approach is employed instead of nonprobabilistic quota sampling as in the Japanese cross-sectional survey. Thus, the samples are more likely to be representative of the population. The sample size calculation of the surveys in 3 markets takes into account the results of the Japanese survey, that is, the IQOS use prevalence is about 1% IQOS in the general population and 63.4% IQOS users use IQOS exclusively in the IQOS sample. This sample size is sufficiently large to allow for the accurate estimates of tobacco use prevalence at the national level. Furthermore, we have implemented a new survey questionnaire. The SQ included in the surveys allows for a comprehensive characterization of tobacco use histories and behaviors according to WHO definitions [40]. By adapting the SQ, the exposure to other novel tobacco products available on the local markets such as IQOS, e-cigarettes, PLOOM, and glo can be assessed. However, measuring the use of novel tobacco products is challenging. As these products have not been on the market long enough, there is no established history of these novel products in surveillance surveys, and often, product-specific nomenclature and measures are not standardized. Most of the ongoing population-based smoking surveys either do not include these products yet and/or do not include exhaustive lists of tobacco products available to the public. In the these surveys, the participants have the opportunity to report all novel tobacco- and nicotine-containing products used, allowing a comprehensive assessment of the exposure to any type of tobacco- and nicotine-containing products. Although the accuracy of the SQ for cigarette smoking has been assessed, the adapted version for other novel tobacco products has not been validated. For the novel products that use tobacco sticks, such as IQOS and glo, the use behavior is quantified in a similar way as cigarette smoking. However, assessing exposure of vaporing products, either with electronic liquid such as e-cigarette or tobacco capsule such as PLOOM tech, is challenging. Vaporing behavior is different from cigarette smoking, as it is not bounded by the time it takes to burn a cigarette. Puffing number, frequency, and duration vary among the users. For e-cigarettes, different device designs and numerous varieties in nicotine levels and compositions are making the use pattern comparisons with other tobacco products even more difficult. Currently, there is no standard tool for assessing vaporing products exposure, although core items for assessing e-cigarette use in population-based surveys have been recommended by Pearson et al [49]. Our survey results on the use behavior of different products will provide valuable information regarding the feasibility of the adapted SQ.

In addition to measuring tobacco use patterns, these surveys will also collect information on motivation to use IQOS and intent to quit, which will allow for establishing the impact of IQOS on cigarette smoking behavior. The surveys will provide data on the attractiveness and acceptability of IQOS in current IQOS users as well as the potential effects of IQOS use on the intention of adult smokers to quit smoking. To assess the role of IQOS in harm reduction, evidence is needed to determine if IQOS has the potential to successfully compete with and replace smoking. Furthermore, these surveys will characterize the profile of dual IQOS users. According to a large, nationally representative, longitudinal study of tobacco use and health in the United States [50], more than a quarter of adults are current users of at least one type of tobacco product. There have been concerns that dual use, for example, dual use of cigarette and e-cigarette, may undermine cigarette smoking cessation, or worse, increase smoking and nicotine dependence [51,52]. However, more and more evidence indicates that e-cigarette use may promote smoking cessation and reduction among dual users [53-56], and dual use of tobacco and e-cigarettes does not necessarily perpetuate or exacerbate smokers' tobacco addiction and use [54]. Currently, we do not know if IQOS use is associated with quitting or reducing smoking in current smokers or dual IQOS users and contributing to relapse prevention in former smokers. If IQOS use has a similar effect as e-cigarettes on smoking cessation and relapse [26,53,57-63], the potential role of IQOS on smoking cessation intervention might be inferred.

In the surveys, the perceived health risk associated with IQOS use will be compared with that of smoking. Risk perceptions related to emerging products possibly have an important influence on how emerging nicotine products are used and by whom [64]. Currently, there are misperceptions of emerging

nicotine products, such as e-cigarettes, regarding the relative harmfulness. It has been shown that in the United Kingdom, only half of the adult smokers believe that e-cigarettes are less harmful than smoking and a majority of smokers and ex-smokers do not think that completely switching from cigarettes to e-cigarettes would lead to major health benefits [26]. The misperception is most likely linked to the perception that most adverse health effects are caused by nicotine [26]. It has been shown that half of the smokers even overestimate the harmfulness of NRTs [26]. Lack of knowledge about the products' health effects may contribute to the misperceptions [64]. Other studies have shown that misperceptions vary consistently by indicators of socioeconomic status, with more disadvantaged smokers and recent ex-smokers having higher rates of misperceptions [26]. The consequence of overestimating the harmfulness of NRTs and e-cigarettes is that smokers may be discouraged from using them in an effort to quit smoking [64]. Public Health England has pointed out that future research should aim at assessing the causes and effects of misperceptions of the relative harmfulness of e-cigarettes and NRTs compared with cigarettes [26]. Compared with e-cigarettes, which contain mainly nicotine liquids, the aerosol generated from THS contains not only nicotine but also other HPHCs, although significantly reduced, compared with cigarette smoke. Consequently, it is likely that the perceived harmfulness of THS will be overestimated, which could impact the acceptance and usage of the product. The survey results will provide evidence on risk perceptions of IQOS users under real-world conditions, although our premarketing assessment has demonstrated that the perceived risk associated with THS is lower than that associated with cigarettes, which is the most hazardous tobacco product, and higher than that associated with NRTs or cessation. The higher risk perception of THS compared with NRTs [29] is consistent with existing literature on risk perception of novel products compared with NRTs [65,66].

Although the weight of evidence that IQOS significantly reduces the exposure to HPHCs is compelling [7-15], higher levels of several components other than HPHCs are found in the aerosol generated by THS. Currently, there is no direct evidence available on the risk reduction of IQOS, and the health impact of the components other than HPHCs is unknown. Recently, Public Health England conducted a systematic review of available evidence and concluded that heated tobacco products may be considerably less harmful than tobacco cigarettes [26]. Our survey results on the perceived general health status [42], self-reported comorbidity [43], and, in particular, the self-reported changes in a number of relevant domains where IQOS users may have potential benefits since starting using the product will allow for exploring the potential health benefits of IQOS.

Limitations

These surveys have some limitations. The cross-sectional studies will not collect prospective data on product-use transition such as changes in the frequency and intensity. However, the SQ, which captures tobacco use history from past 3 months to more than 20 years ago, can provide useful retrospective information on tobacco use. Caution should be taken when evaluating the data, as recall bias and self-reporting bias could impact the data quality and limit the conclusiveness of the results. As IQOS was launched very recently in these 3 markets, a low IQOS use prevalence in the general population sample is expected. The sample size calculation is based on the assumption of 1% IQOS uptake. In case of even lower IQOS use prevalence, the studies are underpowered. However, reasonable inferences can still be made by incorporating information from CIs. The surveys are not designed to claim any causal effects of using tobacco- and nicotine-containing products on self-reported health. The assessment of self-reported, perceived aesthetic changes and current health status will rather allow to characterize different user populations accordingly. Nevertheless, the results of the surveys will provide useful information on the associations between patterns of tobacco product use and associated factors. Furthermore, the limitation of the IQOS user database is that it is most likely not representative of the IQOS user population in each country, as not all IQOS users who purchased the devices are registered in the database. There is also a chance of unintended over-representing of IQOS enthusiasts in the consumer database. Thus, the findings from the IQOS samples cannot be generalized to populations outside the samples, and across-country comparisons are limited. The comparison of IQOS user profiles sampled from the database with those sampled from the general population will provide valuable information regarding the extent of the nonrepresentativeness. However, this comparison is possible only when the IQOS use prevalence in each market has reached a sufficient level. In addition, the number of potential IQOS participants in the IQOS sample is highly dependent on the size of the PMI IQOS database. This might be critical for the Greater London survey, in particular for the first few waves, as the IQOS database in Greater London is currently relatively small compared with those in Italy and Germany. The risk of not being able to enroll a sufficient number of IQOS participants during the first few waves cannot be fully excluded.

Conclusions

The surveys aim to assess the prevalence of tobacco use and will provide insights into use patterns and associated factors. As the surveys will be conducted in 3 markets with similar design and at regular intervals, the results will allow for cross-regional and trend assessments. Most importantly, the results will provide relevant information allowing for assessing the potential health benefit of IQOS in the population.

Conflicts of Interest

All the authors work for PMI, R&D.

Multimedia Appendix 1

UK general population sample survey questionnaire.

[DOCX File, 726KB - resprot_v8i5e12061_app1.docx]

Multimedia Appendix 2

UK IQOS user sample survey questionnaire.

[DOCX File, 20KB - resprot_v8i5e12061_app2.docx]

References

1. World Health Organization. Geneva, Switzerland: WHO; 2015. WHO Global Report On Trends In Prevalence Of Tobacco Smoking 2000-2025 URL: https://www.who.int/tobacco/publications/surveillance/trends-tobacco-smoking-second-edition/en/ [accessed 2019-01-23] [WebCite Cache ID 75e9tz9Tx]

2. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health. US Department of Health and Human Services. 2014. The health consequences of smoking – 50 years of progress: a report of the Surgeon General URL: https://www.surgeongeneral.gov/library/reports/50-years-of-progress/full-report.pdf

3. Jamal A, Homa DM, O'Connor E, Babb SD, Caraballo RS, Singh T, et al. Current cigarette smoking among adults - United States, 2005-2014. MMWR Morb Mortal Wkly Rep 2015;64(44):1233-1260 [FREE Full text] [doi: 10.15585/mmwr.mm6444a2] [Medline: 26562061]

4. Abrams DB, Glasser AM, Pearson JL, Villanti AC, Collins LK, Niaura RS. Harm minimization and tobacco control: reframing societal views of nicotine use to rapidly save lives. Annu Rev Public Health 2018 Apr 1;39:193-213. [doi: 10.1146/annurev-publhealth-040617-013849] [Medline: 29323611]

5. Gottlieb S, Zeller M. A nicotine-focused framework for Public Health. N Engl J Med 2017 Sep 21;377(12):1111-1114. [doi: 10.1056/NEJMp1707409] [Medline: 28813211]

6. Le Houezec J, McNeill A, Britton J. Tobacco, nicotine and harm reduction. Drug Alcohol Rev 2011 Mar;30(2):119-123. [doi: 10.1111/j.1465-3362.2010.00264.x] [Medline: 21375611]

7. Jaccard G, Tafin Djoko D, Moennikes O, Jeannet C, Kondylis A, Belushkin M. Comparative assessment of HPHC yields in the Tobacco Heating System THS2.2 and commercial cigarettes. Regul Toxicol Pharmacol 2017 Nov;90:1-8 [FREE Full text] [doi: 10.1016/j.yrtph.2017.08.006] [Medline: 28818540]

8. Schaller JP, Keller D, Poget L, Pratte P, Kaelin E, McHugh D, et al. Evaluation of the Tobacco Heating System 2.2. Part 2: Chemical composition, genotoxicity, cytotoxicity, and physical properties of the aerosol. Regul Toxicol Pharmacol 2016 Nov 30;81(Suppl 2):S27-S47 [FREE Full text] [doi: 10.1016/j.yrtph.2016.10.001] [Medline: 27720919]

9. Wong ET, Kogel U, Veljkovic E, Martin F, Xiang Y, Boue S, et al. Evaluation of the Tobacco Heating System 2.2. Part 4: 90-day OECD 413 rat inhalation study with systems toxicology endpoints demonstrates reduced exposure effects compared with cigarette smoke. Regul Toxicol Pharmacol 2016 Nov 30;81(Suppl 2):S59-S81 [FREE Full text] [doi: 10.1016/j.yrtph.2016.10.015] [Medline: 27793746]

10. Sewer A, Kogel U, Talikka M, Wong ET, Martin F, Xiang Y, et al. Evaluation of the Tobacco Heating System 2.2 (THS2.2). Part 5: microRNA expression from a 90-day rat inhalation study indicates that exposure to THS2.2 aerosol causes reduced effects on lung tissue compared with cigarette smoke. Regul Toxicol Pharmacol 2016 Nov 30;81(Suppl 2):S82-S92 [FREE Full text] [doi: 10.1016/j.yrtph.2016.11.018] [Medline: 27866933]

11. Oviedo A, Lebrun S, Kogel U, Ho J, Tan WT, Titz B, et al. Evaluation of the Tobacco Heating System 2.2. Part 6: 90-day OECD 413 rat inhalation study with systems toxicology endpoints demonstrates reduced exposure effects of a mentholated version compared with mentholated and non-mentholated cigarette smoke. Regul Toxicol Pharmacol 2016 Nov 30;81(Suppl 2):S93-S122 [FREE Full text] [doi: 10.1016/j.yrtph.2016.11.004] [Medline: 27818348]

12. Kogel U, Titz B, Schlage WK, Nury C, Martin F, Oviedo A, et al. Evaluation of the Tobacco Heating System 2.2. Part 7: Systems toxicological assessment of a mentholated version revealed reduced cellular and molecular exposure effects compared with mentholated and non-mentholated cigarette smoke. Regul Toxicol Pharmacol 2016 Nov 30;81(Suppl 2):S123-SS38 [FREE Full text] [doi: 10.1016/j.yrtph.2016.11.001] [Medline: 27818347]

13. Martin F, Talikka M, Ivanov NV, Haziza C, Hoeng J, Peitsch MC. Evaluation of the tobacco heating system 2.2. Part 9: Application of systems pharmacology to identify exposure response markers in peripheral blood of smokers switching to THS2.2. Regul Toxicol Pharmacol 2016 Nov 30;81(Suppl 2):S151-S157 [FREE Full text] [doi: 10.1016/j.yrtph.2016.11.011] [Medline: 27845159]

14. Haziza C, de La Bourdonnaye G, Merlet S, Benzimra M, Ancerewicz J, Donelli A, et al. Assessment of the reduction in levels of exposure to harmful and potentially harmful constituents in Japanese subjects using a novel tobacco heating system compared with conventional cigarettes and smoking abstinence: A randomized controlled study in confinement. Regul Toxicol Pharmacol 2016 Nov;81:489-499 [FREE Full text] [doi: 10.1016/j.yrtph.2016.09.014] [Medline: 27693654]

XSL•FO
RenderX

15. Haziza C, de La Bourdonnaye G, Skiada D, Ancerewicz J, Baker G, Picavet P, et al. Evaluation of the Tobacco Heating System 2.2. Part 8: 5-Day randomized reduced exposure clinical study in Poland. Regul Toxicol Pharmacol 2016 Nov 30;81(Suppl 2):S139-S150 [FREE Full text] [doi: 10.1016/j.yrtph.2016.11.003] [Medline: 27816672]

16. Li X, Luo Y, Jiang X, Zhang H, Zhu F, Hu S, et al. Chemical analysis and simulated pyrolysis of Tobacco Heating System 2.2 compared to conventional cigarettes. Nicotine Tob Res 2018 Jan 8;21(1):111-118. [doi: 10.1093/ntr/nty005] [Medline: 29319815]

17. Prodanchuk MG, Podrushnyak AE, Malysheva OE, Stroy AM, Zaval'na VV, Moroz TI, et al. Potential risk assessment of the electrically heated tobacco system (EHTS) use. Food Chem Safety 2017(1-2):77-78 [FREE Full text]

18. Gale N, McEwan M, Eldridge AC, Fearon IM, Sherwood N, Bowen E, et al. Changes in biomarkers of exposure on switching from a conventional cigarette to tobacco heating products: a randomized, controlled study in healthy Japanese subjects. Nicotine Tob Res 2018 Jun 15 (forthcoming). [doi: 10.1093/ntr/nty104] [Medline: 29912406]

19. Ruprecht AA, De Marco C, Saffari A, Pozzi P, Mazza R, Veronese C, et al. Environmental pollution and emission factors of electronic cigarettes, heat-not-burn tobacco products, and conventional cigarettes. Aerosol Sci Technol 2017 Feb 28;51(6):674-684 [FREE Full text] [doi: 10.1080/02786826.2017.1300231]

20. Zuck K. Evidence related to the health risk of IQOS use: evaluation of product chemistry. 2018 Presented at: Part of 2018 Tobacco Products Scientific Advisory Committe (TPSAC) Meeting Materials and Information; January 24-25, 2018; USA URL: https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/TobaccoProductsScientificAdvisoryCommittee/UCM594327.pdf

21. UK Committee on Carcinogenicity of Chemicals in Food, CPatEC, UK Committee on Toxicity of Chemicals in Food CPatEC, and UK Committee on Mutagenicity of Chemicals in Food CPatEC. 2017 Dec 12. Statement on the toxicological evaluation of novel heat-not-burn tobacco products URL: http://www.webcitation.org/74Ns80oDK [accessed 2018-12-03] [WebCite Cache ID 74Ns80oDK]

22. Pieper E, Mallock N, Henkler-Stephani F, Luch A. Tabakerhitzer als neues Produkt der Tabakindustrie: Gesundheitliche Risiken. ["Heat not burn" tobacco devices as new tobacco industry products: health risks]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2018 Nov;61(11):1422-1428. [doi: 10.1007/s00103-018-2823-y] [Medline: 30284624]

23. Bekki K, Inaba Y, Uchiyama S, Kunugita N. Comparison of chemicals in mainstream smoke in teat-not-burn tobacco and combustion cigarettes. J UOEH 2017;39(3):201-207 [FREE Full text] [doi: 10.7888/juoeh.39.201] [Medline: 28904270]

24. Rumsey B. Report iQOS investigation Ministry of Health New Zealand. 2017 Nov 17. Investigation into iQOS device, HEETS tobacco sticks and evidence of combustion URL: https://www.pmiscience.com/resources/docs/default-source/NCDC-vs-Morris/nz_crl-energy-ltd---investigation-into-iqos-device-heets-tobacco-sticks-and-evidence-of-combustion_november-2017.pdf [accessed 2018-12-03] [WebCite Cache ID 74O75UkrT]

25. Dutch National Institute for Public Health and the Environment (RIVM). 2018 May 14. IQOS Factsheet URL: https://www.rivm.nl/en/news/addictive-nicotine-and-harmful-substances-also-present-in-heated-tobacco [accessed 2018-12-03] [WebCite Cache ID 74O7AjRxe]

26. McNeill AD, Brose LS, Calder R, Bauld L, Robson D. Public Health England. 2018 Feb 12. Evidence review of e-cigarettes and heated tobacco products 2018 URL: https://kclpure.kcl.ac.uk/portal/en/publications/evidence-review-of-ecigarettes-and-heated-tobacco-products-2018(02a423a9-8fb7-4ee1-aec6-a37165524322).html [WebCite Cache ID 71fHQ2X4Q]

27. St Helen G, Jacob Iii P, Nardone N, Benowitz NL. IQOS: examination of Philip Morris International's claim of reduced exposure. Tob Control 2018 Nov;27(Suppl 1):s30-s36 [FREE Full text] [doi: 10.1136/tobaccocontrol-2018-054321] [Medline: 30158205]

28. Auer R, Concha-Lozano N, Jacot-Sadowski I, Cornuz J, Berthet A. eat-not-burn tobacco cigarettes: smoke by any other name. JAMA Intern Med 2017 Jul 1;177(7):1050-1052 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1419] [Medline: 28531246]

29. Beacher F, Magnani P, Ramazotti A, Weitkunat R, Colman S, Kallischnigg G, et al. Intention to use and perceptions of health and addiction risk for the Tobacco Heating System 2.2. 2016 Presented at: Poster presented at the Behavioural Insights in Research and Policy Making - SABE/IAREP Conference; July 8-10, 2016; Wageningen, the Netherlands URL: https://www.pmiscience.com/resources/docs/default-source/library-documents/beacher_sabeiarep_2016_intention_to_use_and_perception_of_health_risk_and_addiction_risk.pdf?sfvrsn=c6bdf706_2

30. Magnani P, Beacher F, Wilson H, Badoglio S, Ramazzotti A. Philip Morris International Management. 2017. Intent to Use for a new menthol HeatSticks variant in Japan URL: https://www.pmiscience.com/resources/docs/default-source/library-documents/intent-to-use-for-a-new-menthol-heatsticks-variant-in-japan.pdf?sfvrsn=d654c906_0 [accessed 2018-08-14] [WebCite Cache ID 71f9nECub]

31. Stratton K, Shetty P, Wallace R, Bondurant S. Clearing the smoke: the science base for tobacco harm reduction--executive summary. Tob Control 2001 Jun;10(2):189-195 [FREE Full text] [Medline: 11387543]

32. Hatsukami DK, Biener L, Leischow SJ, Zeller MR. Tobacco and nicotine product testing. Nicotine Tob Res 2012 Jan;14(1):7-17 [FREE Full text] [doi: 10.1093/ntr/ntr027] [Medline: 21460383]

XSL•FO
RenderX

33.   Van der Plas A, Prieto L, Skiada D, Dobrynina M, Baker G, Lüdicke F. Prevalence and patterns of tobacco use in Japan after the commercialization of a heat-not-burn alternative (IQOS) to cigarettes. 2017 Aug 20 Presented at: IEA World Congress of Epidemiology; August 20, 2017; Saitama, Japan p. 31 URL: https://www.pmiscience.com/resources/docs/default-source/Presentations_Latest/wce-2017-van-der-plas-prevalence-and-patterns-of-tobacco-use-in-japan-after-the-commercialization-of-a-heat-not-burn.pdf?sfvrsn=decfcd06_0

34.   Langer P, Prieto L, Skiada D, Dobrynina M, Van der Plas A, Afolalu E, et al. Prevalence and patterns of tobacco product use in Japan: first-year results of a repeated cross-sectional survey. 2018 Presented at: Poster presented at the 4th International Conference on Public Health; July 19-21, 2018; Bangkok, Thailand URL: https://publichealthconference.co/wp-content/uploads/2018/08/ICOPH-2018-Book-of-Abstracts.pdf

35.   World Medical Association. 2018 Jul 9. Declaration of Helsinki - Ethical principles for medical research involving human subjects URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2018-08-14] [WebCite Cache ID 71fF6AU02]

36.   IEA (International Epidemiological Association). 2007. Good Epidemiological Practice (GEP): IEA guidelines for proper conduct of epidemiological research URL: https://www.yumpu.com/en/document/view/19409945/iea-guidelines-for-proper-conduct-of-epidemiological-research-hrsa [accessed 2018-08-14] [WebCite Cache ID 71fFE15Zg]

37.   CIOMS (Council for the International Organizations of Medical Sciences) in collaboration with the World Health Organization (WHO). 2008 Feb. International ethical guidelines for epidemiological studies (provisional text) URL: https://www.ufrgs.br/bioetica/cioms2008.pdf [accessed 2018-08-14] [WebCite Cache ID 71fFOuPPz]

38.   Weitkunat R, Coggins CR, Sponsiello-Wang Z, Kallischnigg G, Dempsey R. Assessment of cigarette smoking in epidemiologic studies. Beitr Tabakforsch 2013 Dec 30;25(7):638-648. [doi: 10.2478/cttr-2013-0940]

39.   Sponsiello-Wang Z, de La Bourdonnaye G, David M, Lüdicke F, Weitkunat R. Accuracy of the smoking questionnaire. Beitr Tabakforsch 2017 Dec 29;27(8):224-239. [doi: 10.1515/cttr-2017-0023]

40.   WHO (World Health Organization). WHO Guidelines for Controlling and Monitoring the Tobacco Epidemic. Geneva, Switzerland: WHO; 1998. URL: https://apps.who.int/iris/handle/10665/42049

41.   Kotz D. Heinrich-Heine-University Düsseldorf, Institute of General Practice, Centre for Addiction Research and Clinical Epidemiology. 2017 Jun 30. Recent trends in the use of tobacco and e-cigarettes in Germany: findings from the new DEBRA survey URL: http://www.webcitation.org/71fGK6jW4 [WebCite Cache ID 71fGK6jW4]

42.   Prokhorov AV, Warneke C, de Moor C, Emmons KM, Mullin Jones M, Rosenblum C, et al. Self-reported health status, health vulnerability, and smoking behavior in college students: implications for intervention. Nicotine Tob Res 2003 Aug;5(4):545-552. [Medline: 12959792]

43.   Sangha O, Stucki G, Liang MH, Fossel AH, Katz JN. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. Arthritis Rheum 2003 Apr 15;49(2):156-163 [FREE Full text] [doi: 10.1002/art.10993] [Medline: 12687505]

44.   Cappelleri JC, Bushmakin AG, Baker CL, Merikle E, Olufade AO, Gilbert DG. Confirmatory factor analyses and reliability of the modified cigarette evaluation questionnaire. Addict Behav 2007 May;32(5):912-923. [doi: 10.1016/j.addbeh.2006.06.028] [Medline: 16875787]

45.   Hatsukami DK, Zhang Y, O'Connor RJ, Severson HH. Subjective responses to oral tobacco products: scale validation. Nicotine Tob Res 2013 Jul;15(7):1259-1264 [FREE Full text] [doi: 10.1093/ntr/nts265] [Medline: 23239843]

46.   Chrea C, Emilien G, Salzberger T, Cano S, Alfieri T, Mainy N, et al. Final psychometric evaluation and cultural adaptation of the Perceived Risk Instrument (PRI) to measure perceived risks associated with the use of tobacco and nicotine-containing products. In: Poster presented at the Society for Research on Nicotine and Tobacco Annual Meeting, Chicago, USA. 2016 Presented at: Society for Research on Nicotine and Tobacco Annual Meeting; March 3, 2016; Chicago, USA URL: http://www.webcitation.org/71fGZnlxP

47.   AAPOR (American Association for Public Opinion Research). 2016. Standard definitions: final dispositions of case codes and outcome rates for surveys, Ninth Edition URL: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf [accessed 2018-08-14] [WebCite Cache ID 71fH9KgLV]

48.   Population Reference Bureau. 2010. 2010 World Population Data Sheet URL: https://www.prb.org/wp-content/uploads/2010/11/10wpds_eng.pdf [accessed 2018-08-14] [WebCite Cache ID 71fGt7dGj]

49.   Pearson JL, Hitchman SC, Brose LS, Bauld L, Glasser AM, Villanti AC, et al. Recommended core items to assess e-cigarette use in population-based surveys. Tob Control 2018 May;27(3):341-346. [doi: 10.1136/tobaccocontrol-2016-053541] [Medline: 28624764]

50.   Kasza KA, Ambrose BK, Conway KP, Borek N, Taylor K, Goniewicz ML, et al. Tobacco-product use by adults and youths in the United States in 2013 and 2014. N Engl J Med 2017 Dec 26;376(4):342-353 [FREE Full text] [doi: 10.1056/NEJMsa1607538] [Medline: 28121512]

51.   Grana R, Benowitz N, Glantz SA. E-cigarettes: a scientific review. Circulation 2014 May 13;129(19):1972-1986 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.007667] [Medline: 24821826]

XSL•FO

RenderX

52.   Pulvers K, Hayes RB, Scheuermann TS, Romero DR, Emami AS, Resnicow K, et al. Tobacco use, quitting behavior, and health characteristics among current Electronic Cigarette users in a national tri-ethnic adult stable smoker sample. Nicotine Tob Res 2015 Sep;17(9):1085-1095 [FREE Full text] [doi: 10.1093/ntr/ntu241] [Medline: 25385875]

53.   Maglia M, Caponnetto P, Di Piazza J, La Torre D, Polosa R. Dual use of electronic cigarettes and classic cigarettes: a systematic review. Addict Res Theory 2017 Dec 6;26(4):330-338. [doi: 10.1080/16066359.2017.1388372]

54.   Rahman MA, Hann N, Wilson A, Mnatzaganian G, Worrall-Carter L. E-cigarettes and smoking cessation: evidence from a systematic review and meta-analysis. PLoS One 2015 Mar 30;10(3):e0122544 [FREE Full text] [doi: 10.1371/journal.pone.0122544] [Medline: 25822251]

55.   McRobbie H, Bullen C, Hartmann-Boyce J, Hajek P. Electronic cigarettes for smoking cessation and reduction. Cochrane Database Syst Rev 2014(12):CD010216. [doi: 10.1002/14651858.CD010216.pub2] [Medline: 25515689]

56.   Etter JF, Bullen C. A longitudinal study of electronic cigarette users. Addict Behav 2014 Feb;39(2):491-494. [doi: 10.1016/j.addbeh.2013.10.028] [Medline: 24229843]

57.   Adkison SE, O'Connor RJ, Bansal-Travers M, Hyland A, Borland R, Yong HH, et al. Electronic nicotine delivery systems: international tobacco control four-country survey. Am J Prev Med 2013 Mar;44(3):207-215 [FREE Full text] [doi: 10.1016/j.amepre.2012.10.018] [Medline: 23415116]

58.   Biener L, Hargraves JL. A longitudinal study of electronic cigarette use among a population-based sample of adult smokers: association with smoking cessation and motivation to quit. Nicotine Tob Res 2015 Feb;17(2):127-133 [FREE Full text] [doi: 10.1093/ntr/ntu200] [Medline: 25301815]

59.   Siegel MB, Tanwar KL, Wood KS. Electronic cigarettes as a smoking-cessation: tool results from an online survey. Am J Prev Med 2011 Apr;40(4):472-475. [doi: 10.1016/j.amepre.2010.12.006] [Medline: 21406283]

60.   Berry KM, Reynolds LM, Collins JM, Siegel MB, Fetterman JL, Hamburg NM, et al. E-cigarette initiation and associated changes in smoking cessation and reduction: the Population Assessment of Tobacco and Health Study, 2013-2015. Tob Control 2018 Mar;28(1):42-49. [doi: 10.1136/tobaccocontrol-2017-054108]

61.   Hughes, JR, Hecht S, Carmella S, Murphy S, Callas P. Smoking behaviour and toxin exposure during six weeks use of a potential reduced exposure product: Omni. Tob Control 2004;13(2):175-179 [FREE Full text]

62.   Stead LF, Perera R, Bullen C, Mant D, Hartmann-Boyce J, Cahill K, et al. Nicotine replacement therapy for smoking cessation. Cochrane Database Syst Rev 2012 Nov 14;11:CD000146. [doi: 10.1002/14651858.CD000146.pub4] [Medline: 23152200]

63.   Hajek P, Stead LF, West R, Jarvis M, Hartmann-Boyce J, Lancaster T. Relapse prevention interventions for smoking cessation. Cochrane Database Syst Rev 2013 Aug 20(8):CD003999. [doi: 10.1002/14651858.CD003999.pub4] [Medline: 23963584]

64.   Czoli CD, Fong GT, Mays D, Hammond D. How do consumers perceive differences in risk across nicotine products? A review of relative risk perceptions across smokeless tobacco, e-cigarettes, nicotine replacement therapy and combustible cigarettes. Tob Control 2017 Mar;26(e1):e49-e58. [doi: 10.1136/tobaccocontrol-2016-053060] [Medline: 27625408]

65.   Øverland S, Skogen JC, Lissner L, Bjerkeset O, Tjora T, Stewart R. Snus use and cardiovascular risk factors in the general population: the HUNT3 study. Addiction 2013 Nov;108(11):2019-2028. [doi: 10.1111/add.12313] [Medline: 23909909]

66.   Pepper JK, Emery SL, Ribisl KM, Rini CM, Brewer NT. How risky is it to use e-cigarettes? Smokers' beliefs about their health risks from using novel and traditional tobacco products. J Behav Med 2015 Apr;38(2):318-326 [FREE Full text] [doi: 10.1007/s10865-014-9605-2] [Medline: 25348584]

Abbreviations

CAPI: computer-assisted personal interviews
e-cigarette: electronic cigarette
FDA: Food and Drug Administration
HPHCs: harmful and potentially harmful constituents
NRT: nicotine-replacement therapy
OA: output area
PMI: Philip Morris International
PRI-G: Perceived Risk Instrument-general version
SQ: Smoking Questionnaire
THS: Tobacco Heating System
ToNiPEQ: Tobacco/Nicotine-containing Product Evaluation Questionnaire
WHO: World Health Organization

XSL·FO

RenderX

Protocol

The Impact of Treatment Adherence for Patients With Diabetes and Hypertension on Cardiovascular Disease Risk: Protocol for a Retrospective Cohort Study, 2008-2018

Min Su[1,2], PhD; Victoria Haldane[2], MPH; Ross Upshur[2,3,4], MD; Frank Sullivan[2,4,5], MD; France Légaré[6], PhD; Michelle Greiver[4,7], MSc; Xiaolin Wei[2], PhD

[1]School of Public Administration, Inner Mongolia University, Hohhot, China

[2]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

[3]Lunenfeld Tanenbaum Research Institute, Sinai Health Systems, Toronto, ON, Canada

[4]Department of Family and Community Medicine, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[5]School of Medicine, University of St Andrews, North Haugh, United Kingdom

[6]Faculty of Medicine, Laval University, Quebec, QC, Canada

[7]North York General Hospital, Toronto, ON, Canada

Corresponding Author:
Xiaolin Wei, PhD
Dalla Lana School of Public Health
University of Toronto
155 College Street
Toronto, ON, M5T 3M7
Canada
Phone: 1 416 978 2020
Fax: 1 416 978 1883
Email: xiaolin.wei@utoronto.ca

Abstract

Background: Cardiovascular disease (CVD) is the leading cause of death globally and in Canada. Diabetes and hypertension are major risk factors for CVD events. Despite the increasing availability of effective treatments, the majority of diabetic and hypertensive patients do not have adequate blood pressure and glycemic control. One of the major contributors is poor treatment adherence.

Objective: This study aims to evaluate the impact of treatment adherence for patients with both diabetes and hypertension on acute severe CVD events and intermediate clinical outcomes in Canadian primary care settings.

Methods: We will conduct a population-based retrospective cohort study of patients living with both diabetes and hypertension in Ontario, Canada, between January 1, 2008, and March 31, 2018. The Social Cognitive Theory will be used as a conceptual framework by which to frame the reciprocal relationship between treatment adherence, personal factors, and environmental determinants and how this interplay impacts CVD events and clinical outcomes. Data will be derived from the Diabetes Action Canada National Data Repository. A time-varying Cox proportional hazards model will be used to estimate the impacts of treatment adherence on CVD morbidity and mortality. Multivariable linear regression models and hierarchical regression models will be used to estimate the associations between treatment adherence of different medication categories and intermediate clinical outcomes. Our primary outcome is the association between treatment adherence and the risk of acute severe CVD events, including CVD mortality. The secondary outcome is the association between treatment adherence and intermediate clinical outcomes including diastolic and systolic blood pressures, glycated hemoglobin, low-density lipoprotein cholesterol, and total cholesterol. Owing to data limitation, we use medication prescriptions as a proxy to estimate treatment adherence. We assume that a patient adhered to medications if she or he had any prescription record in the 4 preceding quarters and 1 quarter after each quarter of interest. Acute severe CVD events are defined based on the World Health Organization's Monitoring Trends and Determinants in Cardiovascular Disease Project, including acute coronary heart disease, stroke, and heart failure. As causes of death are not available, the number of CVD deaths will be computed using the most recent systolic blood pressure distributions and the population attributable risks related to systolic blood pressure level.

Conclusions: The findings will be helpful in identifying the challenges of treatment adherence for diabetic and hypertensive patients in primary care settings. This will also help to develop intervention strategies to promote treatment adherence for patients with multi-morbidities.

Introduction

Background

Cardiovascular disease (CVD) is the leading cause of death in Canada, accounting for one-third of deaths nationally [1]. Diabetes and hypertension are major risk factors for CVDs [2-4]. In 2017, approximately 7.3% and 17.8% of Canadians aged 12 years and older reported being diagnosed with diabetes and hypertension, respectively [5]. Diabetes and hypertension are the 2 most common comorbid chronic diseases seen in primary care consultations. Hypertension is reported in over two-thirds of patients with type 2 diabetes [6], whereas nearly 50% of patients with hypertension are diabetic [7]. Despite the increasing availability of effective treatment regiments and guidelines, approximately half of the treated patients do not have adequate blood pressure and glycemic control [8,9]. One of the major contributors to inadequate control is poor treatment adherence [9,10]. Treatment adherence is defined as the degree to which the patient's behavior corresponds with the agreed recommendations from a health care provider [11]. Adherence to antihypertension and antidiabetes medications is proven to reduce CVD morbidity and mortality, hospitalizations, and health expenditure [11-17]. Given the potential impacts of treatment adherence on CVD morbidity and mortality, quantifying treatment adherence and its impacts will help in the development of intervention strategies to improve treatment adherence for patients in primary care settings. This includes the use of patient-centered approaches such as concordance, where doctors elicit patients' views, inform patients of the pros and cons of taking medicine, and involve patients in treatment decision making [18]. This type of informed and shared decision making is believed to improve patient satisfaction, adherence, and treatment outcomes [19]. Unfortunately, the process of patient-doctor communication is often not recorded and cannot be quantified using traditional medical records to understand adherence behaviors.

Treatment adherence can be measured using either subjective or objective methods. One of the most frequently used subjective measures is the Morisky Medication Adherence Scale [20], which is a patient self-reported tool with 8 items related to medication-taking behaviors that can be transformed into an adherence score [21,22]. Morisky scale heavily relies on patient's attitudes toward their medications rather than actual medication-taking behaviors and is vulnerable to significant recall bias [21,22]. Patients tend to underreport their nonadherence to avoid disapproval from their physicians or researchers administering the test [17]. A major limitation of conducting objective measures of adherence is that administrative databases often do not record medication taking. Proxy measures such as prescription refills are used [17] with the assumption that prescription refill patterns correspond to medication-taking behavior [23-25]. However, medication refill patterns can vary substantially across different health care providers and settings.

Up to now, there were limited studies investigating medication treatment adherence and its association with CVD events [26-35] and clinical outcomes [35-40] (Multimedia Appendix 1). Using refill adherence, previous studies have reported that lower adherence (<80%) levels were associated with higher risk of CVD [26], all-cause mortality, and hospitalization for CVD after adjusting for demographic, socioeconomic status (SES), and baseline clinical characteristics [29]; higher adherence to statins (>80%) was associated with significant reduction in low-density lipoprotein cholesterol (LDL-C) in patients with diabetes [36]. In these studies, refill adherence was measured by the medication possession ratio (MPR), reporting the proportion of days with medications on hand during the follow-up. In Canada, there have been 12 studies that were investigated for treatment adherence of patients with diabetes [41-44], hypertension [30-32,34,45,46], or both [47] (Multimedia Appendix 1). Of these studies, 5 reported a negative association between treatment adherence and chronic heart failure [30,31], end-stage renal disease [31], mortality [32], a composite of all-cause death and hospitalization for acute myocardial infarction, heart failure, or stroke [33], and combined CVD events (coronary artery disease, cerebrovascular disease, and chronic heart failure) and hospitalization costs [34]. Using refill adherence, previous studies have showed that lower treatment adherence (<80%) was associated with higher risk of coronary disease [36], cerebrovascular disease [36], and chronic heart failure [36], after adjustment for demographic and SES [30]. These studies had several common limitations. First, few studies examined treatment adherence among patients with multimorbidities, such as both diabetes and hypertension, which accounted for the majority of these patients and is in line with studies showing that multimorbidity and medical complexity increase with age [48]. Second, most studies examined only one type of medication adherence, not considering the combined benefits of adhering to multiple medications in preventing CVD events and mortality [49,50]. Third, many studies were

cross-sectional in design and relied on survey- or hospital-based electronic medical records (EMRs) or had a relatively shorter follow-up time (<10 years), which limited the studies' ability to inform clinical practice at the primary care level.

Theoretical Framework

Treatment adherence is largely viewed, and measured, as a behavior at the individual level [51]. However, adherence is multifactorial and influenced by a host of environmental determinants [52-54]. Although often neglected in adherence studies, environmental determinants have been reported as barriers to adherence at the individual level [55]. Thus, we intend to use Social Cognitive Theory as a lens through which to conceptualize both personal-related factors and environmental determinants that may influence treatment adherence. Social Cognitive Theory explains a reciprocal relationship between that behavior, personal factors, and environmental determinants, and in the case of health, this interplay impacts health outcomes (Figure 1) [56,57].

Other studies of adherence at the individual level have employed the Health Belief Model or Theory of Planned Behavior to understand how individuals may engage in adherence behavior; however, these frameworks are limited in application at the population level and do not adequately explicate environmental determinants of adherence behaviors [58-60]. Similarly, theoretical frameworks that include environmental determinants at the population level, such as the Social-Ecological Model, while useful in revealing the environmental, organizational, and social factors influencing health, are limited in ability to highlight the behavioral mechanisms underpinning adherence [61]. Thus, the Social Cognitive Theory, which explicitly includes environmental determinants and personal factors as contributing to behavior, is well placed to investigate the population factors that impact adherence.

Our study seeks to examine the ways in which medication prescription, as proxy for adherence, impacts on CVD events and clinical outcomes. As shown in our conceptual framework (Figure 1), the behavior of treatment adherence operates not only at the individual level but also cumulatively at the population level. Similarly, health system touchpoints exist and exert impact at the population level as part of the environmental determinants of treatment adherence. These factors also interact with and are influenced by personal factors; however, in our analysis, we are controlling for these personal-level elements to highlight the role of medication prescription as an environmental determinant. We hypothesize that these environmental factors, measured by primary care provider characteristics, medication prescription, and medication regimen complexity over time, in turn, would impact on individual CVD events and clinical outcomes. Ultimately, this lens allows us to consider how population-level considerations, such as health system touchpoints, impact adherence at the population level. The results will provide evidence that may inform health policy and specific health service interventions.

Objectives and Hypotheses

Objective 1: The first objective was to assess the impacts of treatment adherence on acute severe CVD events in Ontario between January 1, 2008, and March 31, 2018 (10 years).

Hypothesis 1: Patients with a lower adherence rate of one or more medications (antihypertension, antidiabetes, statins, and aspirin) are more likely to develop acute severe CVD events (including death from CVD), adjusting for potential confounding factors.

Objective 2: The second objective was to assess the impacts of treatment adherence on intermediate clinical outcomes in Ontario between January 1, 2008, and March 31, 2018.

Hypothesis 2: Patients with a higher adherence rate are associated with significant improvements in clinical outcomes including diastolic blood pressure, systolic blood pressure, glycated hemoglobin (HbA$_{1c}$), LDL-C, and total cholesterol (TC), adjusting for potential confounding factors.

Figure 1. Theoretical framework. CVD: cardiovascular disease.
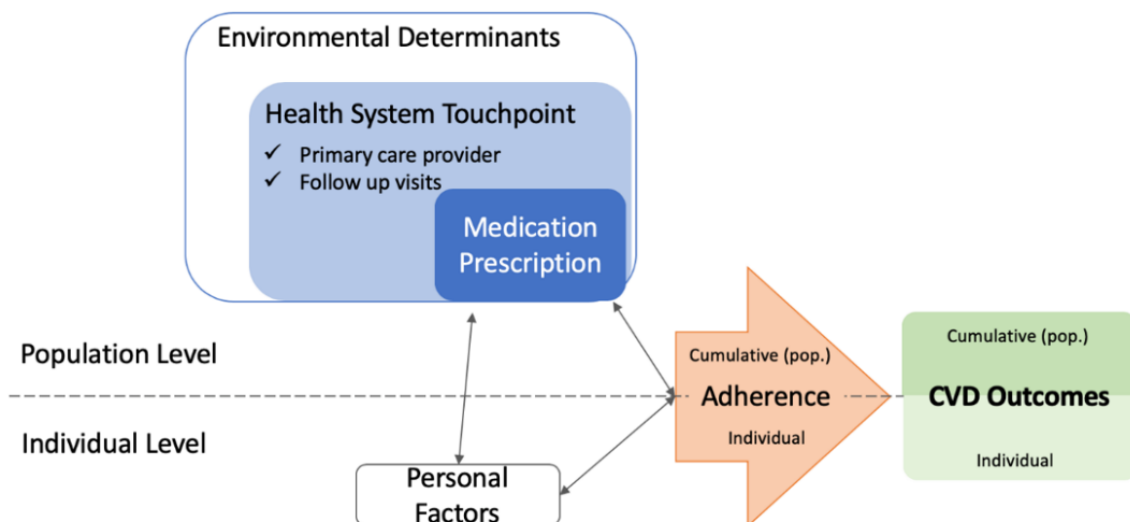
Methods

Study Design and Participants

This is a retrospective cohort study using primary care EMR data. A cohort of patients who were medically diagnosed with both diabetes and hypertension between January 1, 2008, and March 31, 2018, will be included in this study.

Inclusion and Exclusion Criteria for Participants

Time to Enter the Cohort

Cases entered the cohort when a medical diagnosis of both diabetes and hypertension was present and when a prescription for any antihypertensives or antidiabetic medication was provided in the EMR. The exclusion criteria included (1) patients with a past history of any acute severe CVD event, (2) patients who developed CVD events during follow-up where no date of CVD event was present, and (3) patients whose follow-up period was below 6 quarters as we cannot estimate its treatment adherence. A total of 15,642 eligible participants are identified in the finial study population (Figures 2 and 3).

Follow-Up

We will retrospectively follow-up all eligible participants until March 31, 2018 (by months). Follow-up ends when a participant dies, has any acute severe CVD event, or by the end of the study (March 31, 2018). For hypotheses related with clinical outcome, follow-up ends with the participant's latest diastolic blood pressure, systolic blood pressure, HbA_{1c}, LDL-C, and TC outcomes. The follow-up period will be measured approximately in months and treated as an independent variable.

Data Source

We will utilize the Diabetes Action Canadian National Diabetes Repository as the data source. The Repository contains deidentified data from over 100,000 patients living with diabetes, currently from 4 Canadian provinces (Ontario, Manitoba, Quebec, and Alberta). Data are extracted from primary care EMRs of consenting family physicians and nurse practitioners by regional Practice Based Research Networks who are members of the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) and are managed using previously described processes developed through CPCSSN [61]. The Repository provides a Secure Analytic Virtual Environment, which is a privacy compliant research platform in a high-performance computing center. Approved researchers access the Secure Analytic Virtual Environment remotely to analyze datasets derived from Repository data. All projects are reviewed by the Repository's Research Governing Committee, composed of at least 50% patients, to ensure the project's values are consistent with those of patients living with diabetes and of their caregivers. This project was reviewed and approved by the Research Governing Committee.

Deidentified patient data from contributing practices in the Diabetes Repository include the following: (1) patient demographic characteristics, (2) patient health conditions, (3) physical and laboratory examinations, (4) medication prescriptions, (5) risk factors, and (6) comorbidities. A data dictionary that provides information on data elements is available at the Diabetes Action Canada website.

Figure 2. Cohort participant selection for the association between treatment adherence and cardiovascular disease (CVD) morbidity and mortality.
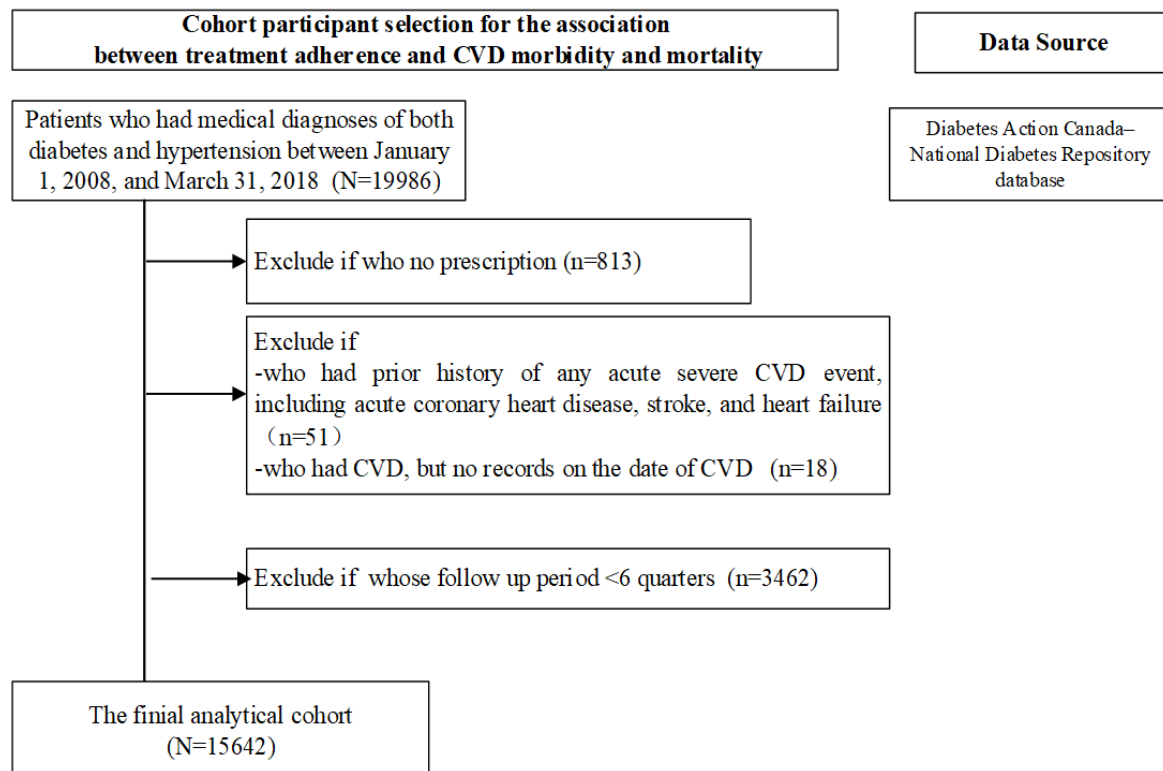
Figure 3. Cohort participant selection for the association between treatment adherence and clinical outcomes. CVD: cardiovascular disease; SBP:systolic blood pressure; DBP: diastolic blood pressure; HbA$_{1c}$: glycated hemoglobin; LDL-C: low-density lipoprotein cholesterol.
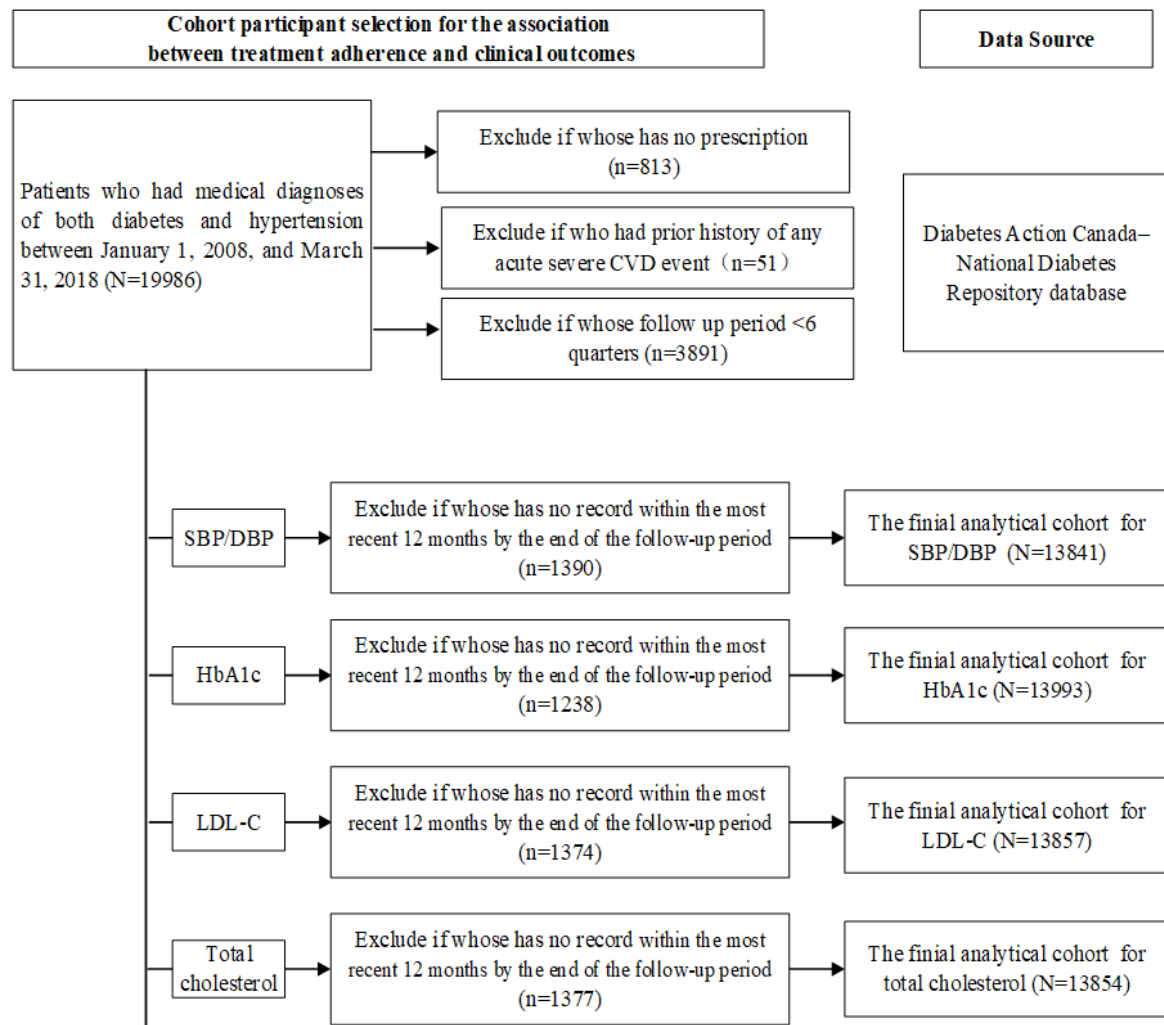
Exposure and Study Outcomes

Treatment Adherence Rate

Medication adherence rate is considered the exposure of interest. Medications will be classified as (1) antidiabetic medications, including metformin, sulfonylurea, and insulin, inhibitors of dipeptidyl peptidase 4, meglitinide, sodium-glucose cotransporter-2 inhibitors, thiazolidinedione, and alpha-glucosidase enzymes, (2) antihypertensive medications, including angiotensin converting enzyme inhibitors, thiazide diuretics, beta-blockers, calcium channel blockers, and angiotensin II receptor blockers, (3) statins, and (4) aspirin. We will measure patient adherence to each type of prescription if prescribed. Theoretically, the adherence rate could be calculated based on the prescription date and refills. However, we cannot depend on the refills as this information varies highly among primary care physicians (eg, some refilled every 3 months, others provided multiple repeats) [62]. Previous research indicated that lag-lead approach is feasible to estimate adherence based on time-dependent associations between different variables (or the same variable) in longitudinal data analysis [63]. We will use this approach to account for the variation in refills. Based on a study using the CPCSSN database, we assume that patients adhered to medications if they had any prescription

record in the 4 preceding quarters and 1 quarter after each quarter of interest (lag4, lead1) [64].

Cardiovascular Disease Morbidity and Mortality

The primary outcome is the risk of any acute severe CVD events including mortality to identify cohort members who developed acute severe CVD events or who died from acute severe CVD events during the follow-up. Acute severe CVD events are defined based on the World Health Organization's Monitoring Trends and Determinants in Cardiovascular Disease Project, including acute coronary heart disease (ICD-9 code 410-412, 414), stroke (ICD-9 code 430-438), and heart failure (ICD-9 code 428) [26,65,66].

We could not access the cause of mortality among patients in this cohort owing to data limitation. Thus, we estimate the cause of mortality based on previous research by Bundy et al who offer a method to estimate the association between systolic blood pressure and CVD mortality [67]. We will apply this method to identify cohort members who died from CVD events during the follow-up. This method assumes that the number of CVD deaths could be increased if the population developed higher systolic blood pressure treatment levels. The number of CVD deaths will be computed using systolic blood pressure distributions and the population attributable risks (PARs) related

to systolic blood pressure level. A given PAR represents the proportion of CVD deaths that could be increased by higher systolic blood pressure levels. We will divide the most recent systolic blood pressure level into 8 categories (<130, 130-134, 135-139, 140-144, 145-149, 150-154, 155-159, and ≥160 mmHg). The PARs will be calculated using the formula given in Figure 4, where p_i is the proportion of the systolic blood pressure category i, HR_i is the hazard ratio of CVD deaths in the systolic blood pressure category i, and k is the total number of systolic blood pressure categories. To estimate hazard ratios for CVD mortality comparing each of the 8 systolic blood

Figure 4. Formula for population attributable risks (PARs).

$$PAR = \frac{\sum_{i=1}^{k} P_i(HR_i - 1)}{1 + \sum_{i=1}^{k} P_i(HR_i - 1)}$$

pressure categories, Bundy et al conducted a network meta-analysis of 42 antihypertensive clinical trials [68]. We will use the hazard ratio of CVD death from the meta-analysis study conducted by Bundy et al. For patients who died during the follow-up period, we assume that the patient died from CVD events if the PAR was higher than 50%.

Clinical Outcomes

The secondary outcomes are the most recent clinical treatment outcomes, including diastolic blood pressure, systolic blood pressure, HbA_{1c}, LDL-C, and TC.

Baseline Covariates and Other Covariates

Based on data availability, we will include the following personal-related factors: (1) patients' demographic and socioeconomic characteristics (ie, sex, age, body mass index, SES, and rurality), (2) risk factors (ie, smoking history, and alcohol history), (3) comorbidities and its duration by the end of the follow-up period (ie, chronic obstructive pulmonary disease, depression, dementia, and Parkinson), and (4) clinical outcomes at baseline (ie, diastolic blood pressure, systolic blood pressure, HbA_{1c}, LDL-C, and TC). Environmental determinants contributing to treatment adherence include those related to primary care physicians and health systems: (1) physicians' demographic characteristics (ie, sex, age, and location type) and (2) complexity of prescription (ie, the types of medication). All covariates will be treated as baseline covariates except that we will measure the incidence of comorbidities and the time to follow-up.

SES is defined according to the Canadian Material Deprivation Index [69]. The Canadian Material Deprivation Index, a proxy for individual-level SES based on the most recent 6-digit residential postal code, is calculated by the average income, percentage without high school graduation, and the employment ratio [70]. SES will be categorized into high, average, and low SES groups.

Statistical Analysis

The analyses will be described separately for objective 1 and objective 2. We will include participants with nonmissing information on treatment adherence. Multiple imputations will be used to replace missing data for baseline covariates, SES, and comorbidities. Finally, a practice site will be used as a random effect in each model. All analyses will be performed in Stata version 13.0 (Stata Corp LP).

Objective 1

For hypothesis 1, a time-varying Cox proportional hazards model will be performed to evaluate the hazard ratio between treatment adherence and the incidence of acute severe CVD and mortality, adjusting for all potential covariates. In addition, the interactions of adherence to multiple medications will be included as a block [71] in the Cox proportional hazards model

to examine the combined benefits of adherence in preventing CVD events and mortality (model 1). Treatment adherence and acute severe CVD risk is considered as the exposure of interest and outcomes, respectively. We will measure follow-up time (in months) from the date of entry into the cohort until the date of incident acute severe CVD events, died, or the end of the follow-up period. Kaplan-Meier survival curves for any acute severe CVD and mortality by treatment adherence level will be generated by pooling the survival estimates.

Objective 2

For hypothesis 2, 4 multivariable linear regression models will be used to test the association between treatment adherence and the most recent clinical outcomes after adjusting for all potential covariates (model 2-model 5). An adjusted risk ratio with 95% CI will be generated. We will build the hierarchical regression [71] into the linear regression models by including the adherence to different type of medications as a block to test the aggregated contribution of adherence to clinical outcomes, including diastolic blood pressure, systolic blood pressure, HbA_{1c}, LDL-C, and TC.

Sensitivity Analysis

Previous evidence has also indicated that there is no single correct lag and lead for estimating time-dependent association [63]. Lag and lead choices are a significantly important issue in the generalizability of results [62,72]. Therefore, we will conduct a sensitivity analysis using lag1 and lead 4 (Multimedia Appendix 2). Additionally, we will examine the consistency of the results regarding the following: (1) at different treatment adherence categories (ie, ≥80%, 60% to 80%, 40% to 60%, and ≤40%) and (2) with and without multiple imputation of missing data on all covariates.

Ethics Approval

The study was approved by the University of Toronto Research Ethics Board (reference number: 36065). Data were deidentified when analyzed.

Results

The project was funded in July 2017 under Diabetes Action Canada (reference number: 503854). The study was approved by the University of Toronto Research Ethics Board (reference number: 36065). The project was started in June 2018. The results are expected to be finished by September 2019.

Discussion

Strengths

CVD events represent a heavy disease burden on individuals and their families, the health system, and society in general. Improving treatment adherence to antihypertensive and antidiabetic medication has been well documented as an effective strategy to prevent CVD events. Compared with previous research, our study has several strengths. First, we will contribute new knowledge on the association between treatment adherence, acute CVD events, and clinical outcomes among patients with both diabetes and hypertension in the primary care setting. Second, we will examine the combined benefits of adhering to multiple medications in preventing CVD events and mortality and clinical treatment outcomes. Finally, our study provides further knowledge by addressing the limitations of previous studies, such as inclusion of important potential confounders such as comorbidities, their duration, and follow-up. For example, previous research reported that mental health conditions (ie, depression, anxiety, and dementia) were important factors when analyzing treatment adherence and CVD events but lacked the ability to identify such mental conditions [73,74]. Many studies have used the Charlson Comorbidity Index [27-29] to explore the combined effects of comorbidities (ie, chronic obstructive pulmonary disease, depression, chronic kidney disease, and dementia), but this approach does not reflect the role of individual comorbidities. Furthermore, these studies did not control for duration of comorbidities. Finally, baseline clinical characteristics such as HbA_{1c}, lipids, blood pressure, and body mass index are also important confounding factors, which may have a direct impact on CVD risk and clinical treatment outcomes [75]. However, few studies control for these factors. We will address the limitations by controlling body mass index, diastolic blood pressure, systolic blood pressure, HbA_{1c}, LDL-C, and TC. Owing to the data availability, the percentage of missing data at baseline is as follows: body mass index (25.25%), diastolic blood pressure (23.54%), systolic blood pressure (18.27%), HbA_{1c} (28.09%), LDL-C (46.21%), and TC (45.92%). We will use multiple imputation and consequently carry out a sensitivity analysis with and without

imputation of missing data on all covariates. Our study will be the first population-based cohort study that systematically investigates the impacts of treatment adherence for patients with both diabetes and hypertension on CVD morbidity and mortality, and clinical treatment outcomes using a longitudinal and large-scale primary care EMR data. Our findings will help to identify challenges in treatment adherence for patients with diabetes and hypertension in primary care settings. Through this study, we hope to provide valuable evidence for policy and practice to inform the design and implementation of primary care health services to support adherence among patients living with diabetes and hypertension.

Limitations

Several limitations should be noted in our study. First, this study is a retrospective study. All data were recorded from routine EMRs with possible errors and omissions. Thus, CVD may not be captured in full. As recording the medical diagnosis of diabetes, hypertension, and CVD is the responsibility of primary care physicians, there may be delays in EMR input. Second, treatment adherence is measured using a proxy, not a real measure. Theoretically, patients with both diabetes and hypertension are regarded as having a high risk of CVD events and should take prescribed medications consistently. We assume that prescription patterns correspond to medication-taking behavior. Third, owing to data limitations, we estimate CVD mortality using systolic blood pressure distributions and the population attributable risks related to the systolic blood pressure level. Fourth, aspirin is an over-the-counter medication that patients can obtain from pharmacies without a prescription. Thus, we may underestimate the adherence rates for aspirin. Finally, we employed the Social Cognitive Theory to explore the reciprocal relationship between that treatment adherence and diabetes/hypertension management outcomes (such as CVD), in the context of personal factors and environmental determinants. However, there were many factors not recorded in our database. For example, there were no variables such as primary language, ethnicity, health literacy, employment status, and marital status, which are factors contributing to patient understanding of the treatment or related to their daily management. There were no process variables recorded such as patient's level of involvement in the treatment decision-making process, understanding of their disease, and family and social support. In addition, there was a lack of reporting on physician-specific variables, such as the level of communication to patients on the benefits and adverse effects of a prescription nor did we have variables related to the health system such as access to primary care and primary care models [15]. These limitations may lead to potential bias.

XSL•FO

RenderX

Frank Sullivan's main affiliation is School of Medicine, University of St Andrews, and also Department of Family and Community Medicine and Dalla Lana School of Public Health, University of Toronto.

Michelle Greiver's main affiliation is North York General Hospital, and also Department of Family and Community Medicine, University of Toronto.

Authors' Contributions

XW obtained funding and led the study. XW, and MS conceived and designed the study. XW and MS managed the literature searches and formulated the research questions. MS, XW, and VH wrote the draft of the protocol. VH, RU, FS, FL, MG, and XW reviewed the protocol. All authors contributed to and approved the current version of the protocol

Conflicts of Interest

None declared.

Multimedia Appendix 1

Study characteristics.

[DOCX File, 149KB - resprot_v8i5e13571_app1.pdf]

Multimedia Appendix 2

Lag lead method.

[PDF File (Adobe PDF File), 361KB - resprot_v8i5e13571_app2.pdf]

References

1. Statistics Canada. Table 13-10-0394-01 Leading causes of death, total population, by age group URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310039401 [accessed 2019-04-30] [WebCite Cache ID 7817NCgwx]
2. Sarnak MJ, Levey AS, Schoolwerth AC, Coresh J, Culleton B, Hamm LL, American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, Epidemiology and Prevention. Kidney disease as a risk factor for development of cardiovascular disease: a statement from the American Heart Association Councils on Kidney in Cardiovascular Disease, High Blood Pressure Research, Clinical Cardiology, and Epidemiology and Prevention. Hypertension 2003 Nov;42(5):1050-1065. [doi: 10.1161/01.HYP.0000102971.85504.7c] [Medline: 14604997]
3. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and trends in diabetes among adults in the United States, 1988-2012. J Am Med Assoc 2015 Sep 8;314(10):1021-1029. [doi: 10.1001/jama.2015.10029] [Medline: 26348752]
4. Tobias DK, Stuart JJ, Li S, Chavarro J, Rimm EB, Rich-Edwards J, et al. Association of history of gestational diabetes with long-term cardiovascular disease risk in a large prospective cohort of US women. JAMA Intern Med 2017 Dec 01;177(12):1735-1742 [FREE Full text] [doi: 10.1001/jamainternmed.2017.2790] [Medline: 29049820]
5. Statistics Canada. Table13-10-0096-09 High blood pressure, by age group URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310009609 [accessed 2019-04-30] [WebCite Cache ID 7817adT4O]
6. Cloutier L, Lamarre-Cliche M. Hypertension in adults with type 2 diabetes: a review of blood pressure measurement methods, targets and therapy. Can J Diabetes 2018 Apr;42(2):188-195. [doi: 10.1016/j.jcjd.2018.01.012] [Medline: 29602406]
7. Cheung BM, Li C. Diabetes and hypertension: is there a common metabolic pathway? Curr Atheroscler Rep 2012 Apr;14(2):160-166 [FREE Full text] [doi: 10.1007/s11883-012-0227-2] [Medline: 22281657]
8. Kennedy-Martin T, Boye KS, Peng X. Cost of medication adherence and persistence in type 2 diabetes mellitus: a literature review. Patient Prefer Adherence 2017;11:1103-1117 [FREE Full text] [doi: 10.2147/PPA.S136639] [Medline: 28721024]
9. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004 May;27(5):1047-1053. [Medline: 15111519]
10. Egan BM, Li J, Qanungo S, Wolfman TE. Blood pressure and cholesterol control in hypertensive hypercholesterolemic patients: national health and nutrition examination surveys 1988-2010. Circulation 2013 Jul 02;128(1):29-41 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.112.000500] [Medline: 23817481]
11. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
12. National Heart, Lung, and Blood Institute. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure URL: https://www.nhlbi.nih.gov/files/docs/guidelines/jnc7full.pdf [accessed 2019-05-06] [WebCite Cache ID 78A27Tyut]
13. The sixth report of the Joint National Committee. The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. Arch Intern Med 1997 Nov 24;157(21):2413-2446. [Medline: 9385294]

XSL•FO

RenderX

14. Yiannakopoulou EC, Papadopulos JS, Cokkinos DV, Mountokalakis TD. Adherence to antihypertensive treatment: a critical factor for blood pressure control. Eur J Cardiovasc Prev Rehabil 2005 Jun;12(3):243-249 [FREE Full text] [Medline: 15942423]

15. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc 2011 Apr;86(4):304-314 [FREE Full text] [doi: 10.4065/mcp.2010.0575] [Medline: 21389250]

16. Cramer JA, Roy A, Burrell A, Fairchild CJ, Fuldeore MJ, Ollendorf DA, et al. Medication compliance and persistence: terminology and definitions. Value Health 2008;11(1):44-47 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00213.x] [Medline: 18237359]

17. Lam WY, Fresco P. Medication adherence measures: an overview. Biomed Res Int 2015;2015:217047 [FREE Full text] [doi: 10.1155/2015/217047] [Medline: 26539470]

18. Elwyn G, Edwards A, Britten N. What information do patients need about medicines? "Doing prescribing": how doctors can be more effective. Br Med J 2003 Oct 11;327(7419):864-867 [FREE Full text] [doi: 10.1136/bmj.327.7419.864] [Medline: 14551110]

19. McDonald HP, Garg AX, Haynes RB. Interventions to enhance patient adherence to medication prescriptions: scientific review. J Am Med Assoc 2002 Dec 11;288(22):2868-2879. [Medline: 12472329]

20. Moon SJ, Lee W, Hwang JS, Hong YP, Morisky DE. Accuracy of a screening tool for medication adherence: a systematic review and meta-analysis of the Morisky Medication Adherence Scale-8. PLoS One 2017;12(11):e0187139 [FREE Full text] [doi: 10.1371/journal.pone.0187139] [Medline: 29095870]

21. Hennein R, Hwang S, Au R, Levy D, Muntner P, Fox CS, et al. Barriers to medication adherence and links to cardiovascular disease risk factor control: the Framingham Heart Study. Intern Med J 2018 Dec;48(4):414-421. [doi: 10.1111/imj.13687] [Medline: 29193523]

22. Morisky DE, Green LW, Levine DM. Concurrent and predictive validity of a self-reported measure of medication adherence. Med Care 1986 Jan;24(1):67-74. [Medline: 3945130]

23. Barner JC. ISPOR Student Forum Presentation. Medication adherence: focus on secondary database analysis URL: https://docplayer.net/41237212-Medication-adherence-focus-on-secondary [accessed 2019-04-29] [WebCite Cache ID 7817sIVcr]

24. van Galen KA, Nellen JF, Nieuwkerk PT. The effect on treatment adherence of administering drugs as fixed-dose combinations versus as separate pills: systematic review and meta-analysis. AIDS Res Treat 2014;2014:967073 [FREE Full text] [doi: 10.1155/2014/967073] [Medline: 25276422]

25. Hess LM, Raebel MA, Conner DA, Malone DC. Measurement of adherence in pharmacy administrative databases: a proposal for standard definitions and preferred measures. Ann Pharmacother 2006;40(7-8):1280-1288. [doi: 10.1345/aph.1H018] [Medline: 16868217]

26. Karlsson SA, Hero C, Svensson A, Franzén S, Miftaraj M, Gudbjörnsdottir S, et al. Association between refill adherence to lipid-lowering medications and the risk of cardiovascular disease and mortality in Swedish patients with type 2 diabetes mellitus: a nationwide cohort study. BMJ Open 2018 Dec 30;8(3):e020309 [FREE Full text] [doi: 10.1136/bmjopen-2017-020309] [Medline: 29602853]

27. Kim J, Kim E, Lee E. Evaluation of the chronic disease management program for appropriateness of medication adherence and persistence in hypertension and type-2 diabetes patients in Korea. Medicine (Baltimore) 2017 Apr;96(14):e6577 [FREE Full text] [doi: 10.1097/MD.0000000000006577] [Medline: 28383439]

28. Shin S, Song H, Oh S, Choi KE, Kim H, Jang S. Effect of antihypertensive medication adherence on hospitalization for cardiovascular disease and mortality in hypertensive patients. Hypertens Res 2013 Nov;36(11):1000-1005. [doi: 10.1038/hr.2013.85] [Medline: 23966057]

29. Yang Q, Chang A, Ritchey MD, Loustalot F. Antihypertensive medication adherence and risk of cardiovascular disease among older adults: a population-based cohort study. J Am Heart Assoc 2017 Jun 24;6(6) [FREE Full text] [doi: 10.1161/JAHA.117.006056] [Medline: 28647688]

30. Perreault S, Dragomir A, White M, Lalonde L, Blais L, Bérard A. Better adherence to antihypertensive agents and risk reduction of chronic heart failure. J Intern Med 2009 Aug;266(2):207-218. [Medline: 19623691]

31. Roy L, White-Guay B, Dorais M, Dragomir A, Lessard M, Perreault S. Adherence to antihypertensive agents improves risk reduction of end-stage renal disease. Kidney Int 2013 Sep;84(3):570-577 [FREE Full text] [doi: 10.1038/ki.2013.103] [Medline: 23698228]

32. Liu Q, Quan H, Chen G, Qian H, Khan N. Antihypertensive medication adherence and mortality according to ethnicity: a cohort study. Can J Cardiol 2014 Aug;30(8):925-931. [doi: 10.1016/j.cjca.2014.04.017] [Medline: 25064583]

33. Verma AA, Khuu W, Tadrous M, Gomes T, Mamdani MM. Fixed-dose combination antihypertensive medications, adherence, and clinical outcomes: a population-based retrospective cohort study. PLoS Med 2018 Jun;15(6):e1002584 [FREE Full text] [doi: 10.1371/journal.pmed.1002584] [Medline: 29889841]

34. Dragomir A, Côté R, Roy L, Blais L, Lalonde L, Bérard A, et al. Impact of adherence to antihypertensive agents on clinical outcomes and hospitalization costs. Med Care 2010 May;48(5):418-425. [doi: 10.1097/MLR.0b013e3181d567bd] [Medline: 20393367]

XSL•FO

RenderX

35.   Wu P, Yang C, Yao Z, Lin W, Wu L, Chang C. Relationship of blood pressure control and hospitalization risk to medication adherence among patients with hypertension in Taiwan. Am J Hypertens 2010 Feb;23(2):155-160. [doi: 10.1038/ajh.2009.210] [Medline: 19927135]

36.   Raebel MA, Dyer W, Nichols GA, Goodrich GK, Schmittdiel JA. Relationships between medication adherence and cardiovascular disease risk factor control in elderly patients with diabetes. Pharmacotherapy 2017 Oct;37(10):1204-1214 [FREE Full text] [doi: 10.1002/phar.1994] [Medline: 28752555]

37.   Frias J, Virdi N, Raja P, Kim Y, Savage G, Osterberg L. Effectiveness of digital medicines to improve clinical outcomes in patients with uncontrolled hypertension and type 2 diabetes: prospective, open-label, cluster-randomized pilot clinical trial. J Med Internet Res 2017 Jul 11;19(7):e246 [FREE Full text] [doi: 10.2196/jmir.7833] [Medline: 28698169]

38.   Cohen HW, Shmukler C, Ullman R, Rivera CM, Walker EA. Measurements of medication adherence in diabetic patients with poorly controlled HbA(1c). Diabet Med 2010 Feb;27(2):210-216 [FREE Full text] [doi: 10.1111/j.1464-5491.2009.02898.x] [Medline: 20546266]

39.   Krapek K, King K, Warren SS, George KG, Caputo DA, Mihelich K, et al. Medication adherence and associated hemoglobin A1c in type 2 diabetes. Ann Pharmacother 2004 Sep;38(9):1357-1362. [doi: 10.1345/aph.1D612] [Medline: 15238621]

40.   Lee JK, Grace KA, Taylor AJ. Effect of a pharmacy care program on medication adherence and persistence, blood pressure, and low-density lipoprotein cholesterol: a randomized controlled trial. J Am Med Assoc 2006 Dec 06;296(21):2563-2571. [doi: 10.1001/jama.296.21.joc60162] [Medline: 17101639]

41.   Gentil L, Vasiliadis H, Préville M, Berbiche D. Adherence to Oral Antihyperglycemic Agents Among Older Adults With Mental Disorders and Its Effect on Health Care Costs, Quebec, Canada, 2005-2008. Prev Chronic Dis 2015 Dec 31;12:E230 [FREE Full text] [doi: 10.5888/pcd12.150412] [Medline: 26719900]

42.   Simard P, Presse N, Roy L, Dorais M, White-Guay B, Räkel A, et al. Persistence and adherence to oral antidiabetics: a population-based cohort study. Acta Diabetol 2015 Jun;52(3):547-556. [doi: 10.1007/s00592-014-0692-x] [Medline: 25524433]

43.   Guénette L, Breton M, Guillaumie L, Lauzier S, Grégoire J, Moisan J. Psychosocial factors associated with adherence to non-insulin antidiabetes treatments. J Diabetes Complications 2016 Mar;30(2):335-342. [doi: 10.1016/j.jdiacomp.2015.10.016] [Medline: 26615905]

44.   Grégoire J, Sirois C, Blanc G, Poirier P, Moisan J. Persistence patterns with oral antidiabetes drug treatment in newly treated patients--a population-based study. Value Health 2010;13(6):820-828 [FREE Full text] [doi: 10.1111/j.1524-4733.2010.00761.x] [Medline: 21054658]

45.   Friedman O, McAlister FA, Yun L, Campbell NR, Tu K, Canadian Hypertension Education Program Outcomes Research Taskforce. Antihypertensive drug persistence and compliance among newly treated elderly hypertensives in ontario. Am J Med 2010 Feb;123(2):173-181. [doi: 10.1016/j.amjmed.2009.08.008] [Medline: 20103027]

46.   Gee ME, Campbell NR, Gwadry-Sridhar F, Nolan RP, Kaczorowski J, Bienek A, Outcomes Research Task Force of the Canadian Hypertension Education Program. Antihypertensive medication use, adherence, stops, and starts in Canadians with hypertension. Can J Cardiol 2012 May;28(3):383-389. [doi: 10.1016/j.cjca.2012.01.014] [Medline: 22522073]

47.   Natarajan N, Putnam W, Van Aarsen K, Beverley Lawson K, Burge F. Adherence to antihypertensive medications among family practice patients with diabetes mellitus and hypertension. Can Fam Physician 2013 Feb;59(2):e93-e100 [FREE Full text] [Medline: 23418264]

48.   Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 7;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]

49.   Aagaard EM, Gonzales R, Camargo CA, Auten R, Levin SK, Maselli J, et al. Physician champions are key to improving antibiotic prescribing quality. Jt Comm J Qual Patient Saf 2010 Mar;36(3):109-116. [Medline: 20235412]

50.   Wei X, Walley JD, Zhang Z, Zou G, Gong W, Deng S, et al. Implementation of a comprehensive intervention for patients at high risk of cardiovascular disease in rural China: a pragmatic cluster randomized controlled trial. PLoS One 2017;12(8):e0183169 [FREE Full text] [doi: 10.1371/journal.pone.0183169] [Medline: 28813512]

51.   Zullig LL, Blalock DV, Dougherty S, Henderson R, Ha CC, Oakes MM, et al. The new landscape of medication adherence improvement: where population health science meets precision medicine. Patient Prefer Adherence 2018;12:1225-1230 [FREE Full text] [doi: 10.2147/PPA.S165404] [Medline: 30034226]

52.   De Geest S, Sabaté E. Adherence to long-term therapies: evidence for action. Eur J Cardiovasc Nurs 2003 Dec;2(4):323. [Medline: 14667488]

53.   Ong SE, Koh JJ, Toh SE, Chia KS, Balabanova D, McKee M, et al. Assessing the influence of health systems on type 2 diabetes mellitus awareness, treatment, adherence, and control: a systematic review. PLoS One 2018;13(3):e0195086 [FREE Full text] [doi: 10.1371/journal.pone.0195086] [Medline: 29596495]

54.   Zullig LL, Peterson ED, Bosworth HB. Ingredients of successful interventions to improve medication adherence. J Am Med Assoc 2013 Dec 25;310(24):2611-2612. [doi: 10.1001/jama.2013.282818] [Medline: 24264605]

55.   Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]

56.   Bandura A. Exercise of human agency through collective efficacy. Curr Dir Psychol Sci 2000;9(3):75-78. [doi: 10.1111/1467-8721.00064]

57.   Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]

58.   Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [Medline: 6392204]

59.   Ajzen I, Madden TJ. Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control. J Exp Soc Psychol 1986 1986 Sep;22(5):453-474. [doi: 10.1016/0022-1031(86)90045-4]

60.   Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]

61.   Stokols D. Translating social ecological theory into guidelines for community health promotion. Am J Health Promot 1996;10(4):282-298. [Medline: 10159709]

62.   Guirguis-Blake J, Keppel GA, Force RW, Cauffield J, Monger RM, Baldwin LM. Variation in refill protocols and procedures in a family medicine residency network. Fam Med 2012 Sep;44(8):564-568 [FREE Full text] [Medline: 22930121]

63.   Sturman M, McCabe D. Choosing whether to lead, lag, or match the market. Journal of Human Resources in Hospitality & Tourism 2006;7(1):85-97. [doi: 10.1300/J171v07n01_05]

64.   Rigobon AV, Kalia S, Nichols J, Aliarzadeh B, Greiver M, Moineddin R, et al. Impact of the Diabetes Canada Guideline Dissemination Strategy on the prescription of vascular protective medications: a retrospective cohort study, 2010-2015. Diabetes Care 2019 Jan;42(1):148-156. [doi: 10.2337/dc18-0935] [Medline: 30389671]

65.   Chen H, Burnett RT, Kwong JC, Villeneuve PJ, Goldberg MS, Brook RD, et al. Spatial association between ambient fine particulate matter and incident hypertension. Circulation 2014 Feb 04;129(5):562-569. [doi: 10.1161/CIRCULATIONAHA.113.003532] [Medline: 24190962]

66.   Tu JV, Chu A, Rezai MR, Guo H, Maclagan LC, Austin PC, et al. The incidence of major cardiovascular events in immigrants to Ontario, Canada: the CANHEART immigrant study. Circulation 2015 Dec 20;132(16):1549–1559 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.015345] [Medline: 26324719]

67.   Bundy JD, Mills KT, Chen J, Li C, Greenland P, He J. Estimating the association of the 2017 and 2014 Hypertension Guidelines with cardiovascular events and deaths in US adults. JAMA Cardiol 2018 Jul 01;3(7):572. [doi: 10.1001/jamacardio.2018.1240] [Medline: 29800138]

68.   Bundy JD, Li C, Stuchlik P, Bu X, Kelly TN, Mills KT, et al. Systolic blood pressure reduction and risk of cardiovascular disease and mortality. JAMA Cardiol 2017 Jul 01;2(7):775. [doi: 10.1001/jamacardio.2017.1421]

69.   Canadian Institute for Health Information. 2008. Reducing gaps in health: a focus on socio-economic status in urban Canada URL: https://secure.cihi.ca/free_products/Reducing_Gaps_in_Health_Report_EN_081009.pdf [accessed 2019-05-06] [WebCite Cache ID 78A3gj0rb]

70.   Pampalon R, Hamel D, Gamache P, Philibert MD, Raymond G, Simpson A. An area-based material and social deprivation index for public health in Québec and Canada. Can J Public Health 2012 Apr 30;103(8 Suppl 2):S17-S22 [FREE Full text] [Medline: 23618066]

71.   Wei X, Barnsley J, Zakus D, Cockerill R, Glazier R, Sun X, et al. Evaluation of a diabetes management program in China demonstrated association of improved continuity of care with clinical outcomes. J Clin Epidemiol 2008 Sep;61(9):932-939. [doi: 10.1016/j.jclinepi.2007.12.013] [Medline: 18619807]

72.   Birtwhistle R, Keshavjee K, Lambert-Lanning A, Godwin M, Greiver M, Manca D, et al. Building a pan-Canadian primary care sentinel surveillance network: initial development and moving forward. J Am Board Fam Med 2009;22(4):412-422 [FREE Full text] [doi: 10.3122/jabfm.2009.04.090081] [Medline: 19587256]

73.   Chao J, Nau DP, Aikens JE, Taylor SD. The mediating role of health beliefs in the relationship between depressive symptoms and medication adherence in persons with diabetes. Res Social Adm Pharm 2005 Dec;1(4):508-525. [doi: 10.1016/j.sapharm.2005.09.002] [Medline: 17138493]

74.   Kilbourne A, Reynolds IC, Good C, Sereika S, Justice A, Fine MJ. How does depression influence diabetes medication adherence in older patients? Am J Geriatr Psychiatry 2005 Mar;13(3):202-210. [doi: 10.1176/appi.ajgp.13.3.202] [Medline: 15728751]

75.   Ekström N, Svensson A, Miftaraj M, Franzén S, Zethelius B, Eliasson B, et al. Cardiovascular safety of glucose-lowering agents as add-on medication to metformin treatment in type 2 diabetes: report from the Swedish National Diabetes Register. Diabetes Obes Metab 2016 Jan;18(10):990-998. [doi: 10.1111/dom.12704] [Medline: 27282621]

Abbreviations

CPCSSN: Canadian Primary Care Sentinel Surveillance Network
CVD: cardiovascular disease
EMR: electronic medical record
HbA$_{1c}$: glycated hemoglobin
LDL-C: low-density lipoprotein cholesterol
PAR: population attributable risk

XSL•FO
RenderX

SES: socioeconomic status
TC: total cholesterol

XSL•FO
RenderX

Protocol

Clinic-Based Delivery of the Young Men's Health Project (YMHP) Targeting HIV Risk Reduction and Substance Use Among Young Men Who Have Sex with Men: Protocol for a Type 2, Hybrid Implementation-Effectiveness Trial

Jeffrey T Parsons[1,2,3], PhD; Tyrel Starks[1,2,3], PhD; Sitaji Gurung[1], MD, MPH; Demetria Cain[1], PhD, MPH; Jonathan Marmo[1], BA; Sylvie Naar[4], PhD

[1]Center for HIV Educational Studies and Training, Hunter College, City University of New York, New York, NY, United States

[2]Hunter Department of Psychology, Hunter College, City University of New York, New York, NY, United States

[3]Health Psychology and Clinical Science Doctoral Program, Graduate Center, City University of New York, New York, NY, United States

[4]College of Medicine, Florida State University, Tallahassee, FL, United States

Corresponding Author:
Sylvie Naar, PhD
College of Medicine
Florida State University
Main Campus
1115 West Call Street
Tallahassee, FL, 32306
United States
Phone: 1 248 207 2903
Email: sylvie.naar@med.fsu.edu

Abstract

Background: Young men who have sex with men (YMSM) are disproportionately at risk for HIV and sexually transmitted infections. Adapting and testing the effectiveness of the Young Men's Health Project (YMHP), an efficacious intervention designed to reduce substance use and condomless anal sex (CAS) among YMSM, at clinics in Miami, Detroit, and Philadelphia has the potential to reduce HIV and STI disparities among urban YMSM.

Objective: This study (Adolescent Medicine Trials Network for HIV/AIDS Interventions [ATN] 145 YMHP) aims to adapt YMHP for clinic and remote delivery by existing clinic staff and compare their effectiveness in real-world adolescent HIV clinics. This protocol is part of the ATN *Scale It Up* program described in a recently published article by Naar et al.

Methods: This is a comparative effectiveness hybrid type-2 trial of the YMHP intervention with 2 delivery formats—clinic-based versus remote delivery—offered following HIV counseling and testing. Phase 1 includes conducting focus groups with youth to obtain implementation feedback about the delivery of the YMHP intervention and intervention components to ensure culturally competent, feasible, and scalable implementation. Phase 2 includes recruitment and enrollment of 270 YMSM, aged 15 to 24 years, 90 at each of the 3 sites. Enrollment will be limited to HIV-negative YMSM who report recent substance use and either CAS or a positive STI test result. Participants will be randomized to receive the YMHP intervention either in person or by remote delivery. Both conditions involve completion of the 4 YMHP sessions and the delivery of pre-exposure prophylaxis information and navigation services. A minimum of 2 community health workers (CHWs) will be trained to deliver the intervention sessions at each site. Sessions will be audio-recorded for Motivational Interviewing Treatment Integrity (MITI) fidelity coding, and CHWs and supervisors will be given implementation support throughout the study period.

Results: Phase 1 focus groups were completed in July 2017 (n=25). Feedback from these focus groups at the 3 sites informed adaptations to the YMHP intervention manual, implementation of the intervention, and recruitment plans for phase 2. Baseline enrollment for phase 2 began in November 2018, and assessments will be at immediate posttest (IP)-, 3-, 6-, 9-, and 12-months after the intervention. Upon collection of both baseline and follow-up data, we will compare the effectiveness and cost-effectiveness of clinic-based versus remote delivery of YMHP in the context of health care access.

XSL•FO
RenderX

Conclusions: We are conducting YMHP in 3 cities with high rates of YMSM at risk for HIV and STIs. When adapted for real-world clinics, this study will help substance-using YMSM at risk for HIV and STIs and allow us to examine differences in effectiveness and cost by the method of delivery.

Trial Registration: ClinicalTrials.gov NCT03488914; https://clinicaltrials.gov/ct2/show/NCT03488914 (Archived by WebCite at http://www.webcitation.org/770WaWWfi)

International Registered Report Identifier (IRRID): DERR1-10.2196/11184

Introduction

Background

Young men who have sex with men (YMSM; aged 16-24 years) are disproportionately at risk for HIV and sexually transmitted infections (STIs). Although new HIV infections have fallen or remained stable among other groups, YMSM have experienced a 132% increase in new infections since 2002 [1]. From 2008 to 2011, YMSM aged 13 to 24 years had the greatest percentage increase (22%) in new HIV infections [2]. YMSM of color are especially at risk: in 2011, among YMSM aged 13 to 24 years with HIV infection, 58% were black and 20% were Latino [2].

Young males aged 15 to 24 years are vastly overrepresented in rates of STIs. In 2013, males aged 15 to 24 years accounted for more than half (57%) of all male cases of Chlamydia trachomatis (CT) infection and 46% of Neisseria gonorrhea (GC) [3]. Young men aged 20 to 24 years had the highest rate of syphilis from 2008 to 2012 [3]. Furthermore, sexual orientation disparities exist—for example, men who have sex with men (MSM) accounted for 75% of all syphilis cases in 2013 [3]. Ethnic and racial disparities exist in the incidence of other STIs among YMSM [3]. Young black men aged 15 to 24 years have a rate of CT infection 5.5-9.5 times higher and GC rates 10.4-13.0 times higher than white men [3]. A recent study of HIV-negative MSM diagnosed with rectal CT or GC at STI clinics between 2008 and 2010 showed that such infections greatly increase HIV incidence [4]. In 2013, MSM accounted for 3 quarters of all primary and secondary syphilis cases diagnosed in the United States—an increase of 10% since 2012 [3].

Location is also a risk factor. YMSM in large urban areas are disproportionally affected. Detroit, Philadelphia, and Miami have high rates of HIV and STIs among males. These cities are geographically diverse and have large numbers of YMSM receiving HIV counseling and testing annually by clinic sites and community partners. This information points to the need to study the effectiveness of the Young Men's Health Project (YMHP), for its significant public health potential in reducing HIV/STI disparities among urban YMSM.

Evidence Base

Substance Use and HIV Risk Among Young Men Who Have Sex With Men

MSM use substances at higher rates than the general population, increasing HIV risk. High rates of drug and alcohol use among MSM relative to the general population have been documented [5-7], and our research has identified higher rates of drug use among YMSM compared with heterosexuals [8-10]. Higher rates of drug use have also been documented among MSM, including YMSM in tandem with sexual activity [11,12]. However, drug-use patterns seem to differ among YMSM with increased rates of cocaine among YMSM, which could have implications for HIV risk given that stimulant use has been linked to condomless anal sex (CAS) [13-16] and higher risk of HIV infection and other STIs [17]. Nearly half of black MSM with newly diagnosed HIV infection (48%) reported substance use during their last anal sex encounter [11]. Substance use has been found to increase sexual risk behavior among MSM [18], placing them at high risk for CAS and HIV seroconversion, exchange sex, and greater number of sexual partners [19]. Our own research using event-level data for the previous 30 days has shown that substance use strongly and significantly predicts the odds of whether YMSM will use a condom [20]. A number of other studies have looked at the impact of substance use on sexual behavior and increased odds of seroconversion [11]. Different substances are associated with sexual risk behavior among specific groups. Among Latino MSM, methamphetamine use [21,22] and among black MSM higher rates of marijuana use have been linked to sexual risk [23]. Therefore, there is a critical need for brief, culturally appropriate, effective behavioral interventions that improve self-management to reduce new HIV infections among substance-using YMSM.

Effectiveness of Motivational Interviewing

Motivational interviewing (MI) has the potential to improve self-management behaviors in terms of promoting sexual health and reducing substance use among YMSM. There is strong evidence that MI is a culturally appropriate and effective approach for working with racial and ethnic minority populations [24] who are disproportionately affected by HIV. One meta-analysis of MI found a greater effect among minorities [25]. MI has been recommended as particularly effective when working with YMSM [26]. MI promotes increased intrinsic motivation to change and, when paired with information regarding health risk behaviors, reinforces the individuals' right and capacity to make well-informed health self-management decisions for themselves [27].

YMHP, a manualized structured 4-session intervention using MI and problem-solving skills building, has been listed as a best evidence intervention by the Center for Disease Control and Prevention (CDC) and is the only intervention to reduce

both CAS and substance use among YMSM [28]. The CDC's endorsement of YMHP was informed by results of the National Institutes of Health (NIH)-funded efficacy trial. A total of 143 HIV-negative YMSM (aged 18-29 years) who reported CAS with a high-risk partner and at least 5 days of drug use in the last 3 months were randomized to receive 4 sessions of YMHP or 4 sessions of health education on sexual risk and substance use. The majority [63% (90/145)] of the sample were YMSM of color (black: 21%, Latino: 28.7%, or multiracial: 13.3%). Retention rates were high: 79% at the 12-month follow-up and 88% completed at least 3 of the 4 sessions. Master's level therapists were trained in MI with ongoing fidelity monitoring using the Motivational Interviewing Treatment Integrity (MITI) coding system. YMSM who received YMHP reduced their cross-time averaged odds of ever using any drug by 67% over the 1-year follow-up. Within-condition analyses showed that YMSM who received YMHP reduced their cross-time averaged odds of ever having CAS by 83% over the 1-year follow-up. YMSM in the YMHP condition were 21% less likely to report CAS on days when they did report drug use relative to men receiving education. This has been the first and only RCT with HIV-negative YMSM to show significant reductions in both CAS and substance use. As such, YMHP has the potential to have a significant impact on YMSM seeking sexual health or HIV counseling and testing services at clinics.

Remote Delivery

Research on telephone-based MI has consistently found that it produces significant improvements in a wide range of physical health challenges [29-31]. Furthermore, telephone-based MI reduces mental health–related problems [32] and alcohol-related problems [33] as well as sexual risk taking [31,34-36] among HIV-positive individuals. Research comparing the effects of telephone-based MI with face-to-face (clinic-delivered) interventions has produced equivocal results. Across studies examining physical activity, mental health, and substance use outcomes, findings suggest no significant differences in delivery method [34-37]. Carey et al examined the relative efficacy of a telephone versus face-to-face intervention for alcohol use and observed a significant interaction of delivery method with gender [38]. Women on average had better outcome in the face-to-face condition, whereas men responded equally well to both delivery methods.

Notably, the issue of health care access has not yet been examined as a moderator of relative effectiveness. One advantage of telephone-based MI is that it significantly reduces patient burden. It is, therefore, plausible that it will show superior effects among YMSM who experience barriers to health care access. However, it is also possible that remote delivery (via phone or video chat using Skype or FaceTime) will decrease engagement and the quality of the relationship between the clinicians and the participant. This might result in clinic-based delivery being superior among YMSM who have better access to health care. Understanding how health care access intersects with delivery method will substantially inform implementation decisions at clinics and other agencies seeking to utilize YMHP.

Community Health Workers Intervention Delivery

Integrating implementation science into a comparative effectiveness trial (CET) can minimize the science-practice gap [39]. MI providers need not be clinicians—1 study conducted by our team comparing community health workers (CHWs) with clinicians found both were equally effective in providing high-quality MI and that clients were more likely to be retained in HIV care when working with CHWs [40]. CHWs are commonly integrated into clinics and often play a central role in providing HIV prevention services, including HIV counseling and testing. Training CHWs to deliver evidence-based interventions is a critical step toward realistic and cost-effective implementation [41]. CDC has called for expanded use of CHWs in services for chronic disease [42] with attention to implementation and training [43-50].

Several steps can be taken in a CET of YMHP to promote adoption and sustainability. First, using staff embedded in the existing clinic setting can build capacity for implementation. Second, CHWs have long been the cornerstone of integrating support services into HIV-related prevention efforts [51]. Training CHWs to deliver interventions is a critical step toward realistic and cost-effective implementation. Research has documented the amount of training needed to obtain MI fidelity, concluding that initial training followed by ongoing coaching is required [52-55]. Such training can be costly when relying on outside trainers. Thus, a *train the trainer* model where expert trainers provide local supervisors with MI coaching skills might be more sustainable [56]. Finally, CET designs help gain information about implementation [57]. In a *Hybrid 2* trial, the goal is to dually determine which treatments work in which settings and to simultaneously answer implementation science questions about the potential barriers/facilitators to a treatment's widespread and continued implementation.

Aims

The aim of this paper is to describe Adolescent Medicine Trials Network for HIV/AIDS Interventions (ATN) 145 YMHP to study the scale-up of evidence-based practices in multidisciplinary adolescent HIV care settings while balancing flexibility and fidelity. The protocol is part of the *Scale it Up* research program focusing on implementation of self-management interventions to impact the adolescent HIV prevention and care cascades [58]. The purpose of this study is to adapt YMHP for clinic and remote delivery by existing HIV clinic staff, CHWs, who work with YMSM aged 15 to 24 years. This study also aims to (1) compare the effectiveness of clinic-based versus remote delivery of YMHP in the context of health care access by hypothesizing that remote-delivery will yield significantly better results among youth who experience barriers to health care access, whereas clinic-based delivery will yield significantly better results among youth who do not, (2) assess the cost-effectiveness of both formats of delivery to increase the likelihood of uptake for this intervention, and (3) assess the 5 components of the self-management model (ie, problem solving, decision making, resource utilization, forming of a patient and health care provider partnership, and taking action) and how these components vary over time, are directly improved by the interventions, and mediate intervention effects.

Finally, a sustainable model of YMHP will be tested in real-world adolescent clinics utilizing the *Exploration, Preparation, Implementation, Sustainment model (EPIS;* see ATN 153 EPIS protocol paper [59]).

Methods

Overview of Content and Delivery

This is a comparative effectiveness hybrid type-2 trial of the YMHP intervention designed to achieve aims over 2 phases. Phase 1 was completed by conducting focus groups with youth as part of its formative research to obtain implementation feedback about the delivery of the YMHP intervention and intervention components to ensure culturally competent, feasible, and scalable implementation both in the clinic setting and via remote condition, when delivered by CHWs. For phase 2, the target sample size is 270 YMSM, aged 15 to 24 years, 90 at each of the 3 sites. Participants are randomized at the end of their baseline assessment into 1 of the 2 intervention conditions: (1) delivery of the YMHP intervention in person at the clinic site or (2) delivery of the YMHP intervention by phone or video chat using apps such as Skype or FaceTime. Everyone will receive session 1 in person immediately upon completion of the baseline assessment regardless of their randomization to the 2 intervention conditions. Those randomized to face-to-face clinic-based delivery will schedule and complete their remaining 3 sessions in person, whereas those randomized to the remote condition will schedule and complete the remaining 3 sessions remotely.

The study will employ a stratified randomization procedure based on city, minority status, and health care access, so that within each city, youth who experience barriers to health care access are distributed equally across conditions. We will also stratify by interventionist and whether a youth has used marijuana only versus other drugs in the past 3 months. Research staff will complete a survey on Qualtrics that will randomize participants into 1 of the 2 study groups with no masking to the research staff or participant. Intervention sessions will occur approximately once per week for 4 weeks beginning 1 week after completion of the baseline assessment. There is a 12-week window for intervention completion, that is, all 4 sessions must be completed within 12 weeks of the baseline assessment. The immediate posttest (IP) assessment will occur 3 months after baseline and subsequent follow-ups will be scheduled 3 months thereafter until the 12-month postintervention (ie, 15 months postbaseline) assessment. Baseline, 3-month, and 9-month assessments consist of self-report data collection (CASI) and HIV and STI testing for CT, GC, and syphilis. The IP, 6-month, and 12-month follow-up assessments require completing the CASI only and can be completed remotely via email Qualtrics

link. All enrolled participants could receive up to US $275 by the end of this study—US $50 for baseline, which includes CASI, bio-testing, and session 1; US $30 for sessions 2 to 4; US $20 as a bonus for completing all 4 sessions; US $25 for CASI-only assessments (ie, for IP, 6-month, and 12-month assessments); and US $50 for assessments with both CASI and bio-testing (ie, for 3-month and 9-month assessments). The participant flow diagram is presented in Figure 1.

Recruitment and Eligibility

Recruitment

The Center for HIV Educational Studies and Training (CHEST) at Hunter College will use a variety of recruitment strategies to recruit participants for this study. First, as a result of CHEST's role as the Management Core and Recruitment and Retention Center (REC) for *Scale It Up*, we will utilize 3 Subject Recruitment Venues (SRVs) in our network (Wayne State University Prevention in Detroit, University of Miami, and Children's Hospital of Philadelphia) to complete clinic- and field-based recruitment. All 3 SRVs, as well as CHEST, have extensive relations with the gay, lesbian, bisexual, and transgender communities; community service organizations; health service organizations; and providers for MSM. In this aspect, recruitment will occur from routine walk-in visits for HIV testing. Partnering clinics will include offering this study through outreach activities and through their HIV testing services they provide in their clinics. Information about the study will be included in the institutional review board (IRB)-approved palm cards, brochures, and flyers at each clinic site (see Figure 2). This information will be displayed in waiting rooms and exam rooms at the clinic as well as at locations of mobile testing events and outreach shifts to be passed to potential participants to both encourage HIV testing and promote the YMHP study. This method has worked well in the past for numerous studies and takes advantage of when patients have more idle time to learn about the study.

YMSM who test HIV negative at the 3 SRVs in Miami, Detroit, and Philadelphia or through mobile testing efforts provided by community collaborators will be offered the opportunity to participate in YMHP. Those interested will be asked to complete a brief *Study Screener* on a study iPad to collect demographic and behavior questions related to eligibility criteria. Participant ID numbers are generated and assigned through Qualtrics during the screening process to all potential participants, including those who screen ineligible. If eligible, a screen will be displayed informing the potential participant of his eligibility, and the site study staff person will then schedule the potential participant for a baseline assessment. We anticipate enrollment of 5 participants per month, per SRV, and have allocated staffing resources to ensure this rate.

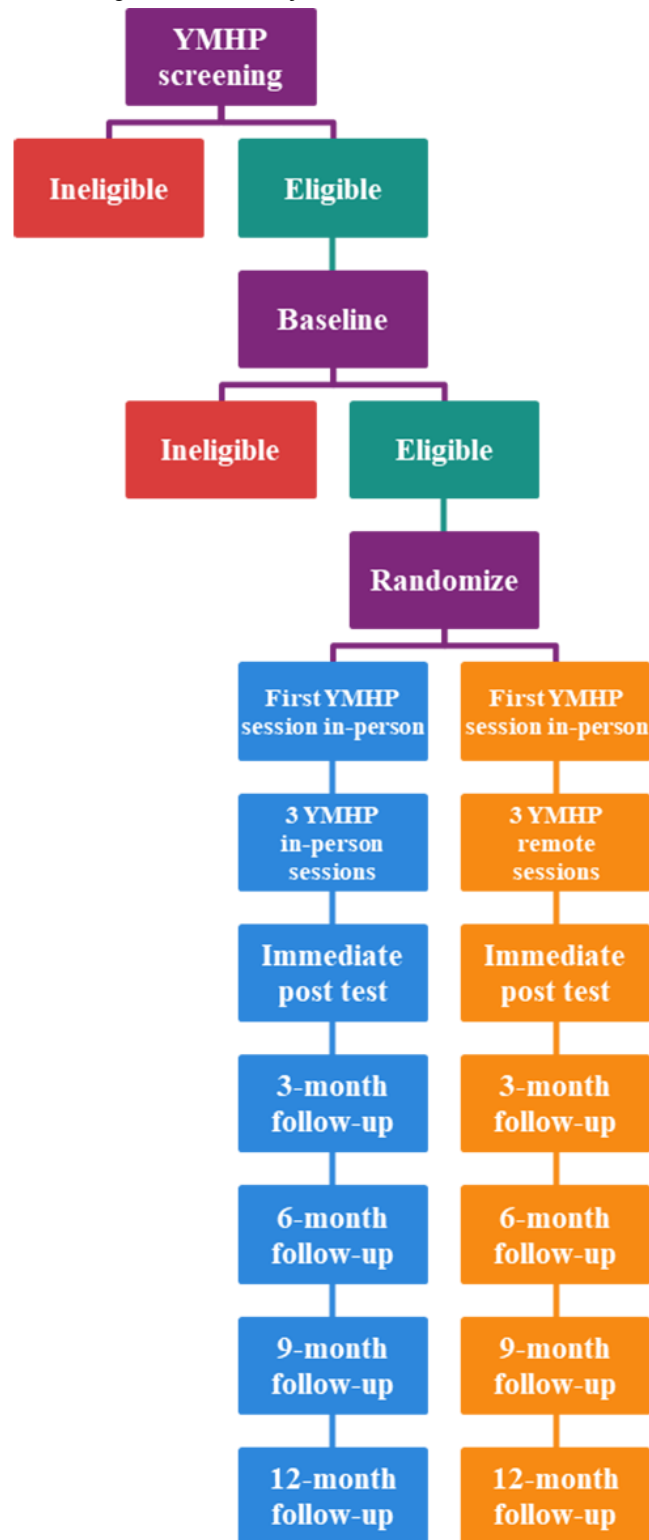**Figure 1.** Participant flow diagram. YMHP: Young Men's Health Project.

Figure 2. Young Men's Health Project recruitment flyer.

In addition, CHEST will assist in referring potentially eligible participants to the YMHP study through existing Web-based recruitment efforts. CHEST utilizes the Hunter College IRB-approved online master screener (OMS) used to preliminarily screen individuals who are interested in participating in studies being conducted through CHEST. If an individual is preliminarily eligible for a study, the individual is asked to provide contact information to CHEST for follow-up. For the purposes of this study, the OMS will be used to refer potentially eligible YMSM to HIV testing sites by sending them an email referral informing them about the YMHP study and where to go to determine eligibility after completing the OMS. Potential participants will also be called by CHEST staff to complete the *Study Screener* over the phone. If they meet the study criterion, they will be scheduled by the study staff to attend the clinic for HIV testing and YMHP enrollment. The OMS, in this instance, will primarily be used as a referral mechanism for the study, directing participants to which study they might be eligible for, including YMHP.

Eligibility Criteria

All interested participants are assessed for eligibility by completing a brief *Study Screener*. Study inclusion criteria include (1) being aged between 15 and 24 years; (2) currently identifying as male (regardless of birth sex); (3) receiving an HIV-negative test result from a study site or mobile HIV testing

in the past 90 days; (4) having sex with men in the past 90 days; (5) reporting at least 3 days of illicit drug use or heavy drinking (5 or more drinks) in the past 90 days; (6) reporting at least one episode of CAS with a male partner in the past 90 days or a positive STI test result in the past 90 days; (7) living in the Detroit, Miami, or Philadelphia metropolitan areas; and (8) having the ability to communicate in English. Exclusionary criteria include the following: (1) participants whose mental, physical, or emotional capacity does not permit them to complete the protocol as written; (2) currently taking Truvada as pre-exposure prophylaxis (PrEP); and (3) 5 or more days of injection drug use in the past 90 days.

Management and Tracking Study Visits

There are 2 different types of visits during the study timeline. Research visits include all assessments including the baseline visit, the posttest assessment, and all other follow-up assessments. All participants, in either condition, are expected to complete 6 research visits as part of full participation in the study. Research visits will be tracked using REDCap, a secure Web app for managing Web-based surveys and databases. This system allows both SRV study staff and the REC staff to monitor the completion of study visits and surveys and generate reports on enrollment and retention as needed. SRV study staff will track all completed study components in REDCap immediately after completing an assessment no later than the end of the

business day. Intervention visits include all YMHP sessions between enrolled study participants and trained CHWs. Intervention visits will be largely managed by the CHWs with their individual study participants. CHWs will notify SRV study staff of whether visits were completed for tracking purposes, provide receipts for session compensation to SRV study staff, and upload audio files to Dropbox Business as per study procedures. CHWs will deliver the first YHMP session at the completion of each baseline assessments immediately, regardless of randomization, and will schedule the subsequent session in person or remotely.

Intervention Design

YMHP has been listed as a best evidence intervention by the CDC and is the only intervention to show significant effects on both CAS and substance use among YMSM. MI techniques are utilized to provide personalized feedback for reducing CAS and substance use among YMSM. The YMHP intervention will be delivered by MI-trained CHWs employed at SRVs, primarily by former HIV counseling and testing counselors, health educators, and trained program peers. Participants will be randomized to receive the intervention either in person or remotely by phone or via video chat using Skype or FaceTime. Both conditions involve completion of the 4 YMHP sessions and the delivery of PrEP information and navigation services to interested participants.

Four Sessions of Young Men's Health Project

In session 1, youth will choose which behavior to discuss first (sexual risk or substance use), and the CHW will elicit the participant's view of the problem using standard MI techniques, building motivation for change by eliciting and reinforcing change talk and clarifying the youth's own personal priorities (through a structured card sorting activity). The CHW will assess and reflect the participant's readiness to make changes in target behavior. If the participant is willing to proceed, goals on healthy behaviors (ie, substance moderation, sex risk harm reduction strategies) are set. The session ends with MI strategies to evoke the youth's ideas about how to take steps toward change, consolidate the youth's commitment to the plan, and problem solving. Session 2 follows the same format as session 1 but revolves around the second target behavior. Session 3 includes a discussion about how PrEP might fit within a youth's goal for healthy behaviors. In sessions 3 and 4, the CHW will review the change plan, continue to elicit and reinforce change talk, problem-solve barriers, consolidate commitment, and address maintenance of behavior change.

Training of Interventionists

YMHP intervention training will occur before the initiation of phase 2, with ongoing coaching and supervision and training of new interventionists, as required. The interventionist training team includes 2 members of the Motivational Interviewing Network of Trainers from the CHEST, and this training procedure includes (1) an initial 3-day training for CHWs and local supervisors; (2) a 2- to 3-month training period of role-play practice, coding and feedback, and supervision modeling, including mock sessions with *standardized participants* role played by CHEST research assistants; (3) 1-hour weekly

supervision sessions between local supervisors and CHWs; (4) monthly supervision calls between local supervisors, the interventionist training team, and the protocol lead including a quarterly Skype booster training; and (5) ongoing quality assurance and feedback using MITI coding.

All materials (eg, slides, training exercises, supervisory tools) will be packaged for potential dissemination. Before dissemination, any copyrighted media will be removed from these materials. The initial 3-day training was held for all CHWs and their supervisors in Miami and followed a curriculum developed from previous NIH-funded effectiveness trials. CHWs and supervisors participated together in days 1 and 2 of the training. The third day of training was split so CHWs could have more practice with the YMHP protocol and supervisors could focus on coaching MI. CHEST provided external MITI coding for the supervisors to use as feedback. Following the 3-day training workshop, all CHWs and supervisors submitted audio recordings of all intervention sessions. Mock sessions are completed as in-person and remote delivery with mock participants. These sessions were MITI coded and CHEST trainers provided coaching and feedback. Once beginner competency was met, the local SRV supervisor took over weekly individual supervision of the CHWs. Throughout YMHP, the interventionist training team will continue to provide support to supervisors for assistance in supervision and will focus on having them practice listening for skills and then model for them how to use MITI feedback.

Fidelity Monitoring

All sessions (clinic-based and remote-based) will be audio-recorded, and one recording per CHW will be randomly selected for MITI coding by the research team on a regular basis. For the full trial, a random selection of 10% of the interviews will be independently coded. Supervisors will complete fidelity checklists for the supervision session so the team can monitor implementation. The protocol lead and the interventionist training team will facilitate quarterly boosters via group Skype for supervisors. Before the quarterly boosters, supervisors will submit a recording of a supervision session for review. Boosters will cover successes and challenges, MITI scores, updated MI skill development plans for each CHW, and role-plays of supervision skills. The protocol lead and the interventionist training team will join supervision sessions via Skype if MITI scores fall below competency without remediation. They will also lead annual in-person booster trainings covering MI skills and specific delivery of YMHP for supervisors and CHWs. All boosters will be recorded and qualitatively analyzed.

Results

Phase 1

Phase 1 was conducted in 2017, with site visits to Miami, Detroit, and Philadelphia where focus groups were conducted with youth. Focus groups participation breakdown is provided in Table 1. The feedback from the focus groups at each clinic has been used to modify the YMHP intervention before the launch of phase 2. Phase 1 focus groups were conducted with a total of 25 youth across the 3 SRVs to gather information that

would be used to better implement YMHP. Youth were divided into 2 age groups, 15 to 17 years and 18 to 24 years. The first focus group was conducted in Detroit between June 7th and 9th of 2017 with 14 scheduled potential participants. Out of those, 7 participants provided consent and participated in the group discussion. The second focus group was conducted in Philadelphia between July 10th and 11th with 23 scheduled potential participants. Out of those, 8 participants provided consent and participated in the group discussion. The final group was conducted in Miami between July 12th and 14th with 20 scheduled potential participants. Out of those, 10 participants provided consent and participated in the group discussion. A total of 25 participants who reported their HIV status as negative attended focus groups and provided their feedback on the YMHP intervention.

Participants in focus groups were asked to provide feedback on HIV testing and counseling experiences and how to incorporate screening into the testing process, the YMHP intervention, and barriers to completing participation in the study. Participants reported positive experience working with the 3 sites and limited negative experiences overall with testing and counseling. Many participants expressed a lack of discussion about substance use but a desire to engage in discussions about it with CHWs, especially in the context of peer pressure from older partners. Participants at all SRVs reported cocaine and ecstasy as commonly used substances in their cities. When participants were asked about the intervention, they expressed interest in the remote delivery option. Many thought that the advantages to sessions over the phone or video chat such as Skype and FaceTime are that it eliminates transportation as a barrier to session completion and the resistance of talking to a therapist face-to-face. Participants also felt that individual characteristics (eg, race, gender) of the CHW delivering the YMHP intervention did not matter as long as they were confident with their knowledge and the resources they are offering. In addition, they perceived scheduling and discretion were the 2 biggest barriers to completing sessions because of school and needing parents to possibly transport them to and from appointments at the SRV, while also having to explain the study to strict parents for youth less than or equal to 18 years as an obstacle.

Table 1. Young Men's Health Project focus groups.

Site	Date	Participants	
		Age group 15-17 years (n=3)	Age group 18-24 years (n=22)
Detroit	June 7-9, 2017	0	7
Philadelphia	July 10-11, 2017	2	6
Miami	July 12-14, 2017	1	9

Phase 2

On the basis of the results from phase 1, for phase 2 we will enroll 270 YMSM, with 90 participants per site and 135 participants in each of the 2 conditions. Recruitment for phase 2 began in October 2018, and all participant components are projected to end in December 2020.

Intervention Outcomes and Measures

This study examines 4 coprimary outcomes related to substance use (1 outcome) and sexual health management (3 outcomes). These are measured during baseline; immediate postintervention; and 3-, 6-, 9-, and 12-months follow-up assessments. Specifically, sexual health management is measured by 3 outcomes: (1) decreased STIs, (2) decreased CAS, and (3) increased PrEP uptake/adherence. HIV/STI testing occurs after immediate postintervention at 3- and 9-month follow-up assessments.

Substance use is measured using the *Alcohol, Smoking and Substance Involvement Screening Test* (ASSIST) [60]. ASSIST is often utilized as a screener in primary care settings for substance abuse. Specifically, ASSIST assesses participants' use of alcohol, tobacco, cannabis, sedative, hallucinogens, inhalants, opioids, and other drugs. The primary substance use outcome will be substance use days during the past 90 days.

To measure CAS, participants complete a series of questions pertaining to their sexual behavior with main and casual partners [61]. Participants estimate their total number of sex partners in the past 3 months. Participants also indicate if they had unprotected, receptive or insertive, anal sex with partners who were HIV positive or of unknown HIV status.

The *Motivational PrEP Cascade* is a series of 21 questions that are designed to assess PrEP treatment uptake and adherence [62-64]. Participants report their familiarity with PrEP; experiences and acceptability; as well as PrEP contemplation, preparation, action, and maintenance utilizing the Transtheoretical Model of Change framework. This includes assessing willingness and intentions for PrEP uptake. The *Motivational PrEP Cascade* also assesses when participants begin and stop taking PrEP [65].

Putative Moderators and Mediators of Intervention Effects

Moderators and mediators of intervention effects will be assessed with health care access and other self-management constructs for decision making, problem solving, self-regulation, and provider and health care relationship. Health care access is a series of 9 questions adapted from Williams and Chapman's unmet health and mental health need [66]. Items included whether youth were not able to get health care services (ie, routine physical examination, see a provider for sexual health care or get access to PrEP, psychological or emotional counseling, and counseling for drug or alcohol use) in the past year when they felt they needed those services.

Self-management constructs included the Behavior Rating Inventory of Executive Function (BRIEF-A) [67] and the Patient

Activation Measure [68]. BRIEF-A is a series of 22 questions asking patients to rate whether certain problems (eg, bothered by having to deal with changes, do not plan ahead for tasks, and problems completing my work) occurred in the past month. Responses included never, sometimes, and often. For patient activation measure, participants were asked on a 4-point scale, ranging from strongly disagree to strongly agree or not applicable, a series of 10 questions related to their health. Examples of some of the questions include, "When all is said and done, I am the person who is responsible for taking care of my health" and "I am confident that I can follow through on medical treatments I may need to do at home."

Quantitative Analysis Plan

The primary hypothesis is that, although the main effect of YMHP delivery modality will be nonsignificant, there will be a significant interaction between access to health care and delivery modality. Specifically, it is hypothesized that remote-based YMHP will demonstrate greater improvements in sexual health management (as measured by decreased STIs, CAS, and increased PrEP uptake/adherence) as well as greater reductions in substance use, compared with clinic-based YMHP, among YMSM who do report barriers to health care access. In contrast, it is hypothesized that clinic-based YMHP will demonstrate greater improvements in sexual health management and reduced substance use among YMSM who do not report barriers to health care access.

The protocol lead and analytic team will test the effect of clinic-based versus remote-based delivery on STI rates, alcohol and drug use behavior, CAS, and PrEP uptake through multilevel growth mixture modeling (GMM). A separate model will be run for each outcome. Each model will be a 2-level model in which individuals (level I) are nested in clinics (level II). This approach controls for the nonindependence of individuals within clinics. As 3 sites provide extremely limited predictive power at level II, no site covariates are included in the model. The models are empty at level II.

In a GMM, a latent growth curve with an intercept and linear slope factor is specified. Latent class analysis is applied to these 2 growth components (intercept and slope) to identify groups of individuals who share trajectories. For example, with regard to drug or alcohol use, *immediate and sustained responders* might have the lowest postintervention intercept and a flat slope. Meanwhile, *non-responders* might have the highest postintervention intercept and a flat slope. In contrast, *delayed responders* might have a high postintervention intercept but a significant negative slope, indicating reductions in missed medication over the follow-up period.

If modeling results indicate that discrete classes are not present, we will proceed with analyses in which the growth factors (intercept and slope) are predicted directly by intervention condition (and demographic factors found to be associated with condition after randomization or with attrition over follow-up). GMMs can subsequently incorporate predictors of class membership. These analyses can be conceptualized as a multinominal logistic regression with the latent trajectory-class membership constituting the outcome. The predictors of primary interest will be intervention condition, the presence of any

barriers to health care access, and the interaction between condition and barriers to access. We will include as covariates any demographic variables that were associated with condition after randomization or with attrition over the follow-up period.

Power Analysis and Sample Size

Power analysis was conducted based on the 4 outcomes of STIs, PrEP uptake, alcohol and drug use, and CAS. First, we analyzed power assuming independence of participant observations (assuming that the nesting of people within clinic was irrelevant). With regard to STIs, we utilized the *repeated measures* module of PASS 13.0 [NCSS Statistical Software] to examine power to detect odds ratio differences in a repeated measures design. Specifying compound symmetry, we allowed ρ to vary between .2 and .5. Assuming the prevalence of STIs in remote delivery condition varies between .05 and .15, the proposed design (N=270) has power .80 to detect an odds ratio of 0.20 to 0.50. Similarly, with regard to the odds of PrEP uptake, allowing the rate of uptake in the remote delivery condition to vary between .05 and .20 and allowing ρ to vary between .4 and .7, the proposed design has power of .80 to detect an odds ratio of 1.9 to 2.3. With regard to number of alcohol and drug use days, we utilized the Tests for Two Poisson Means module in PASS 13.0. On the basis of the data from our previous studies, we allowed the rate of heavy drinking in the remote delivery condition to vary between 7 and 9 days during a 30-day assessment. The study is adequately powered to detect a 3% reduction in the number of heavy drinking days in the in-person condition at any single follow-up point. Allowing the rate of substance use in the remote delivery condition to vary between 3 and 5 days during a 30-day assessment period, the study is adequately powered to detect a 4% reduction in drug use in the in-person condition. A similar analysis was conducted with respect to CAS. On the basis of our previous research, we allowed the rate of CAS in the remote delivery condition to vary between 2 and 4 acts in a 30-day assessment period. The study is powered to detect a difference as small as 5% between the remote delivery and in-person YMHP conditions. The nesting of individuals within sites has the potential to reduce power because substantial variability in outcome across sites can obscure level-II treatment effects [69,70]. The design effect can be used to tailor power analyses calculated under assumptions of independence. In the case of a level-I predictor with a fixed effect that is uncorrelated with other covariates in the model, the design effect is equal to the 1–ρ, where ρ is the intraclass correlation or the percentage of variance accounted for by variability between sites. In previous ATN intervention trials, between clinic site variability in HIV-related outcomes did not differ significantly from 0. We anticipate an absence of variability across clinics, suggesting that the design effect would result in a negligible reduction in power. Finally, a sample size of 270 is sufficient to detect a moderation effect with an f^2 of 0.02. Cohen [71] designates this as a small effect; however, recent work has characterized an effect of this size as moderate to large as applied to moderation [72].

Cost-Effectiveness

To enhance the likelihood of uptake if effective, the cost-effectiveness of 2 delivery models of YMHP in reducing

sexual risk and substance use will be assessed utilizing CDC's guidelines for cost-effectiveness analysis on HIV infections averted. The economic analysis will have 2 components: (1) a cost analysis of the YMHP intervention and (2) an incremental cost-effectiveness analysis that compares the value of clinic-delivery of YMHP over remote delivery. We will first estimate the marginal costs of delivering the 2 formats of YMHP. Using data from the modified Drug Abuse Treatment Cost Analysis Program [73,74] and study contact and expenditure records, key statistics from the cost evaluation will include the total annual economic cost for each program, weekly economic cost per client, and total economic cost per intervention session [75-77]. To highlight the relative contribution of the various cost components and necessary future budgeting, we will also perform a descriptive analysis of the cost accounted for by resource category.

The mean aggregate cost of the interventions will be used as inputs in the cost-effectiveness model. Cost-effectiveness will be modeled for clinic-delivery of YMHP as compared with remote-delivery for the differences in CAS and predicted through Markov modeling for 5 and 10 years and over a lifetime using varying assumptions about decay of the effect of the intervention over time. Modeling will be performed from the perspectives of (1) a third-party payer, (2) the medical care system, and (3) society.

Intervention Effects on Self-Management and Tests of Putative Mediation

Using GMM models similar to those described above in primary outcome analyses, we will examine the cross-time effects of intervention on the 5 dimensions of self-management. Where a significant between-condition difference in self-management is detected, we will explore mediation by examining whether intercept and/or growth factors for that dimension of self-management in turn predict outcomes. Indirect effects in GMM will be evaluated using bootstrapping estimation where possible. When this is not possible, a constraint approach will be employed. This approach involves comparing the fit of 2 models: one model in which the product of constituent pathways is constrained to be 0 and another in which the product of the direct effects is unconstrained. A significant Chi-square test associated with this comparison indicates that the constraint significantly diminishes model fit and constitutes evidence of the significance of the indirect pathway [78].

Discussion

Principal Findings

The goals of this YMHP intervention are to better understand HIV prevention–focused self-management behaviors among HIV-negative YMSM and to study the implementation of YMHP to improve portability and scalability. The SRVs will help to assess and address practical problems at the frontline of service provision to pave the way for a comprehensive program to reduce HIV infection among YMSM that reflects the complexities of real-world adolescent HIV clinics. If proven successful, this intervention delivery could help YMSM across the United States.

On the basis of previous studies, YMSM are at an even higher risk for HIV and STIs than older MSM [3,23,25]. YMSM living in urban areas and YMSM of color are especially at risk [6]. The need to lower these rates makes this study important. Similar interventions have been effective in the past. In a study of youth currently living with HIV, results showed that participants who had attended MI sessions were more likely to pursue behavior changes compared with participants in the control condition. The more sessions a participant attended, the better the results [5]. MI was also beneficial in a study of field outreach for young black men for HIV counseling and testing. The study found that outreach workers who had implemented MI were more likely to encourage youth to learn about their HIV status.

Implementing YMHP MI-based interventions targeting MSM at risk has been shown to be effective in past studies. This study is beneficial because it addresses the needs of the YMSM population. In urban areas, access to care (information about care or transportation means), PrEP information, and knowledge of sexual health and substance use might not occur in schools or with primary care providers; introducing all of this in a specially tailored intervention will give youth the information to lower their risk of HIV infection and STIs. It is important to implement this strategy drawing on the success of YMHP interventions in the past [79,80] because of the fact that YMSM as the population continues to see a rise in new HIV infection (132%), whereas rates among other groups have remained stable [6].

To address concerns related to differential drop out in that more remote delivery participants might never receive first session and that it is quite possible that retention across the 4 sessions is still higher in the remote condition because of transportation issues for in-person delivery, everyone will be asked to complete their session 1 in person regardless of their randomization to the intervention conditions. This would also address concerns some of the focus group youth had about developing a rapport over the phone or video chat. We estimate that retention for the IP and the 3-, 6-, 9-, and 12-month assessments will be 97%, 94%, 91%, 88%, and 85%, respectively, based on our previous study on YMHP [80,81]. Through REDCap, each clinic site will be able to generate reports for when YMSM are due for follow-up assessments, and the lead site, CHEST Hunter College, will provide extensive training to research assistants on retention efforts.

In addition, as YMHP is being tested in conjunction with other evidence-based practices in the Scale It Up program of research [58], intervention applicability and affordability will be determined through implementation science research methods and cost-effectiveness analysis. These components will be studied using EPIS to generate knowledge about the barriers and facilitators to the implementation and sustainment of the intervention into adolescent HIV prevention and clinical care settings.

Limitations

This study has several possible limitations based on the population and locations involved. Another limitation regarding eligibility criteria is self-reporting of 3 or more days of substance

use and 1 episode of CAS in the past 90 days. However, with the 3 SRVs HIV epicenters, intervention would be beneficial even as a preventative measure for this population. Although this study has a waiver of parental consent, parents or guardians might also come up as a barrier to participation. If parents do not approve of participation, they might discourage potential participants from participating or enrolled participants from completing the full study by refusing to cooperate if

transportation to and from the clinic is dependent on the parent. Transportation overall might limit which participants can make it to appointments when they have to go to SRVs for testing, such as in Detroit where the lack of efficient public transportation is a barrier to participation in intervention studies. This is one of the main reasons why remote-based delivery is being tested in these real-world adolescent clinics.

Acknowledgments

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. Atlanta, GA; 2015 Feb. HIV Surveillance Report, 2013 URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-report-2013-vol-25.pdf [accessed 2019-03-19] [WebCite Cache ID 770EAqL7I]

2. Centers for Disease Control and Prevention. Atlanta, GA HIV surveillance in men who have sex with men (MSM), 2012 URL: https://www.cdc.gov/healthyyouth/sexualbehaviors/pdf/hiv_factsheet_ymsm.pdf [accessed 2019-03-19] [WebCite Cache ID 770EPhqwX]

3. Centers for Disease Control and Prevention. Atlanta, GA; 2014. Sexually Transmitted Disease Surveillance, 2014 URL: https://www.cdc.gov/std/stats/ [accessed 2019-03-19] [WebCite Cache ID 770MEjZe1]

4. Pathela P, Braunstein SL, Blank S, Schillinger JA. HIV incidence among men with and those without sexually transmitted rectal infections: estimates from matching against an HIV case registry. Clin Infect Dis 2013 Oct;57(8):1203-1209. [doi: 10.1093/cid/cit437] [Medline: 23800942]

5. Corliss HL, Rosario M, Wypij D, Wylie SA, Frazier AL, Austin SB. Sexual orientation and drug use in a longitudinal cohort study of US adolescents. Addict Behav 2010 May;35(5):517-521 [FREE Full text] [doi: 10.1016/j.addbeh.2009.12.019] [Medline: 20061091]

6. Parsons JT, Grov C, Kelly BC. Club drug use and dependence among young adults recruited through time-space sampling. Public Health Rep 2009 Mar;124(2):246-254 [FREE Full text] [doi: 10.1177/003335490912400212] [Medline: 19320366]

7. Parsons JT, Halkitis PN, Bimbi DS. Club drug use among young adults frequenting dance clubs and other social venues in New York City. J Child Adolesc Subst Abuse 2006 Feb 27;15(3):1-14. [doi: 10.1300/J029v15n03_01]

8. Parsons JT, Kutnick AH, Halkitis PN, Punzalan JC, Carbonari JP. Sexual risk behaviors and substance use among alcohol abusing HIV-positive men who have sex with men. J Psychoactive Drugs 2005 Mar;37(1):27-36 [FREE Full text] [doi: 10.1080/02791072.2005.10399746] [Medline: 15916249]

9. Purcell DW, Moss S, Remien RH, Woods WJ, Parsons JT. Illicit substance use, sexual risk, and HIV-positive gay and bisexual men: differences by serostatus of casual partners. AIDS 2005 Apr;19(Suppl 1):S37-S47. [doi: 10.1097/01.aids.0000167350.00503.db] [Medline: 15838193]

10. Purcell DW, Parsons JT, Halkitis PN, Mizuno Y, Woods WJ. Substance use and sexual transmission risk behavior of HIV-positive men who have sex with men. J Subst Abuse 2001;13(1-2):185-200. [Medline: 11547619]

11. Ostrow DG, Plankey MW, Cox C, Li X, Shoptaw S, Jacobson LP, et al. Specific sex drug combinations contribute to the majority of recent HIV seroconversions among MSM in the MACS. J Acquir Immune Defic Syndr 2009 Jul 1;51(3):349-355 [FREE Full text] [doi: 10.1097/QAI.0b013e3181a24b20] [Medline: 19387357]

12. Peterson JL, Coates TJ, Catania JA, Middleton L, Hilliard B, Hearst N. High-risk sexual behavior and condom use among gay and bisexual African-American men. Am J Public Health 1992 Nov;82(11):1490-1494. [doi: 10.2105/AJPH.82.11.1490] [Medline: 1443298]

13. Edlin BR, Irwin KL, Faruque S, McCoy CB, Word C, Serrano Y, et al. Intersecting epidemics--crack cocaine use and HIV infection among inner-city young adults. Multicenter Crack Cocaine and HIV Infection Study Team. N Engl J Med 1994 Nov 24;331(21):1422-1427. [doi: 10.1056/NEJM199411243312106] [Medline: 7969281]

XSL•FO
RenderX

14. Marx R, Aral SO, Rolfs RT, Sterk CE, Kahn JG. Crack, sex, and STD. Sex Transm Dis 1991;18(2):92-101. [doi: 10.1097/00007435-199118020-00008] [Medline: 1862466]

15. Shoptaw S, Reback CJ, Frosch DL, Rawson RA. Stimulant abuse treatment as HIV prevention. J Addict Dis 1998;17(4):19-32. [doi: 10.1300/J069v17n04_03] [Medline: 9848029]

16. Word CO, Bowser B. Background to crack cocaine addiction and HIV high-risk behavior: the next epidemic. Am J Drug Alcohol Abuse 1997 Feb;23(1):67-77. [doi: 10.3109/00952999709001688] [Medline: 9048148]

17. Mayer KH, Wang L, Koblin B, Mannheimer S, Magnus M, del RC, et al. Concomitant socioeconomic, behavioral, and biological factors associated with the disproportionate HIV infection burden among black men who have sex with men in 6 US cities. PLoS One 2014;9(1):e87298 [FREE Full text] [doi: 10.1371/journal.pone.0087298] [Medline: 24498067]

18. Centers for Disease Control and Prevention. Atlanta, GA; 2013. HIV substance use in the United States URL: https://www.cdc.gov/hiv/risk/substanceuse.html [WebCite Cache ID 6ypNIZypu]

19. Plankey MW, Ostrow DG, Stall R, Cox C, Li X, Peck JA, et al. The relationship between methamphetamine and popper use and risk of HIV seroconversion in the multicenter AIDS cohort study. J Acquir Immune Defic Syndr 2007 May 1;45(1):85-92 [FREE Full text] [doi: 10.1097/QAI.0b013e3180417c99] [Medline: 17325605]

20. Parsons JT, Lelutiu-Weinberger C, Botsko M, Golub SA. Predictors of day-level sexual risk for young gay and bisexual men. AIDS Behav 2013 May;17(4):1465-1477 [FREE Full text] [doi: 10.1007/s10461-012-0206-1] [Medline: 22614745]

21. Garofalo R, Mustanski B, Johnson A, Emerson E. Exploring factors that underlie racial/ethnic disparities in HIV risk among young men who have sex with men. J Urban Health 2010 Mar;87(2):318-323 [FREE Full text] [doi: 10.1007/s11524-009-9430-z] [Medline: 20155329]

22. Newcomb ME, Ryan DT, Greene GJ, Garofalo R, Mustanski B. Prevalence and patterns of smoking, alcohol use, and illicit drug use in young men who have sex with men. Drug Alcohol Depend 2014 Aug 1;141:65-71 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.05.005] [Medline: 24907774]

23. Centers for Disease Control and Prevention (CDC). Vital signs: HIV infection, testing, and risk behaviors among youths - United States. MMWR Morb Mortal Wkly Rep 2012 Nov 30;61(47):971-976 [FREE Full text] [doi: 10.1111/ajt.12172] [Medline: 23190571]

24. Hettema J, Steele J, Miller WR. Motivational interviewing. Annu Rev Clin Psychol 2005;1:91-111. [doi: 10.1146/annurev.clinpsy.1.102803.143833] [Medline: 17716083]

25. Lundahl BW, Kunz C, Brownell C, Tollefson D, Burke BL. A meta-analysis of motivational interviewing: twenty-five years of empirical studies. Res Soc Work Pract 2010 Jan 11;20(2):137-160. [doi: 10.1177/1049731509347850]

26. Outlaw AY, Naar-King S, Parsons JT, Green-Jones M, Janisse H, Secord E. Using motivational interviewing in HIV field outreach with young African American men who have sex with men: a randomized clinical trial. Am J Public Health 2010 Apr 1;100(Suppl 1):S146-S151. [doi: 10.2105/AJPH.2009.166991] [Medline: 20147689]

27. Miller W, Rollnick S. Motivational Interviewing: Helping People Change, 3rd Edition (Applications of Motivational Interviewing). New York, NY: The Guilford Press; 2012.

28. Parsons JT, Lelutiu-Weinberger C, Botsko M, Golub SA. A randomized controlled trial utilizing motivational interviewing to reduce HIV risk and drug use in young gay and bisexual men. J Consult Clin Psychol 2014 Feb;82(1):9-18 [FREE Full text] [doi: 10.1037/a0035311] [Medline: 24364800]

29. Larsen MH, Krogstad AL, Aas E, Moum T, Wahl AK. A telephone-based motivational interviewing intervention has positive effects on psoriasis severity and self-management: a randomized controlled trial. Br J Dermatol 2014 Dec;171(6):1458-1469. [doi: 10.1111/bjd.13363] [Medline: 25143061]

30. Ream E, Gargaro G, Barsevick A, Richardson A. Management of cancer-related fatigue during chemotherapy through telephone motivational interviewing: modeling and randomized exploratory trial. Patient Educ Couns 2015 Feb;98(2):199-206. [doi: 10.1016/j.pec.2014.10.012] [Medline: 25468398]

31. Stevens J, Hayes J, Pakalnis A. A randomized trial of telephone-based motivational interviewing for adolescent chronic headache with medication overuse. Cephalalgia 2014 May;34(6):446-454. [doi: 10.1177/0333102413515336] [Medline: 24322483]

32. Lovejoy TI. Telephone-delivered motivational interviewing targeting sexual risk behavior reduces depression, anxiety, and stress in HIV-positive older adults. Ann Behav Med 2012 Dec;44(3):416-421. [doi: 10.1007/s12160-012-9401-6] [Medline: 22956397]

33. Borsari B, Short EE, Mastroleo NR, Hustad JTP, Tevyaw TO, Barnett NP, et al. Phone-delivered brief motivational interventions for mandated college students delivered during the summer months. J Subst Abuse Treat 2014 May;46(5):592-596 [FREE Full text] [doi: 10.1016/j.jsat.2014.01.001] [Medline: 24512944]

34. Clement S, van Nieuwenhuizen A, Kassam A, Flach C, Lazarus A, de Castro M, et al. Filmed v live social contact interventions to reduce stigma: randomised controlled trial. Br J Psychiatry 2012 Jul;201(1):57-64 [FREE Full text] [doi: 10.1192/bjp.bp.111.093120] [Medline: 22157800]

35. Patten CA, Croghan IT, Meis TM, Decker PA, Pingree S, Colligan RC, et al. Randomized clinical trial of an Internet-based versus brief office intervention for adolescent smoking cessation. Patient Educ Couns 2006 Dec;64(1-3):249-258. [doi: 10.1016/j.pec.2006.03.001] [Medline: 16616449]

36.   Paxton SJ, McLean SA, Gollings EK, Faulkner C, Wertheim EH. Comparison of face-to-face and internet interventions for body image and eating problems in adult women: an RCT. Int J Eat Disord 2007 Dec;40(8):692-704. [doi: 10.1002/eat.20446] [Medline: 17702020]

37.   Steele R, Mummery WK, Dwyer T. Using the internet to promote physical activity: a randomized trial of intervention delivery modes. J Phys Act Health 2007 Jul;4(3):245-260. [doi: 10.1123/jpah.4.3.245] [Medline: 17846455]

38.   Carey KB, Carey MP, Henson JM, Maisto SA, DeMartini KS. Brief alcohol interventions for mandated college students: comparison of face-to-face counseling and computer-delivered interventions. Addiction 2011 Mar;106(3):528-537 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.03193.x] [Medline: 21059184]

39.   Sullivan P, Goldmann D. The promise of comparative effectiveness research. J Am Med Assoc 2011 Jan 26;305(4):400-401. [doi: 10.1001/jama.2011.12] [Medline: 21266687]

40.   Naar-King S, Outlaw A, Green-Jones M, Wright K, Parsons JT. Motivational interviewing by peer outreach workers: a pilot randomized clinical trial to retain adolescents and young adults in HIV care. AIDS Care 2009 Jul;21(7):868-873. [doi: 10.1080/09540120802612824] [Medline: 20024744]

41.   Lauer PA, Christopher DE, Firpo-Triplett R, Buchting F. The impact of short-term professional development on participant outcomes: a review of the literature. Professional Dev Educ 2013 Apr 10;40(2):207-227. [doi: 10.1080/19415257.2013.776619]

42.   Lewin SA, Dick J, Pond P, Zwarenstein M, Aja G, van Wyk B, et al. Lay health workers in primary and community health care. Cochrane Database Syst Rev 2005 Jan 25(1):CD004015. [doi: 10.1002/14651858.CD004015.pub2] [Medline: 15674924]

43.   Conway PH, Clancy C. Comparative-effectiveness research--implications of the Federal Coordinating Council's report. N Engl J Med 2009 Jul 23;361(4):328-330. [doi: 10.1056/NEJMp0905631] [Medline: 19567829]

44.   Federal Coordinating Council for Comparative Effectiveness Research. Office of Science Policy. Washington, DC: US Department of Health and Human Services; 2009 Jun 30. Report to the President and Congress URL: https://osp.od.nih.gov/wp-content/uploads/FCCCER-Report-to-the-President-and-Congress-2009.pdf [accessed 2019-03-19] [WebCite Cache ID 770FBKWJg]

45.   Iglehart JK. Prioritizing comparative-effectiveness research--IOM recommendations. N Engl J Med 2009 Jul 23;361(4):325-328. [doi: 10.1056/NEJMp0904133] [Medline: 19567828]

46.   Institute of Medicine. Initial National Priorities for Comparative Effectiveness Research. Washington, DC: The National Academies Press; 2009.

47.   Luce BR, Kramer JM, Goodman SN, Connor JT, Tunis S, Whicher D, et al. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. Ann Intern Med 2009 Aug 4;151(3):206-209. [doi: 10.7326/0003-4819-151-3-200908040-00126] [Medline: 19567619]

48.   Proeschold-Bell RJ, Heine A, Pence BW, McAdam K, Quinlivan EB. A cross-site, comparative effectiveness study of an integrated HIV and substance use treatment program. AIDS Patient Care STDS 2010 Oct;24(10):651-658. [doi: 10.1089/apc.2010.0073] [Medline: 20846009]

49.   Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. Ann Intern Med 2009 Aug 4;151(3):203-205. [doi: 10.7326/0003-4819-151-3-200908040-00125] [Medline: 19567618]

50.   Volpp KG, Das A. Comparative effectiveness--thinking beyond medication A versus medication B. N Engl J Med 2009 Jul 23;361(4):331-333. [doi: 10.1056/NEJMp0903496] [Medline: 19625714]

51.   Mukherjee JS, Eustache FE. Community health workers as a cornerstone for integrating HIV and primary healthcare. AIDS Care 2007 Mar;19(Suppl 1):S73-S82. [doi: 10.1080/09540120601114485] [Medline: 17364390]

52.   Miller WR, Yahne CE, Moyers TB, Martinez J, Pirritano M. A randomized trial of methods to help clinicians learn motivational interviewing. J Consult Clin Psychol 2004 Dec;72(6):1050-1062. [doi: 10.1037/0022-006X.72.6.1050] [Medline: 15612851]

53.   Mitcheson L, Bhavsar K, McCambridge J. Randomized trial of training and supervision in motivational interviewing with adolescent drug treatment practitioners. J Subst Abuse Treat 2009 Jul;37(1):73-78. [doi: 10.1016/j.jsat.2008.11.001] [Medline: 19150203]

54.   Moyers TB, Manuel JK, Wilson PG, Hendrickson SM, Talcott W, Durand P. A randomized trial investigating training in motivational interviewing for behavioral health providers. Behav Cognit Psychother 2007 Nov 22;36(02). [doi: 10.1017/S1352465807004055]

55.   Söderlund LL, Madson MB, Rubak S, Nilsen P. A systematic review of motivational interviewing training for general health care practitioners. Patient Educ Couns 2011 Jul;84(1):16-26. [doi: 10.1016/j.pec.2010.06.025] [Medline: 20667432]

56.   Martino S, Brigham GS, Higgins C, Gallon S, Freese TE, Albright LM, et al. Partnerships and pathways of dissemination: the National Institute on Drug Abuse-Substance Abuse and Mental Health Services Administration Blending Initiative in the Clinical Trials Network. J Subst Abuse Treat 2010 Jun;38(Suppl 1):S31-S43 [FREE Full text] [doi: 10.1016/j.jsat.2009.12.013] [Medline: 20307793]

57.   Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]

58. Naar S, Parsons JT, Stanton BF. Adolescent Trials Network for HIV-AIDS Scale It Up Program: rational and overviews. JMIR Res Protoc 2019 Feb 1;8(2):e11204 [FREE Full text] [doi: 10.2196/11204] [Medline: 30707102]

59. Carcone AI, Coyle K, Gurung S, Cain D, Dilones RE, Jadwin-Cakmak L, et al. Effectiveness-Implementation Research to enhance HIV-related self-management among youth—Exploration-Preparation-Implementation-Sustainment (EPIS). JMIR Res Protoc 2019 (forthcoming). [doi: 10.2196/11202]

60. WHO ASSIST Working Group. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. Addiction 2002 Sep;97(9):1183-1194. [doi: 10.1046/j.1360-0443.2002.00185.x] [Medline: 12199834]

61. Grov C, Parsons JT, Bimbi DS. Sexual risk behavior and venues for meeting sex partners: an intercept survey of gay and bisexual men in LA and NYC. AIDS Behav 2007 Nov;11(6):915-926. [doi: 10.1007/s10461-006-9199-y] [Medline: 17206536]

62. Golub SA, Kowalczyk W, Weinberger CL, Parsons JT. Preexposure prophylaxis and predicted condom use among high-risk men who have sex with men. J Acquir Immune Defic Syndr 2010 Aug;54(5):548-555 [FREE Full text] [doi: 10.1097/QAI.0b013e3181e19a54] [Medline: 20512046]

63. Grov C, Whitfield THF, Rendina HJ, Ventuneac A, Parsons JT. Willingness to take PrEP and potential for risk compensation among highly sexually active gay and bisexual men. AIDS Behav 2015 Dec;19(12):2234-2244. [doi: 10.1007/s10461-015-1030-1] [Medline: 25735243]

64. Parsons JT, Rendina HJ, Lassiter JM, Whitfield THF, Starks TJ, Grov C. Uptake of HIV pre-exposure prophylaxis (PrEP) in a national cohort of gay and bisexual men in the United States. J Acquir Immune Defic Syndr 2017 Dec 1;74(3):285-292 [FREE Full text] [doi: 10.1097/QAI.0000000000001251] [Medline: 28187084]

65. McCumber M, Cain D, LeGrand S, Mayer K, Murphy D, Psioda M, et al. Adolescent Medicine Trials Network for HIV/AIDS interventions data harmonization: rationale and development of guidelines. JMIR Res Protoc 2018 Dec 21;7(12):e11207 [FREE Full text] [doi: 10.2196/11207] [Medline: 30578242]

66. Williams KA, Chapman MV. Unmet health and mental health need among adolescents: the roles of sexual minority status and child-parent connectedness. Am J Orthopsychiatry 2012 Oct;82(4):473-481. [doi: 10.1111/j.1939-0025.2012.01182.x] [Medline: 23039345]

67. Roth R, Isquith P, Gioia G. PAR-iConnect. Behavior Rating Inventory of Executive Function® Adult Version (BRIEF®-A) URL: https://www.parinc.com/Products/Pkey/25 [accessed 2019-03-19] [WebCite Cache ID 770Ev6w50]

68. Insignia Health. The Patient Activation Measure ® (PAM®). URL: https://www.insigniahealth.com/products/pam-survey [accessed 2019-03-19] [WebCite Cache ID 770EzbqzQ]

69. Snijders T. Power and sample size in multilevel modeling. In: Everitt BS, Howell DC, editors. Encyclopedia of Statistics in Behavioral Science, Volume 3. Chicester: Wiley; 2005:1570-1573.

70. Ling E, Cotter D. Statistical power in comparative aquaculture studies. Aquaculture 2003 Jun;224(1-4):159-168. [doi: 10.1016/S0044-8486(03)00225-4]

71. Cohen J. Statistical Power Analysis For The Behavioral Sciences, Second Edition. United States: Routledge; 1988.

72. Aguinis H, Beaty J, Boik R, Pierce C. Effect size and power in assessing moderating effects of categorical variables using multiple regression: a 30-year review. J Appl Psychol 2005 Jan;90(1):94-107. [doi: 10.1037/0021-9010.90.1.94] [Medline: 15641892]

73. French MT, Salomé HJ, Carney M. Using the DATCAP and ASI to estimate the costs and benefits of residential addiction treatment in the State of Washington. Soc Sci Med 2002 Dec;55(12):2267-2282. [doi: 10.1016/S0277-9536(02)00060-6] [Medline: 12409139]

74. French M, Salomé HJ, Sindelar J, McLellan A. Benefit-cost analysis of addiction treatment: methodological guidelines and empirical application using the DATCAP and ASI. Health Serv Res 2002 Apr;37(2):433-455 [FREE Full text] [doi: 10.1111/1475-6773.031] [Medline: 12036002]

75. Simpson KN. Economic modeling of HIV treatments. Curr Opin HIV AIDS 2010 May 25;5(3):242-248. [doi: 10.1097/COH.0b013e3283384aed] [Medline: 20539081]

76. Simpson KN, Baran RW, Kirbach SE, Dietz B. Economics of switching to second-line antiretroviral therapy with lopinavir/ritonavir in Africa: estimates based on DART trial results and costs for Uganda and Kenya. Value Health 2011 Dec;14(8):1048-1054 [FREE Full text] [doi: 10.1016/j.jval.2011.06.011] [Medline: 22152173]

77. Simpson KN, Roberts G, Hicks CB, Finnern HW. Cost-effectiveness of tipranavir in treatment-experienced HIV patients in the United States. HIV Clin Trials 2008 Jan;9(4):225-237. [doi: 10.1310/hct0904-225] [Medline: 18753117]

78. Muthén L, Muthén B. Mplus User's Guide Eight Edition. Los Angeles, CA: Muthén & Muthén; 2017.

79. Naar-King S, Parsons JT, Murphy D, Kolmodin K, Harris DR, ATN 004 Protocol Team. A multisite randomized trial of a motivational intervention targeting multiple risks in youth living with HIV: initial effects on motivation, self-efficacy, and depression. J Adolesc Health 2010 May;46(5):422-428 [FREE Full text] [doi: 10.1016/j.jadohealth.2009.11.198] [Medline: 20413077]

80. Samet JH, Horton NJ, Meli S, Dukes K, Tripps T, Sullivan L, et al. A randomized controlled trial to enhance antiretroviral therapy adherence in patients with a history of alcohol problems. Antivir Ther 2005;10(1):83-93. [Medline: 15751766]

XSL•FO

RenderX

81. Morgenstern J, Longabaugh R. Cognitive-behavioral treatment for alcohol dependence: a review of evidence for its hypothesized mechanisms of action. Addiction 2000 Oct;95(10):1475-1490. [doi: 10.1046/j.1360-0443.2000.951014753.x] [Medline: 11070524]

Abbreviations

ASSIST: Alcohol, Smoking and Substance Involvement Screening Test
ATN: Adolescent Medicine Trials Network for HIV/AIDS Interventions
BRIEF-A: Behavior Rating Inventory of Executive Function
CAS: condomless anal sex
CDC: Center for Disease Control and Prevention
CET: comparative effectiveness trial
CHEST: Center for HIV Educational Studies and Training
CHWs: community health workers
CT: Chlamydia trachomatis
EPIS: Exploration, Preparation, Implementation, Sustainment model
GC: Neisseria gonorrhea
GMM: growth mixture modeling
IP: immediate posttest
MI: motivational interviewing
MITI: Motivational Interviewing Treatment Integrity
MSM: men who have sex with men
NIH: National Institutes of Health
OMS: online master screener
PrEP: pre-exposure prophylaxis
REC: recruitment and retention center
SRVs: subject recruitment venues
STIs: sexually transmitted infections
YMHP: Young Men's Health Project
YMSM: young men who have sex with men

XSL•FO
RenderX

Protocol

Implementation Science Research Examining the Integration of Evidence-Based Practices Into HIV Prevention and Clinical Care: Protocol for a Mixed-Methods Study Using the Exploration, Preparation, Implementation, and Sustainment (EPIS) Model

April Idalski Carcone[1], PhD; Karin Coyle[2], PhD; Sitaji Gurung[3], MD, MPH; Demetria Cain[3], PhD, MPH; Rafael E Dilones[3], BSc; Laura Jadwin-Cakmak[4], MPH; Jeffrey T Parsons[3,5,6], PhD; Sylvie Naar[7], PhD

[1]Department of Family Medicine and Public Health Sciences, Wayne State University, Detroit, MI, United States

[2]Education, Training, and Research, Scotts Valley, CA, United States

[3]Center for HIV Educational Studies and Training, Hunter College, City University of New York, New York, NY, United States

[4]Department of Health Behavior and Health Education, University of Michigan School of Public Health, Ann Arbor, MI, United States

[5]Hunter Department of Psychology, Hunter College, City University of New York, New York, NY, United States

[6]Health Psychology and Clinical Science Doctoral Program, Graduate Center, City University of New York, New York, NY, United States

[7]College of Medicine, Florida State University, Tallahassee, FL, United States

Corresponding Author:
Sylvie Naar, PhD
College of Medicine
Florida State University
Main Campus
1115 West Call Street
Tallahassee, FL, 32306
United States
Phone: 1 248 207 2903
Email: sylvie.naar@med.fsu.edu

Abstract

Background: The Exploration, Preparation, Implementation, and Sustainment (EPIS) model is an implementation framework for studying the integration of evidence-based practices (EBPs) into real-world settings. The EPIS model conceptualizes implementation as a process starting with the earliest stages of problem recognition (Exploration) through the continued use of an EBP in a given clinical context (Sustainment). This is the first implementation science (IS) study of the integration of EBPs into adolescent HIV prevention and care settings.

Objective: This protocol (ATN 153 EPIS) is part of the Scale It Up program, a research program administered by the Adolescent Medicine Trials Network for HIV/AIDS Interventions (ATN), described in this issue by Naar et al. The EPIS study is a descriptive study of the uptake of 4 EBPs within the Scale It Up program. The goal of EPIS is to understand the barriers and facilitators associated with the Preparation, Implementation, and Sustainment of EBPs into HIV prevention and clinical care settings.

Methods: The EPIS study is a convergent parallel mixed-methods IS study. Key implementation stakeholders, that is, clinical care providers and leaders, located within 13 ATN sites across the United States will complete a qualitative interview conducted by telephone and Web-based surveys at 3 key implementation stages. The Preparation assessment occurs before EBP implementation, Implementation occurs immediately after sites finish implementation activities and prepare for sustainment, and Sustainment occurs 1 year postimplementation. Assessments will examine stakeholders' perceptions of the barriers and facilitators to EBP implementation within their clinical site as outlined by the EPIS framework.

Results: The EPIS baseline period began in June 2017 and concluded in May 2018; analysis of the baseline data is underway. To date, 153 stakeholders have completed qualitative interviews, and 91.5% (140/153) completed the quantitative survey.

Conclusions: The knowledge gained from the EPIS study will strengthen the implementation and sustainment of EBPs in adolescent prevention and clinical care contexts by offering insights into the barriers and facilitators of successful EBP implementation and sustainment in real-world clinical contexts.

XSL•FO
RenderX

Introduction

Background

Over the past 25 years, behavioral scientists have developed a number of efficacious interventions to reduce HIV transmission and improve self-management among those living with HIV. Between 2003 and 2014, the overall incidence of HIV in the United States decreased by 25%, yet youth aged 13 to 24 experienced a 43% increase [1] and accounted for a quarter (26%) of new HIV infections. More than half of (60%) of youth living with HIV are unaware of their HIV status. Once diagnosed, less than two-thirds are linked to HIV clinical care within 1 year, and just over half (54%) achieve viral suppression. Hence, fewer than 10% of US youth are and remain virally suppressed [2]. These data clearly illustrate that implementation of efficacious interventions in settings that serve youth has not yet been fully realized.

Implementation science is the study of methods and factors influencing the translation of research and other evidence-based practices (EBPs) into routine care [3]. Multiple implementation theories and models have been proposed for the prediction or explanation of the process of adopting and sustaining EBPs within the social sector. Where theories seek to generalize predictable pathways of translating knowledge into practice, determinant models and frameworks attempt to explain the factors that influence various stages of adoption, implementation, and sustainability in specific fields and contexts [4]. Determinant models originating from child welfare and mental health fields may be particularly pertinent to the HIV field because of the similar ways in which social context influences program delivery to youth and the adoption of new practices by the clinical care providers.

The Exploration, Preparation, Implementation, Sustainment (EPIS) model [5,6] is an implementation framework studying the integration of EBPs into real-world settings. A strength of the EPIS model is its view of EBP implementation across 4 phases [7]. The Exploration phase involves the recognition of a concern or opportunity for improvement. In Preparation, there is a decision to adopt an EBP. Implementation refers to the active integration of the EBP into routine care, whereas Sustainment examines the continued use of the new EBP. Within each phase, EPIS outlines and highlights the interplay between critical inner (internal to the organization, eg, organizational leadership and clinician characteristics) and outer (external systems, eg, political environment, funding, and other resources)

contextual factors likely to impact EBP implementation. A number of reliable, validated measures of these inner and outer contextual factors have been published in the research literature (see Measures section for a description of selected measures), making the EPIS model an ideal framework for the study of EBP implementation in HIV clinical care settings [8,9]. Finally, the EPIS model has been successfully used to study EBP uptake in similar multisite effectiveness trials such as the JJ-TRIALS and SAT2HIV [10-12] studies.

Aims and Objectives

This paper describes the EPIS research protocol, a study being conducted by the Adolescent Medicine Trials Network for HIV/AIDS Interventions (ATN; referenced as ATN 153 EPIS). EPIS is a mixed-methods implementation science (IS) research study of the uptake of 4 EBPs across the United States at ATN research sites. Thus, EPIS is 1 study within a larger program of research, "Scale It Up," to improve HIV-related self-management among youth living with or at risk of contracting HIV [13]. The 4 EBPs include sequential multiple assignment randomized trial (referred to as ATN 144 SMART), an adaptive intervention that combines short message service text messaging and cell phone support to increase antiretroviral therapy adherence among youth living with HIV (see the study by Belzer et al [14] in this issue). Scale It Up also includes a comparative effectiveness trial of clinic- versus telephone-delivery of the Young Men's Health Project (referred to as ATN 145 YMHP), a 4-session intervention to reduce the risk of HIV infection among young men who have sex with men (see the study by Parsons et al [15] in this issue). The tailored motivational interviewing (MI) study (referred to as ATN 146 TMI) aims to scale up the use of an EBP, MI, in adolescent HIV clinical care settings (see the study by Naar et al [16] in this issue). Finally, a comparative effectiveness trial to assess the additive benefit of communication training during couples' HIV testing and counseling (referred to as ATN 156 We Test; see the study by Starks et al [17] in this issue). The goal of the EPIS study is to describe the inner and outer contextual factors impacting the uptake of these 4 EBPs across 3 implementation phases. In years 1 to 2, as sites prepare for the integration of EBPs into their clinical care routines, the EPIS study will assess several providers and organizational characteristics that may impact the implementation and sustainment of EBPs at each clinical site (Table 1). Years 3 to 4 will focus on understanding the barriers and facilitators sites experienced during Implementation, and year 5 will assess plans for Sustainment.

XSL•FO

RenderX

Table 1. Exploration, Preparation, Implementation, and Sustainment (EPIS) model Inner (I) and Outer (O) context factors to be explored in the EPIS protocol.

Factors	Data source	EPIS phase/timeline for data collection		
		Preparation (Years 1 to 2)	Implementation (Years 2 to 3)	Sustainment (Years 4 to 5)
Leadership (I[a])	Survey	✓[b]	✓	✓
Organizational culture and climate (I)	Interviews; survey	✓	✓	✓
Fiscal viability and resources (I, O[c])	Interviews; survey	✓	✓	✓
Experience with evidence-based practices (I)	Interviews	✓	✓	✓
Attitudes toward evidence-based practices, including perceived barriers and facilitators (I)	Interviews; survey	✓	✓	✓
Facilitator/provider characteristics (I)	Survey	✓	✓	✓
Intervention fit (I)	Interviews; survey	✓	✓	✓
Interorganizational networks (O)	Interviews	✓	✓	✓
Fidelity monitoring and support[d]	Clinical records	—[e]	✓	✓
Perceived client outcomes	Interviews	✓	✓	✓

[a]I: inner context factor.

[b]Factor collected at a given EPIS phase/timeline.

[c]O: outer context factor.

[d]Fidelity data (defined as the extent to which providers adhere to treatment protocols) will be collected as part of the Scale It Up individual study protocols.

[e]Not applicable.

Methods

Design

This study will use a convergent parallel mixed-methods design [18] with data collected at 3 critical implementation phases: preimplementation (Prepare), postimplementation (Implementation), and sustainment. Participants will be enrolled in the EPIS study for up to 40 months. Preimplementation interviews will be conducted before EBP implementation, beginning in June 2017 and concluding in March 2018. The postimplementation interviews are scheduled to coincide with the sites' completion of the implementation phase, beginning in March 2019. Sustainment interviews will begin in March 2020 to capture participant perceptions of sustainment 1 year postimplementation. At each phase, participants will complete a qualitative interview by telephone and a quantitative survey via electronic data capture. Questions will focus on participants' perceptions of the barriers and facilitators to EBP implementation within their clinical site as outlined by the EPIS model.

Participants and Targeted Sites

All medical providers and staff with patient contact ("Key Stakeholders") at 13 ATN sites participating in the

aforementioned Scale It Up research projects will be eligible to participate (Table 2). Patient contact is defined as having direct patient interaction across several points of care, including prevention, counseling and testing, linkage to care, HIV primary care, services to promote retention and adherence to medications, and other medical or psychosocial services. Key stakeholders will also include administrative and research staff with key decision-making roles (eg, division chief and clinic director) who will provide input on prevention and care services and site operations. Each site will identify a clinical leader ("Site PI") to represent the organizational leadership perspective. There are no exclusion criteria. Participant turnover will be managed by maintaining the participant's responses collected up to the point of separation as a part of the study data corpus, but participants will not be retained in the study post separation. Similarly, if a site discontinues its participation, participants associated with that site will remain part of the study data corpus. Newly hired medical providers and staff at the 2 follow-up points will be invited to participate. Different sites participated in different Scale It Up projects because of the differing nature of each EBP being tested and the hybrid design selected for each effectiveness-implementation trial (see the study by Naar et al [13] in this issue). For example, ATN 146 used providers as the participants, but the other 3 trials primarily used patients as the unit of analysis.

Table 2. Scale It Up projects and participating sites in the Exploration, Preparation, Implementation, and Sustainment protocol.

Site	City, State	ATN[a] 144 SMART[b]	ATN 145 YMHP[c]	ATN 146 TMI[d]	ATN 156 We Test
Johns Hopkins University	Baltimore, MD	X[e]	—[f]	X	—
University of Alabama at Birmingham/Birmingham AIDS Outreach	Birmingham, AL	X	—	X	—
Center for HIV Educational Studies and Training at Hunter College[g]	New York, NY	—	—	—	X
State University of New York Downstate Medical Center	Brooklyn, NY	X	—	X	—
Wayne State University Prevention	Detroit, MI	—	X	—	X
Children's Hospital of Los Angeles	Los Angeles, CA	X	—	X	—
St. Jude Children's Research Hospital	Memphis, TN	X	—	X	—
University of Miami	Miami, FL	X	X	X	X
Tulane University[h]	New Orleans, LA	X	—	X	—
Children's Hospital of Philadelphia	Philadelphia, PA	X	X	X	—
University of California, San Diego	San Diego, CA	X	—	X	X
University of South Florida	Tampa, FL	X	—	X	—
Children's National Health System	Washington, D.C.	X	—	X	—

[a]ATN: Adolescent Medicine Trials Network for HIV/AIDS Interventions.

[b]SMART: Sequential Multiple Assignment Randomized Trial.

[c]YMHP: Young Men's Health Project.

[d]TMI: Tailored Motivational Interviewing Implementation Intervention.

[e]Site is participating in given SIU project and receives relevant questions for Exploration, Preparation, Implementation, and Sustainment model.

[f]Not applicable.

[g]Postimplementation and sustainment phase only.

[h]Preimplementation phase only.

Before each data collection effort, each site will provide a list of the medical providers and staff with direct patient contact. This list will include names, contact information (phone number and email), and role(s) within the clinic. Potential participants will receive an initial "enrollment email" introducing them to the EPIS model and study and providing them with instructions for scheduling their qualitative interview through a Web-based scheduling system. After the initial email, potential participants will be sent reminders every 2 weeks throughout the baseline study period about project enrollment. All sites have agreed to permit participants to participate in EPIS data collection efforts during their regularly scheduled work hours. Participants will be provided a list of available interview times from which they can choose an interview time most convenient for their schedule and availability. Participants will also be given the option of directly emailing their availability to arrange the most convenient interview. Interviewers are centralized, providing available times for all sites and will call participants at the scheduled time.

Upon completion of the interview, participants receive a link to complete the survey in Qualtrics. Participants who complete both the qualitative interview and quantitative survey receive a US $10 Amazon e-gift card. If a participant completes all 3 assessments (ie, preimplementation, implementation, and sustainment), they can receive a total of US $30 in Amazon e-gift cards. Participants who have not completed the quantitative survey will receive periodic reminders to do so for the duration of the data collection window.

All study procedures were approved by the institutional review board of the Scale It Up principal investigator's (PI) academic institution. All participants provided oral informed consent before the initiation of any study activity.

Assessments

Assessments will elicit participants' perceptions of barriers and facilitators to EBP implementation and sustainment at 3 critical implementation phases: preimplementation (Prepare), postimplementation (Implementation), and sustainment. The baseline assessment (June 2017-March 2018) will capture preimplementation feedback on anticipated barriers and facilitators for the specific EBPs each site will be implementing. The first follow-up assessment will occur postimplementation (March 2019-February 2020) and will assess barriers and facilitators experienced during EBP implementation and query anticipated barriers and facilitators to sustaining the EBPs. The second follow-up assessment (March 2020-February 2021) will assess barriers and facilitators experienced during the initial (1 year postimplementation) sustainment period.

Interviews

Trained interviewers will conduct interviews by telephone using a semistructured interview guide. Interview domains will include gathering information about the participant's professional background and experience, clinical site organization and structure, familiarity with EBPs in general, familiarity with the specific EBPs being implemented, and perceived barriers and facilitators to implementing the specific EBPs. In addition, site PIs will be asked about organizational history with EBPs, internal (organizational) and external (community and state) leadership structures, and their site's political context (policies and funding mechanisms) and fiscal considerations. It is estimated that key stakeholder interviews will require 30 min to 60 min to complete. Site PI interviews will require 60 min to 90 min to complete and thus will be completed in 2 parts (30 min to 60 min each).

Interviewer training will include prework for priming before the training and a 2-part live virtual training with modeling. Follow-up support will include interviewers conducting 2 mock interviews with self-assessment and trainer ratings and feedback following each mock interview; the rating forms were adapted from the study by Amico [19]. Interviewers who do not achieve adequate ratings on the second mock interview will complete a third to determine if they are fit for the interviewer role. Once data collection begins, the project team will hold monthly interview support calls that focus on reviewing and problem-solving issues raised by interviewers or identified through a review of transcripts. Interviewers will also be able to trigger immediate support through a Web-based technical assistance support form. Training procedures will be initiated 1 month before each data collection point.

Interviews will be audio-recorded and, immediately upon completion of the interview, uploaded to a secure server for storage. Audio files will be electronically transferred to a professional transcription service. Transcriptionists will provide a verbatim, deidentified transcript of the interview. Deidentification will involve removing participant and clinic staff member names. Research staff will review transcripts for quality (ie, accuracy) and confidentiality (ie, deidentification) before releasing the data for coding. Interview data will be uploaded to NVivo Version 12 (QSR International, Inc) for analysis.

Survey

Key stakeholders' and Site PIs' attitudes toward the adoption of EBPs will be assessed with the *Evidence-Based Practice Attitude Scale* (EBPAS; Aarons) [20]. The EBPAS assesses 4 attitudinal dimensions with strong internal consistency reliability: intuitive *Appeal* of EBP (alpha=.80), likelihood of adopting EBP given *Requirements* to do so (alpha=.90), *Openness* to new practices (alpha=.78), and perceived *Divergence* from usual practice with research-based / academically developed interventions (alpha=.59). They will also complete an updated version of the scale, the *Evidence-Based Practice Attitude Scale-50* (EBPAS-50; Aarons et al), which assesses 8 additional attitudinal domains [21]. The EBPAS-50 assesses the following: EBPs *Limitations* and their inability to address client needs (alpha=.92), EBP *Fit* with the

values and needs of the client and clinician (alpha=.88), negative perceptions of *Monitoring* or supervision (alpha=.87), the *Balance* of skills and the role of science in treatment (alpha=.79), time and administrative *Burden* associated with learning EBPs (alpha=.77), likelihood of increased *Job Security* or professional marketability provided by learning an EBP (alpha=.82), *Organizational Support* for learning an EBP (alpha=.85), and positive perceptions of receiving *Feedback* related to service delivery (alpha=.82).

Participants' perceptions of organizational climate will be assessed with 3 measures. Key stakeholders' and Site PIs' perceptions of organization climate, in general, will be assessed with the *Organizational Climate Measure* (OCM; Patterson et al) [22]. The OCM assesses organizational policies, practices, and procedures that provide a contextual backdrop for interactional patterns and behaviors that foster creativity, innovation, safety, or service within the organization, in other words, teamwork. Subscales will include the emphasis given to *Quality* procedures (alpha=.80), *Training* or a concern with developing employee skills (alpha=.83), and *Performance Feedback* (alpha=.78), which refers to the measurement and feedback of job performance. They will also complete the *Implementation Climate Scale* (ICS; Ehrhart et al) [23]. The ICS reliably assesses the extent to which a clinic fosters EBP implementation across 6 dimensions: *Focus* on EBP (alpha=.91), *Educational Support* for EBP (alpha=.84), *Recognition* for EBP (alpha=.88), *Rewards* for EBP (alpha=.81), *Selection* for EBP (alpha=.89), and *Selection* for openness (alpha=.91). Key stakeholders will only complete the *Perceived Organizational Support Scale* (POS; Rhoades et al) [24]. The POS assesses general beliefs about the extent to which an organization values employees' contributions and cares about their well-being (alpha=.90).

Key stakeholders and Site PIs will also evaluate the role of leadership in the implementation of EBPs using 2 scales: the *Director Leadership Scale*, (DLS; Broome et al) [25] and the *Implementation Leadership Scale* (ILS; Aarons et al) [26]. The DLS is a brief global assessment of organizational leadership with strong internal consistency (alpha=.90). The ILS assesses strategic leadership support for EBP implementation with 4 subscales: *Proactive* leadership (alpha=.95), *Knowledgeable* leadership (alpha=.96), *Supportive* leadership (alpha=.95), and *Perseverant* leadership (alpha=.96).

The extent to which the strategies, procedures, and elements of the 4 EBPs being implemented in the Scale It Up program match the values, needs, skills, and available resources (contextual fit) will be assessed with an adapted version of the *Self-Assessment of Fit in Schools* [27]. Key stakeholders and Site PIs will rate the extent to which they have the skills required to implement the EBPs, their comfort with the different elements of the EBPs, consistency of the EBPs with current clinical practices, ease of implementation including availability of resources and administrative support for the implementation of the EBPs, and perceived efficacy of the EBPs.

Site PIs will assess the extent to which their staff contributes to EBP implementation by demonstrating behaviors that go beyond minimum requirements using the *Implementation*

Citizenship Behavior Scale (ICBS; Ehrhart et al) [28]. The ICBS assesses 2 domains: helping others (alpha=.93) and keeping informed (alpha=.91). Finally, all participants will complete an investigator-developed survey to collect basic demographic information, such as position, years in position, race, ethnicity, gender identity, and current caseload. It is estimated that it will require participants 60 min to 90 min to complete the survey.

Analysis Plan

The analyses will focus on understanding the barriers and facilitators located within sites' inner and outer context that is associated with implementing and sustaining EBPs into HIV care settings. Analyses will be guided by the following questions: (1) How do inner context factors (eg, organizational culture and climate and leadership) influence EBP implementation and sustainment? (2) How do outer context factors (eg, fiscal viability and interorganizational networks) influence EBP implementation and sustainment? (3) To what extent do the perceptions of key stakeholders and clinical leaders (ie, site PIs) vary, and how does that variation affect EBP implementation and sustainment? (4) To what extent do stakeholder perceptions (key stakeholder and site PI combined) vary by site (ie, organizational structure)?

Qualitative Analysis Plan

First, consistent with Morgan's [29] recommendations for qualitative content analyses and Hsieh and Shannon's [30] directed qualitative content analytic approach, standard definitions of the concepts of interest will be developed on the basis of the EPIS model. Each interview will be systematically reviewed at each time point for all thematic mentions of the following: (1) features of the inner and outer context per EPIS that have the potential to influence implementation of an EBP, (2) people who have the potential to influence implementation of an EBP, and (3) personal perceptions of the EBP in question and other EBPs that have the potential to improve patient outcomes. Within these longer thematic lists, we will then separate out specific categories of work-setting characteristics (eg, leadership, incentives, and disincentives for innovating), people (eg, patients, nurses, physicians, administrators, experts, and novices), and perceptions of evidence-based interventions (eg, feasible and advantageous), initially using existing theory to guide categorization but also allowing themes to emerge from the data through open coding procedures [31,32]. This combined inductive and deductive coding approach will allow us to both validate and extend the EPIS model. Revision of our initial coding categories will occur iteratively until we reach saturation in the identification of new codes. During this iterative process, categories and their definitions will be refined and subcategories of codes will be consolidated, consistent with an axial-coding process. At this point, we will return to each interview and systematically apply the final, revised set of codes. In addition, case codes will be applied to each interview to reflect clinic role, site, cluster, and relevant demographic characteristics of the respondent.

The coding team will be led by the EPIS study PI, an experienced PhD-level mixed-methods researcher. A total of 3 coders, 2 research assistants with, at minimum, a baccalaureate degree, and 1 postdoctoral fellow with qualitative coding experience will code all the data. Coders will undergo initial training to familiarize themselves with the EPIS model, its constructs, and the operational definitions developed for the study. Coders will also be trained in the analytic approach, including the coding software. Coders will first collaboratively code 6 interviews (3 site PI and 3 key implementers) to familiarize themselves with the data and finalize the working codebook. An initial assessment of intercoder reliability will be conducted on 2 interviews (1 site PI and 1 key implementer). Coders will not be released for independent coding unless their intercoder reliability is at a minimum of 0.60 or higher as assessed by Cohen kappa [33]. To ensure intercoder reliability is maintained, a random selection of 30% of the interviews will be co-coded to ensure that the kappa coefficient remains 0.60 or higher [33]. After each intercoder reliability assessment, coders will meet to discuss and resolve coding discrepancies. Finally, the coding team is supported by 3 consultants with expertise in IS and/or HIV qualitative research.

The coded data will be comparatively analyzed both within and across time to examine differences at the setting and provider-level in quality and extent of EBP implementation. Examining the segments of text that are associated with differences in the frequency of categories between, for example, high-fidelity and low-fidelity sites, and examination of patterns in the presence and absence of thematic categories will allow us to provide empirically grounded explanations for differences in study outcomes.

Quantitative Analysis Plan

Analysis will begin by examining the psychometric properties, for example, internal consistency reliability using Cronbach alpha for all scales and subscales of established measures. Measures demonstrating insufficient reliability (eg, internal consistency $<.70$) in the study sample will be further examined with an exploratory factor analysis using Promax oblique rotation. Items with loadings $<.40$ or strong cross-loadings may be excluded for further analyses. Intercorrelations among items within each subscale and subscales within each measure will be examined; the correlation among measures will also be examined. Once reliability in the sample is established, descriptive analyses will be used to summarize the inner and outer contextual factors within and across sites and by informant (eg, site PIs and key implementers; clinical care providers and administrative staff). At baseline, we will examine mean differences in perceptions of intervention fit and attitudes toward EBP across site PIs and key implementers. We will also assess how perceptions vary as a function of Implementer demographics. At each follow-up, a comparison of changes in the inner and outer context factors over time (ie, from baseline to postintervention and sustainment) using a multivariate analysis will be conducted. Mixed linear effects models, adjusting for covariates, including age, time in position, role in clinic, experience level, and site-level factors, will be used to explore the impact these factors have on the overall implementation and sustainability of Scale It Up projects across sites and patient outcomes.

Mixed-Methods Analysis Plan

To offer findings in ways that move beyond the particularistic view of EBP implementation within the sites, once all of the data are coded across all time points, we will adopt the innovation profile approach [34] originally developed for classroom research. The approach results in a multidimensional rubric to classify where an organization is in the process of developing its capacity to engage in a particular set of activities, in this case, the integration of EBPs into routine patient care. The dimensions and subdimensions of the matrix we develop, as well as descriptions of the behavioral indicators of exemplary, intermediate, emerging, and low capacity to integrate EBPs, will be derived from aggregating the data produced during the analysis to the site level. These data will be integrated with quantitative fidelity data collected by the intervention protocol teams with equal weight given to qualitative and quantitative data sources [35]. We will follow best practices for conducting mixed-method designs in the health sciences as outlined by the Office of Behavioral and Social Sciences Research [36]. These include employing rigorous procedures in the methods of data collection and analysis and integrating the multiple sources of data toward the goal of obtaining rich, descriptive output.

Results

EPIS data collection was launched in June 2017 and, at the writing of this paper, the first phase (Preparation) of data collection has concluded, and analyses are underway. A total of 140 of 282 eligible stakeholders completed both components of the first EPIS data collection. The baseline data collection window closed with a small proportion of providers (13, 8.5%) having partially completed the baseline assessment, that is, the qualitative component was completed, and the quantitative survey remains outstanding. About 20% (56) declined to participate and the remaining stakeholders did not respond to the enrollment invitation before the closure of the baseline data collection window. Figure 1 illustrates completion rates by site. Follow-up data collections are scheduled to begin in March 2019 (Implementation) and March 2020 (Sustainment).

Figure 1. Total participant enrollment status per site till May 2018.
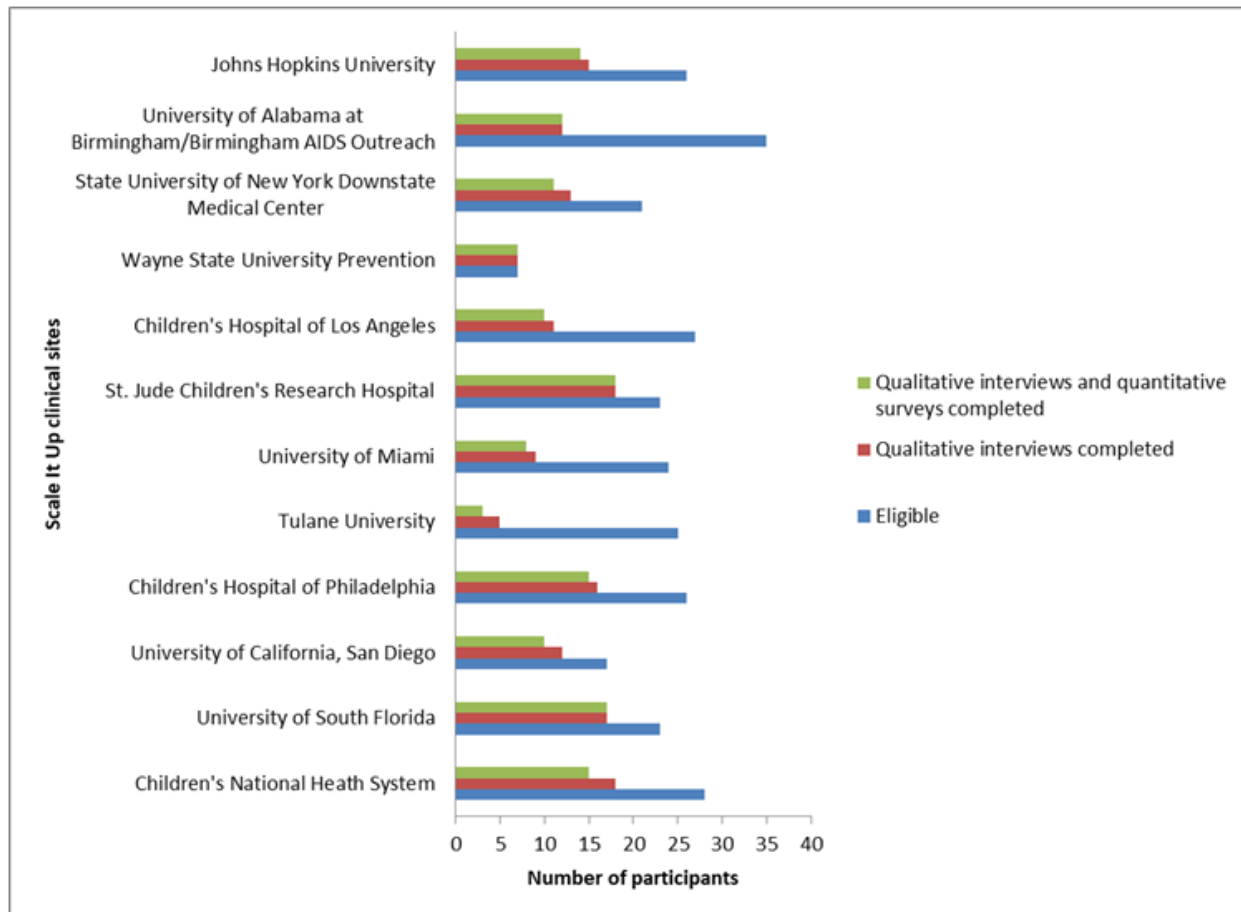
Discussion

Protocol Goal

Although EBPs have demonstrated success in the academic setting, many challenges can prevent an EBP's successful implementation and sustainment in real-world clinical contexts. The goal of the EPIS IS study is to generate knowledge about the barriers and facilitators to the implementation and sustainment of EBPs into adolescent HIV prevention and clinical care settings. Understanding the factors that impact organizations, clinics, and practitioners throughout the EBP implementation process will facilitate the adoption of EBPs by tailoring implementation to fit within the needs and culture of the organization and/or clinic.

Limitations

The EPIS sample is limited to the 13 participating ATN clinics and the medical providers and staff with direct patient contact within these clinical settings. These participants may not be representative of service providers in other contexts. This study and the Scale It Up program are focused on the implementation of EBPs in multidisciplinary adolescent HIV settings. The EPIS model was developed in child welfare [6,37] and has begun to be applied to other service sectors including behavioral health care [38] and juvenile justice [10]. In general, the findings from this research will add to the growing literature on IS and particularly the EPIS model, which will facilitate translation to other clinical settings.

Implications

This study is the first IS study of EBP implementation in adolescent HIV settings. The knowledge gained from the EPIS study will strengthen the implementation and sustainment of EBPs in both adolescent prevention and clinical care contexts by offering insights into the barriers and facilitators of successful EBP implementation and sustainment in real-world clinical contexts.

Conflicts of Interest

None declared.

References

1. Frieden TR, Foti KE, Mermin J. Applying public health principles to the HIV epidemic--how are we doing? N Engl J Med 2015 Dec 3;373(23):2281-2287. [doi: 10.1056/NEJMms1513641] [Medline: 26624243]
2. Zanoni BC, Mayer KH. The adolescent and young adult HIV cascade of care in the United States: exaggerated health disparities. AIDS Patient Care STDS 2014 Mar;28(3):128-135 [FREE Full text] [doi: 10.1089/apc.2013.0345] [Medline: 24601734]
3. Eccles MP, Mittman BS. Welcome to implementation science. Implementation Sci 2006 Feb 22;1(1) [FREE Full text] [doi: 10.1186/1748-5908-1-1]
4. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
5. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Adm Policy Ment Health 2011 Jan;38(1):4-23 [FREE Full text] [doi: 10.1007/s10488-010-0327-7] [Medline: 21197565]
6. Aarons GA, Green AE, Willging CE, Ehrhart MG, Roesch SC, Hecht DB, et al. Mixed-method study of a conceptual model of evidence-based intervention sustainment across multiple public-sector service settings. Implement Sci 2014 Dec 10;9:183. [doi: 10.1186/s13012-014-0183-z] [Medline: 25490886]
7. Hanson RF, Self-Brown S, Rostad WL, Jackson MC. The what, when, and why of implementation frameworks for evidence-based practices in child welfare and child mental health service systems. Child Abuse Negl 2016 Mar;53:51-63 [FREE Full text] [doi: 10.1016/j.chiabu.2015.09.014] [Medline: 26547360]
8. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
9. Hurlburt M, Aarons GA, Fettes D, Willging C, Gunderson L, Chaffin MJ. Interagency collaborative team model for capacity building to scale-up evidence-based practice. Child Youth Serv Rev 2014 Apr;39:160-168 [FREE Full text] [doi: 10.1016/j.childyouth.2013.10.005] [Medline: 27512239]
10. Becan JE, Bartkowski JP, Knight DK, Wiley TR, DiClemente R, Ducharme L, et al. A model for rigorously applying the Exploration, Preparation, Implementation, Sustainment (EPIS) framework in the design and measurement of a large scale

collaborative multi-site study. Health Justice 2018 Apr 13;6(1):9 [FREE Full text] [doi: 10.1186/s40352-018-0068-3] [Medline: 29654518]

11. Garner BR, Zehner M, Roosa MR, Martino S, Gotham HJ, Ball EL, et al. Testing the implementation and sustainment facilitation (ISF) strategy as an effective adjunct to the Addiction Technology Transfer Center (ATTC) strategy: study protocol for a cluster randomized trial. Addict Sci Clin Pract 2017 Dec 17;12(1):32 [FREE Full text] [doi: 10.1186/s13722-017-0096-7] [Medline: 29149909]

12. Moullin J, Dickson K, Stadnick NA, Rabin B, Aarons GA. Systematic review of the Exploration, Preparation, Implementation, Sustainment (EPIS) framework. Implement Sci 2019 Jan 5;14(1):1 [FREE Full text] [doi: 10.1186/s13012-018-0842-6] [Medline: 30611302]

13. Naar SP, Parsons JT, Stanton BF. Adolescent trials network for HIV-AIDS Scale It Up program: protocol for a rational and overview. JMIR Res Protoc 2019 Feb 1;8(2):e11204 [FREE Full text] [doi: 10.2196/11204] [Medline: 30707102]

14. Belzer ME, MacDonell KK, Ghosh S, Naar S, McAvoy-Banerjea J, Gurung S, et al. Adaptive antiretroviral therapy adherence interventions for youth living with HIV through text message and cell phone support with and without incentives: protocol for a sequential multiple assignment randomized trial (SMART). JMIR Res Protoc 2018 Dec 20;7(12):e11183-e11150 [FREE Full text] [doi: 10.2196/11183] [Medline: 30573448]

15. Parsons JS, Starks TJ, Gurung S, Cain D, Marmo J, Naar S. Comparative effectiveness trial of clinic-based delivery of the Young Men's Health Project (YMHP) targeting HIV risk reduction and substance use among youth men who have sex with men (YMSM). JMIR Res Protoc 2019. [doi: 10.2196/11184]

16. Naar S, MacDonnell K, Chapman JE, Todd L, Gurung S, Cain D, et al. Testing a motivational interviewing implementation intervention in adolescent HIV clinics: protocol for a type 3, hybrid implementation-effectiveness trial. JMIR Res Protoc 2019 (forthcoming)(forthcoming). [doi: 10.2196/11200]

17. Starks TF, Feldstein Ewing SW, Lovejoy T, Gurung S, Cain D, Fan C, et al. Enhancing sexual safety: couples' communication and HIV testing among young men who have sex with men. JMIR Res Protoco 2019 (forthcoming)(forthcoming). [doi: 10.2196/11186]

18. Caruth G. Demystifying mixed methods research design: a review of the literature. Mevlana Int J Educ 2013 Aug;3(2):112-122 [FREE Full text] [doi: 10.13054/mije.13.35.3.2]

19. Amico K. Scribd. 2015. Interviewer Training Intensive Tool v1 URL: https://www.scribd.com/document/395458381/interviewer-training-intensive-tool-v1#download&from_embed

20. Aarons GA. Mental health provider attitudes toward adoption of evidence-based practice: the Evidence-Based Practice Attitude Scale (EBPAS). Ment Health Serv Res 2004 Jun;6(2):61-74. [doi: 10.1023/B:MHSR.0000024351.12294.65] [Medline: 15224451]

21. Aarons GA, Cafri G, Lugo L, Sawitzky A. Expanding the domains of attitudes towards evidence-based practice: the evidence based practice attitude scale-50. Adm Policy Ment Health 2012 Sep;39(5):331-340 [FREE Full text] [doi: 10.1007/s10488-010-0302-3] [Medline: 20607597]

22. Patterson MG, West MA, Shackleton VJ, Dawson JF, Lawthom R, Maitlis S, et al. Validating the organizational climate measure: links to managerial practices, productivity and innovation. J Organiz Behav 2005 Jun;26(4):379-408 [FREE Full text] [doi: 10.1002/job.312]

23. Ehrhart MG, Aarons GA, Farahnak LR. Assessing the organizational context for EBP implementation: the development and validity testing of the Implementation Climate Scale (ICS). Implement Sci 2014 Oct 23;9:157 [FREE Full text] [doi: 10.1186/s13012-014-0157-1] [Medline: 25338781]

24. Rhoades L, Eisenberger R, Armeli S. Affective commitment to the organization: the contribution of perceived organizational support. J Appl Psychol 2001 Oct;86(5):825-836. [doi: 10.1037/0021-9010.86.5.825] [Medline: 11596800]

25. Broome KM, Knight DK, Edwards JR, Flynn PM. Leadership, burnout, and job satisfaction in outpatient drug-free treatment programs. J Subst Abuse Treat 2009 Sep;37(2):160-170 [FREE Full text] [doi: 10.1016/j.jsat.2008.12.002] [Medline: 19339143]

26. Aarons GA, Ehrhart MG, Farahnak LR. The Implementation Leadership Scale (ILS): development of a brief measure of unit level implementation leadership. Implement Sci 2014 Apr 14;9(1):45 [FREE Full text] [doi: 10.1186/1748-5908-9-45] [Medline: 24731295]

27. Horner RH, Salentine S, Albin RW. Positive Behavioral Interventions & Supports (PBIS). 2003. Self Assessment of Contextual fit in Schools URL: https://www.pbis.org/resource/248/self-assessment-of-contextual-fit-in-schools [accessed 2019-02-11] [WebCite Cache ID 766LYGnZ6]

28. Ehrhart MG, Aarons GA, Farahnak LR. Going above and beyond for implementation: the development and validity testing of the Implementation Citizenship Behavior Scale (ICBS). Implement Sci 2015 May 7;10:65 [FREE Full text] [doi: 10.1186/s13012-015-0255-8] [Medline: 25948489]

29. Morgan DL. Qualitative content analysis: a guide to paths not taken. Qual Health Res 1993 Feb;3(1):112-121. [doi: 10.1177/104973239300300107] [Medline: 8457790]

30. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]

31. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage Publications; 2008.

32. Miles MB, Huberman AM, Saldaña J. Qualitative Data Analysis: A Methods Sourcebook. Thousand Oaks, CA: Sage Publications; 2014.

33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310] [Medline: 843571]

34. Leithwood K, Montgomery D. Improving Classroom Practice Using Innovation Profiles. Toronto, ON: The Ontario Institute for Studies in Education; 1987.

35. Leech NL, Onwuegbuzie AJ. A typology of mixed methods research designs. Qual Quant 2007 Mar 27;43(2):265-275. [doi: 10.1007/s11135-007-9105-3]

36. Creswell JW, Klassen AC, Plano Clark VL, Clegg Smith K, Office of Behavioral and Social Sciences Research. Best Practices for Mixed Methods Research in the Health Sciences. Bethesda, MD: National Institutes of Health (NIH); Aug 2011.

37. Willging CE, Green AE, Gunderson L, Chaffin M, Aarons GA. From a "perfect storm" to "smooth sailing": policymaker perspectives on implementation and sustainment of an evidence-based practice in two states. Child Maltreat 2015 Feb;20(1):24-36 [FREE Full text] [doi: 10.1177/1077559514547384] [Medline: 25125232]

38. Willging CE, Sommerfeld DH, Aarons GA, Waitzkin H. The effects of behavioral health reform on safety-net institutions: a mixed-method assessment in a rural state. Adm Policy Ment Health 2014 Mar;41(2):276-291 [FREE Full text] [doi: 10.1007/s10488-012-0465-1] [Medline: 23307162]

Abbreviations

ATN: Adolescent Medicine Trials Network for HIV/AIDS Interventions
DLS: Director Leadership Scale
EBP: evidence-based practice
EBPAS: Evidence-Based Practice Attitude Scale
EBPAS-50: Evidence-Based Practice Attitude Scale-50
EPIS: Exploration, Preparation, Implementation, and Sustainment Model
ICBS: Implementation Citizenship Behavior Scale
ICS: Implementation Climate Scale
ILS: Implementation Leadership Scale
IS: implementation science
MI: motivational interviewing
OCM: Organizational Climate Measure
PI: principal investigator
POS: Perceived Organizational Support Scale

<u>Protocol</u>

Cascade Monitoring in Multidisciplinary Adolescent HIV Care Settings: Protocol for Utilizing Electronic Health Records

Amy L Pennar[1], PhD; Tyra Dark[1], PhD; Kit N Simpson[2], DrPH; Sitaji Gurung[3], MD, MPH; Demetria Cain[3], PhD, MPH; Carolyn Fan[3], BA; Jeffrey T Parsons[3,4,5], PhD; Sylvie Naar[1], PhD

[1]College of Medicine, Florida State University, Tallahassee, FL, United States

[2]Department of Healthcare Leadership and Management, Medical University of South Carolina, Charleston, SC, United States

[3]Center for HIV Educational Studies and Training, Hunter College, City University of New York, New York, NY, United States

[4]Hunter College Department of Psychology, Hunter College, City University of New York, New York, NY, United States

[5]Health Psychology and Clinical Science Doctoral Program, Graduate Center, City University of New York, New York, NY, United States

Corresponding Author:
Sylvie Naar, PhD
College of Medicine
Florida State University
Main Campus
1115 West Call Street
Tallahassee, FL, 32306
United States
Phone: 1 248 207 2903
Email: sylvie.naar@med.fsu.edu

Abstract

Background: Past research shows that youth living with HIV (YLH) are not as engaged in the HIV treatment cascade as other HIV-positive populations. To achieve the health benefits of rapid and widespread testing and advanced pharmacologic treatment, YLH must be fully engaged in every stage of the treatment cascade. Cascade monitoring provides an opportunity to assess the youth care cascade, including engagement in care and when youth commonly drop out of care, across 10 clinical sites in the United States. Collecting electronic health record (EHR) data for prevention and care across participant recruitment venues within the Adolescent Medicine Trials Network (ATN) allows for monitoring of the prevention and care cascades within the ATN, for comparing the ATN population to large-scale surveillance, for future integration of technology-based interventions into EHRs, and for informing ATN strategic planning.

Objective: The aim of this protocol study is to examine the trends in treatment cascade, including whether patients are receiving antiretroviral therapy, adhering to regimens, attending care appointments, and maintaining suppressed viral loads, to guide new protocol development and to facilitate community engagement. This protocol is part of the ATN Scale It Up (SIU) program described in this issue.

Methods: Deidentified EHR data of YLH, aged 15 to 24 years, will be collected annually (2017 to 2022) from 10 ATN clinical sites, resulting in patient data from 2016 to 2021. These data will be transferred and stored using Dropbox Business, a Health Insurance Portability and Accountability Act–compliant site and then analyzed by the SIU analytic core.

Results: This study was launched in December 2017 in 10 clinical sites, with 2016's EHR data due on January 31, 2017. All 10 sites electronically uploaded their EHR data. The mandatory variables requested to monitor cascade of care include date of visit, age, gender, height, weight, race, ethnicity, viral load, and International Classification of Diseases codes for other diagnosis. In total, 70% of the sites provided data for all mandatory variables. The remaining mandatory variables were manually extracted.

Conclusions: This study will provide a platform to determine how YLH across the nation progress through or drop out of the HIV treatment cascade. It will also provide a foundation for assessing impact of SIU projects on treatment cascade outcomes.

International Registered Report Identifier (IRRID): DERR1-10.2196/11185

(JMIR Res Protoc 2019;8(5):e11185) doi:10.2196/11185

XSL·FO
RenderX

Introduction

Background

In the United States, youth living with HIV (YLH)—especially ethnic and racial minority youth and men who have sex with men (MSM)—are experiencing disproportionately high rates of morbidity [1], which places them at risk for early mortality. Despite the fact that dramatic decreases in HIV transmission are achievable with currently available treatments and interventions (such as antiretroviral treatment [ART], pre-exposure prophylaxis [PrEP], and rapid and widespread testing), such decreases have not been realized among youth. Only 42% of YLH aged 13 to 24 years are aware of their sero-status, compared with 83.5% of all HIV-positive Americans [2].

Therefore, it is clear that the HIV health care systems are failing to engage youth at risk for HIV and YLH at one or more steps of the cascade. The HIV prevention cascade includes routine HIV and sexually transmitted infection (STI) testing, as well as PrEP knowledge, access, uptake, and adherence when warranted [3-5]. The HIV treatment cascade describes care for those living with HIV, and includes diagnosis, linkage to care, timely initiation of care, persistence and adherence to ART, and sustained viral suppression.

Even when youth enter the treatment cascade through diagnosis, knowledge of HIV status does not necessarily result in linkage to care. Compared with 75% of adults, less than two-thirds of YLH are linked to care within 6 to 12 months after diagnosis [6]. Importantly, even linkage to care does not guarantee quality or effective care. The physician must initiate ART early, even if there is erratic behavior on the part of the youth, and YLH must recognize that ART is a lifelong commitment and requires a high degree of adherence. Even with appropriate ART initiation and adherence, drug resistance is present (even at baseline) in approximately 10% of youth [7-9]. Although effectiveness trials report viral suppression rates of 80% or more among YLH receiving ART, observational studies (ie, real-world occurrence) find much lower rates, closer to 50% [9]. Zanoni and Mayer presented the HIV care cascade of care for adolescent and young adults in the United States with declining numbers for those diagnosed, engaged in medical care, initiated antiretroviral therapy, and achieved viral suppression [9].

Ultimately, the goal of those involved in HIV prevention and care is to achieve maximum commitment to the HIV prevention and treatment cascades. Mathematical modeling indicates that even with 90% detection of HIV infections, followed by 90% engagement in care of YLH, 90% appropriate treatment of those in care, and 90% viral suppression of treated individuals, approximately 34% of YLH will remain viremic (ie, presence of a virus in the blood) [3]. In summary, implementation of sustainable service delivery interventions along multiple points in the cascade is necessary to achieve maximal benefits of increased access to ART and increase viral suppression [10].

To achieve sustained viral suppression, a youth must determine their HIV status and effectively engage in points along the HIV treatment cascade. Engagement should begin at the prevention cascade level, in that youth should have routine HIV testing and consider postexposure prophylaxis or PrEP if warranted. If HIV positive, the youth must immediately engage with the care system, initiate ART and comprehend and embrace the necessity of proper adherence, be retained in care, and maintain viral suppression. Collectively, these steps comprise proper HIV and sexual health self-management, defined as "strategies to help individuals...and their caregivers better understand and manage their illness and [/or] improve their health behaviors" [11]. Every step of the HIV treatment cascade requires at-risk youth or YLH to make decisions to engage with the system or to modify their behaviors, that is, every step requires active *self-management*. In fact, large numbers of at-risk youth are not seeking HIV/STI testing or other prevention services; if positive, they are not engaging in care. Even when they are in care, many youths are not sustaining adequate HIV care and treatment. They may be engaging in other risky behaviors, such as substance use, that interfere across all points in the cascade.

The Scale It Up (SIU) Research Projects [12] were funded as part of the Adolescent Medicine Trials Network (ATN) and designed to advance the field of implementation science by employing 3 types of effectiveness implementation hybrid designs, all addressing self-management and inner and outer context variables including sociopolitical culture, organizational characteristics, culture and climate, leadership, dynamics of the multidisciplinary team, facilitator characteristics and attitudes, training fidelity monitoring and support, efforts, intervention fit, and fiscal viability that are involved in successful implementation. SIU utilizes motivational interviewing (MI) to provide a clear framework for improving patient-provider communication and promoting behavior change (ie, improved self-management) using client-centered methods for enhancing motivation and self-efficacy [13]. These are provider-driven strategies to meet patients where they are currently at regarding behaviors and build motivation for increased change. High-quality patient-provider relationships are associated with a greater likelihood of patients' receiving ART, ART adherence, attending appointments, and having lower viral load [14-18]. Through enhancing implementation of treatment protocols and patient-provider relationships, SIU focuses on improving youths' outcomes along the prevention and care cascades. Examining trends in the care cascade is therefore critical for providing a foundation for outcomes assessment across SIU projects.

Cascade monitoring (CM) ATN 154 is a center-wide protocol in SIU focusing on the HIV care cascade as well as health factors (eg, cardiovascular functioning) known to be consequential for the long-term health outcomes of individuals living with HIV. Information collected in CM will monitor impact from other SIU protocols on potential gains in the treatment cascade. CM specifically collects only data that can be extracted from electronic health records (EHRs); therefore, CM does not include HIV prevention data, as the participating

clinical sites generally did not have prevention data available for extraction. Documenting trends in the care cascade will facilitate a deeper comprehension of when in the care cascade patients most commonly drop out, and what populations commonly disengage from the cascade, thereby providing researchers, clinicians, and community stakeholders a pulse on the adolescent HIV epidemic. Furthermore, the use of electronically extracted variables that measure general patient conditions by International Statistical Classification of Diseases and Related Health Problems-10th revision (ICD-10) codes, as well as cascade-specific variables, enables us to identify clusters of data points that are meaningful for monitoring differences in treatment cascade retention. These clusters of EHR variables, called informatics *EHR phenotypes*, can then be used to identify similar patients in other electronic data systems. The EHR phenotypes identified for YLH through CM will be used in the cost-effectiveness analysis for the SIU studies and used with large archival databases to estimate the prevalence of cascade failures to YLH in other community settings not included in the CM study.

Aims

CM provides an opportunity to assess trends to guide new protocol development and to facilitate community engagement. These data will support the Adolescent Medicine Trials Network (ATN) to have a finger on the pulse of the epidemic and will provide feedback to sites and community partners. Using the EHR phenotypes from CM, the patterns observed in the 10 SIU clinical sites can be extended to identify YLH in other settings, which can help policymakers and service providers improve systems and services and address barriers to care to better support individuals as they move from one step in the HIV care continuum to the next.

- Aim 1: Develop a common data model of variables of relevance for measuring cascade outcomes across 10 multidisciplinary adolescent HIV clinics and implement EHR data extraction (downloading) annually for 2016 to 2021 calendar years.
- Aim 2: Utilize data submission reports to identify a systematic EHR data extraction process to be utilized by clinic sites.
- Aim 3: Using extracted EHR data, determine how many of those retained in care and receiving ART keep clinic appointments and achieve viral suppression.
- Aim 4A: Determine when patients most commonly drop out of the HIV care continuum, and what populations commonly do so, in the SIU clinical sites.
- Aim 4B: Develop a phenotype for HIV-infected adolescents from the EHR CM data. Use the phenotypes applied to large national datasets to estimate variations in rates of cascade failure for YLH in the United States to help policymakers and service providers improve systems and services to better support individuals as they move from one step in the HIV care continuum to the next.
- Aim 5: Measure the effect of clinic-based interventions in SIU on longitudinal HIV care cascade outcomes across the 10 multidisciplinary adolescent HIV care clinics (approximately 1200 patients). This center-wide protocol will support the aims of the separate SIU projects.

- Aim 6: Apply the phenotype for HIV-infected adolescents from the EHR CM data to national databases to estimate costs weights for use in the assessment of cost-effectiveness of the SIU interventions.

Methods

Overview

This study (CM, ATN 154) is part of the SIU program as described in the overview paper in this issue [12]. Deidentified EHR data will be collected retrospectively from 10 clinical sites, also known as Subject Recruitment Venues (SRVs), participating in the SIU. The sites are as follows:

1. Site 1: Baltimore—Johns Hopkins University, Maryland.
2. Site 2: Birmingham—University of Alabama at Birmingham, Alabama.
3. Site 3: Brooklyn—SUNY Downstate Medical Center, New York.
4. Site 6: Los Angeles—Children's Hospital Los Angeles, California.
5. Site 7: Memphis—St. Jude Children's Research Hospital, Tennessee.
6. Site 8: Miami—University of Miami, Florida.
7. Site 10: Philadelphia—Children's Hospital of Philadelphia, Pennsylvania.
8. Site 11: San Diego—University of California San Diego, California.
9. Site 12: Tampa—University of South Florida, Florida.
10. Site 13: Washington, DC—Children's National Health System, District of Columbia.

The first extracts will contain standard of care and treatment visits in the full year of 2016 and associated data (see Table 1) for all YLH, aged 15 to 24 years, treated at sites. Subsequently, 1-year data extracts will be requested from sites annually, with the final year of data uploaded in 2022 for the full year of 2021 (see Figure 1).

Data Acquisition

Data collection will occur at all SIU clinical sites providing care to YLH by site personnel. The scientific and leadership teams have determined that there will not be resources allotted to do chart review for the CM measures; therefore, the CM protocol team will receive only measures available for extraction from the clinics' electronic data systems. The provision of electronically downloaded data, relative to hand extracted, also reduces the risk of data entry error, particularly for complex variables (eg, ICD-10 codes). However, given that longitudinal surveillance of viral load is critical to the success of the protocol, hand-extracted viral load data are the only exception when it cannot be electronically downloaded. All data transferred by the sites to the CM protocol team will be deidentified at the sites before data transfer. Having the sites create and update a master list of patients and unique identifiers is crucial for conducting longitudinal data analysis and following patient attributes and outcomes over time. Sites will be provided with the instructions on the unique identifiers that should be used to create participant identification numbers.

Table 1. Measures.

Type and variable	Metrics and notes
Time varying	
Patient ID	Unique to study, should not match patient's medical record file (ie, deidentified); see table below
Visit date	MM/DD/YYYY
Antiretroviral Medications—Name	Name of medication; include all prescribed
Viral load results	Quantitative value (parts per cubic ml)
CD4 cell count result	Absolute count
Height	Meters
Weight	Kg
Blood pressure—systolic	mmHg
Blood pressure—diastolic	mmHg
Cholesterol panel—total	mg/dL
Cholesterol panel—low density lipoprotein	mg/dL
Cholesterol panel—high density lipoprotein	mg/dL
Cholesterol panel—triglycerides	mg/dL
Blood glucose—hemoglobin A_{1c}	mg/dL
Blood glucose—fasting	mg/dL
Tobacco use—current status	current/past/never
Tobacco use—how much	Number of packs per day
Tobacco use—how long	Number of years
Chlamydia test result	Positive/negative OR reactive/nonreactive; if titer please enter positive or reactive
Gonorrhea test result	Positive/negative OR reactive/nonreactive; if titer please enter positive or reactive
Herpes test result	Positive/negative OR reactive/nonreactive; if titer please enter positive or reactive
Syphilis test result	Positive/negative OR reactive/nonreactive; if titer please enter positive or reactive
Human papilloma virus test result	Positive/negative OR reactive/nonreactive; if titer please enter positive or reactive
Other non-HIV medication—Name	Name of medication; include all prescribed
Patient zip code	Last known/documented
International Statistical Classification of Diseases and Related Health Problems-10th revision Diagnoses Codes	Include all from visit; extracted as left-hand justified character variables
CPT[a]-4 Procedure Codes	Include all from visit; extracted as 5-digit numerical variables
Health care common procedure coding system Procedure Codes	Include all from visit; if used instead of CPT codes; extracted as 5-digit character variable
Time invariant	
Patient ID	Unique to study, should not match patient's medical record file (ie, deidentified)
Current age	Age as of January 1 of the year of data download (eg, 01/01/2016 for 2016 data download)
Age at HIV diagnosis	Years
Date of HIV diagnosis	MM/DD/YYYY
Gender	Self-report; for example, male, female, transgender male-to-female, transgender female-to-male, other

Type and variable	Metrics and notes
Race	Self-report
Ethnicity	Hispanic/non-Hispanic/other as available /self-report
Sexual-orientation	Self-report
Mode of transmission	For example, men who have sex with men, intravenous drug use, high-risk heterosexual
Date of entry into care	MM/DD/YYYY
Date of death (if applicable)	MM/DD/YYYY

[a]CPT: Current Procedural Terminology.

Figure 1. Cascade monitoring study design.
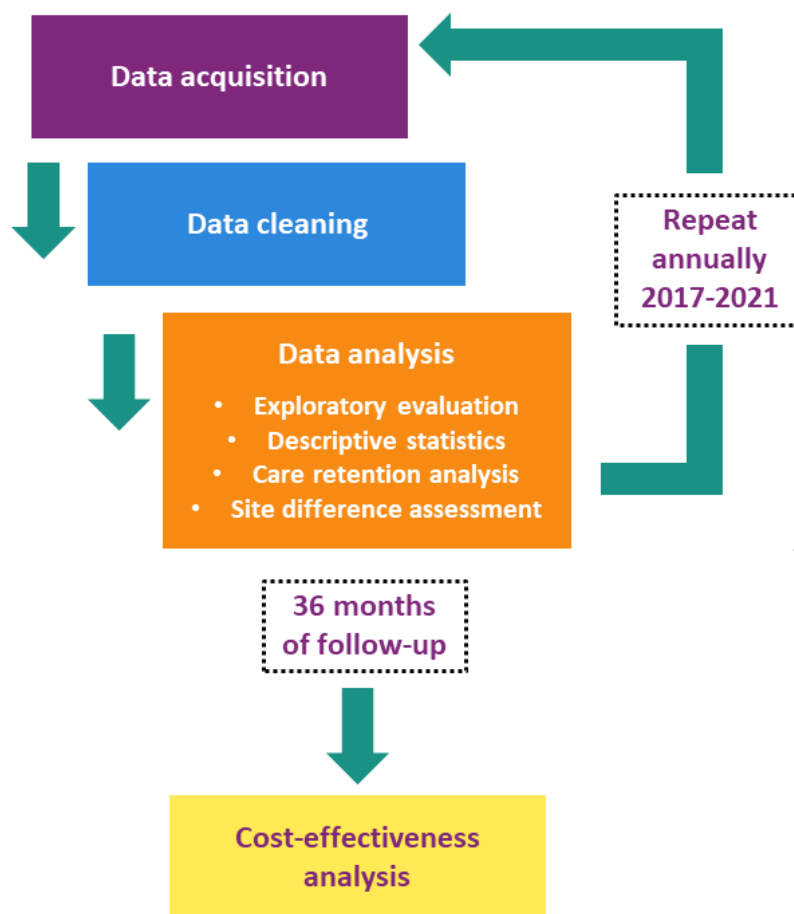
The SIU Recruitment and Enrollment Center (REC) is a subsidiary of the management core and works with the CM protocol team to communicate and work with SRVs to obtain EHR data and troubleshoot challenges. The REC issues launch materials including example templates to sites before download, specifying data elements, formats, and ordering. Data structure will likely vary by site and depend on the resources and capabilities of each site. The REC is equipped to process various data structures and formats, in this case.

The CM protocol requires data to be electronically downloaded, not hand-extracted. If a site cannot, at a minimum, download viral load, date of appointment, age, height, weight and ICD-10 diagnoses codes, then the CM protocol team will consider dropping the site. The site needs to discuss with their information technology team/department to determine if downloading is possible, and any associated cost. The CM protocol team will then determine whether it is feasible for the site to remain as part of CM and, if applicable, an acceptable date of data transfer.

EHR data on cost-effectiveness variables include ICD-10 diagnoses codes (extracted as left-hand justified character variables), current procedural terminology (version 4) procedure codes (extracted as 5-digit numerical variables), and health care common procedure coding system procedure codes (if used instead of CPT codes; extracted as 5-digit character variables). A data explanation form is used to document information relevant to the conduct of the study that is not captured on other study forms (eg, explanation for missing variables or explanation

for using a different metric than the one specified in the study summary). This form also includes a column for sites to indicate which method they used to obtain the data (ie, hand extraction for viral load only vs downloaded).

Data Transfer

Data transfer will occur annually as site data extracts are uploaded and processed. Dropbox Business software will be utilized to allow for the transfer of relevant data files from sites to the REC and from the REC to the CM protocol leads and analytic core (AC) that are encrypted and meet all standards for Health Insurance Portability and Accountability Act (HIPAA) compliance. The Dropbox Business program and its rigorous HIPAA-compliant system for managing the sharing of individual folders will allow for the development of permissions that restrict access to each file to site-specific personnel, protecting participant confidentiality. The system uses a combination of user authentication, file audit trails, device and user permissions, emergency access protocols, and login/logoff checks to manage the integrity of files. The Dropbox Business program encrypts each file uploaded onto the Dropbox servers and then, with the appropriate permissions via *keys*, allows users to download and unencrypt the file on their local machine. This means that all files stored on Dropbox are completely encrypted such that anyone without the appropriate *key* on their local machine will be unable to open the file after downloading it—this process can be done live through shared folders or through Web-based file delivery (similar to a file transfer protocol delivery). To upload, a user drags and drops files to the appropriate folder or selects them via the menu, just as they normally would. As Dropbox Business will be used for data storage and the software employs secure encryption upon upload and Secure Sockets Layer (SSL) for data transmission, device encryption is not required at the sites.

Regarding Dropbox Business and encryption, the key features of Dropbox Business (a HIPAA-compliant cloud encryption software) are as follows:

1. Dropbox employs file-level encryption on devices and the cloud with Advanced Encryption Standard 256-bit encryption and uses SSL for transmission.
2. Any file placed in Dropbox is automatically encrypted. As Dropbox synchronizes only encrypted versions of files, the data are consistently protected.

Data transfer will occur annually, as site data extracts are uploaded and processed.

Data Storage

Dropbox Business will be the primary method of data storage, and data files will be uploaded and downloaded securely by the REC, CM protocol lead and colead, and AC. Data files will not be saved on local machines under any circumstances. Regarding analyses, data will be analyzed by the CM protocol team and the AC using secure computer systems.

Data Management

The protocol lead and colead will provide data management support for the AC, working closely with the clinical sites and the management core. One of their roles is specifically

management of EHRs to develop a common data model across all 10 clinical sites.

The leads will ensure that analysis-ready datasets are regularly produced for the AC through ongoing data management. They will consult with the scientific team and with the AC to create a data management protocol, which details how finalized data will be structured. The data management protocol will be used in conjunction with the variable codebook to guide the data management process. Data will be exported from relevant sources and then imported into Statistical Analysis System (SAS Institute) and IBM Statistical Package for the Social Sciences (SPSS Inc). Using both SAS and SPSS, the leads will develop automated syntax files that will be used to clean data files by calculating scale scores, recoding variables, addressing missing data, and merging datasets together. Final datasets will be delivered to the AC in a preferred file format (SPSS, SAS, or CSV). Protocol leads will also provide ongoing support for cleaned data and perform additional data cleaning tasks as requested by the AC.

Data Analysis Plan

Upon initial receipt of the data and at regular intervals, exploratory data analysis will be conducted to understand the breadth and specificity of available data. Descriptive statistics, including measures of central tendency, will be generated to identify outliers and track data consistency for future downloads. For the ATN as a whole, and also for sites individually, the proportion of individuals in each key stage of the cascade (retained in care, prescribed ART, and virally suppressed) will be described at baseline and over time.

The AC will perform longitudinal data analyses with advanced analytic procedures such as mixed-effects regression models, generalized estimation equations with a Poisson distribution, and log link function to model care retention based on both patient-related and clinical characteristics to identify the relevant predictors associated with dropout at any stage of the cascade. Moreover, these models will be utilized to identify trends and patterns of change over the course of data collection. We will also conduct additional analyses to assess baseline differences among sites as a function of previous training or participation in ATN studies. On the basis of the findings, the AC will provide suggestions to the SIU leadership, the protocol leads, sites, and community partners to support ATN protocols and development going forward.

Cost-Effectiveness Analysis Plan

The AC will specify costs of implementation for budgeting further scale up of SIU interventions but will also identify the incremental benefit of SIU interventions on cascade outcomes over time. The cost-effectiveness analysis for the study is designed to measure costs and consequences of changes in the implementation over the 48 months of study follow-up. Furthermore, the goal is to help inform the investigators of the economic consequences of the varying amount of resources used in the exploration, preparation, implementation, and sustainability components of the study [19]. The data collected through the CM study and the resulting EHR phenotypes for YLH will be used to construct episode costs from large billing

databases that are relevant for modeling the cost-effectiveness of interventions aimed at YLH. The AC will use a previously developed cost utility model to estimate the cost per quality-adjusted life year over a 10-year time horizon expected from cascade outcomes of viral suppression and retention in care. As part of the identification of the larger cost-of-illness burden of cascade lapses for YLH, we will use the EHR phenotypes and archival data from Medicaid and/or privately insured populations to model the extent of cascade interruptions present in other practice settings. The data will be combined with the individual cost weights to estimate the variations in the economic burden that cascade disruptions for YLH place on US communities.

Results

The first year of data extraction for the 2016 data was launched on December 4, 2017, in 10 nationwide clinic sites. The REC sent launch emails to each site with instructions for data formatting and transfer, as well as created individual Dropbox Business accounts for each site to ensure secure transfer and storage of data. Sites varied in the length of time to obtain and download data (from 2 months to 6 months post study launch). All sites successfully uploaded data by May of 2018; all sites were asked to download the following: age as of January 1, 2016, appointment dates, viral load, ICD-10 diagnoses codes, height, and weight. Additional variables were downloaded if available for electronic extraction, such as STI test results, cholesterol panel, race, and CD4 counts. In total, 70% of the sites provided data for all mandatory variables via electronic download. The remaining sites will provide data for all mandatory variables via electronic and hand extraction in the future. These variables may include various lab tests, gender, race/ethnicity, and weight. Therefore, we will have complete information for all mandatory variables from all sites for future data submissions. Although a few sites were not able to electronically download demographic information and specific lab results, the protocol team worked with each site to develop a plan for hand extraction of those variables.

The second year of data extraction for 2017 data was launched during October of 2018. The deadline for site data submission is scheduled for December of 2018. Subsequent data extractions will take place yearly in October through 2021. Data for the years of 2016 through 2021 will be available for data analysis as requested by specific SIU protocol team.

These uploads mark significant progress toward protocol goals to examine trends in the HIV treatment cascade. Through preliminary evaluation of uploaded data and correspondence with sites through the SIU Website's Support Request Form system, the REC has learned valuable information that will facilitate the ATN's extended knowledge on the pulse of the epidemic.

Discussion

Summary of Key Innovations

This study has several strengths. One is its sample size—by capturing data from 10 different clinic sites across the nation,

it creates a robust dataset of YLH in the HIV treatment cascade. Moreover, the longitudinal design increases the amount of data available across the 10 sites and allows tracking youths' progress over time, including an understanding of those who enter or exit the system between 2016 and 2021.

The longitudinal data also allow the SIU scientific team to track the impact of SIU clinic-based interventions, including ATN 146, Tailored Motivational Interviewing Implementation Intervention [20]. The EHR is beneficial as it reduces secondary data entry error when creating data files shared between sites and SIU personnel and does not rely on youth self-report of health. This is especially crucial for YLH, who may not yet be familiar with the details and nuances of their health information. With EHR data, we are able to obtain data that are specific and verified and gathered relatively quickly in large quantities. With the development of the phenotypes, we will be able to extend our study findings to identify lapses in the treatment cascade for YLH using large national databases. This process will permit the enumeration of this patient phenotype nationwide and the description of related annual economic costs of care. This will enable us to identify meaningful variations between communities that can then serve as the basis for targeting interventions.

Finally, the method through which we receive data is also secure and safe for participants. All datasets are deidentified at the clinic sites thereby minimizing risks to breach of confidentiality. Using Dropbox Business further secures patient data, as the service encrypts all data and is HIPAA compliant.

Limitations and Conclusions

Despite significant benefits and innovation, study limitations exist. First, although there are 10 ATN clinic sites participating in this study, the majority of them are located in the Eastern part of the United States. There are only 2 sites west of Memphis and both (Los Angeles and San Diego) are in California. As a result, YLH who live in other parts of the United States, mainly the Midwest, Southwest, and West, are not included in this study. In addition, all clinics are situated in major cities. YLH living in smaller cities or in rural areas are also excluded. This potentially biases the sample, as YLH in nonurban areas experience different barriers to accessing and staying in care. For instance, those in urban areas may have greater access to public transportation and do not have to rely on cars or parking to visit a clinic. In addition, those in nonurban areas may have fewer clinics nearby, experience greater stigma from their community, or receive less structural support (in the form of policy or provider availability) for HIV-related care. Finally, the ATN clinics themselves are located within major academic research institutions. This suggests that the providers have more funding and support, as well as greater ability to adopt new evidence-based practices, than those without university support. As a result, the outcomes from this study will likely represent the best-case scenario for YLH in treatment. However, this limitation will be modified somewhat because of our plan to use the CM-developed EHR phenotypes to document cascade variations among communities represented by large archival databases, such as data from Medicaid and private insurers. Although such assessments will have errors and omissions at the level of individual patients, these analyses will provide

important data at the population level. Even with the extended modeling, our findings will still be limited by the fact that the YLH in our CM study and by extension in the large national databases already know their status and have entered the HIV treatment cascade. There are many YLH who are still unaware of their sero-status or may not have the means or ability to seek care. This particular study is only able to investigate YLH who have already engaged in care in some way.

Despite these limitations, this study has the capacity to provide a wealth of crucial data for improving our understanding of YLH and their engagement in the HIV treatment cascade.

Specifically, data from this study will indicate patterns of youths' engagement in care across 10 different clinic sites and the manner in which this population changes over the course of 6 years. This research has the ability to provide high-quality data on trends in the HIV treatment cascade; specifically, how many of those retained in care and receiving ART adhere to their treatment plan and achieve viral suppression. Likewise, it will be possible to determine who falls out of treatment, providing clinics and policymakers critical information for how to better support these communities. These data will also be crucial for measuring the effect MI has on YLH outcomes over time.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. Atlanta, GA: Division of HIV/AIDS Prevention, National Center for HIV/AIDS, Viral Hepatitis, Sexual Transmitted Diseases and Tuberculosis Prevention, Centers for Disease Control and Prevention; 2017. HIV Among Youth URL: https://www.cdc.gov/hiv/group/age/youth/index.html [accessed 2019-03-21] [WebCite Cache ID 772qDPO67]

2. Centers for Disease Control and Prevention. 2016. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2014 URL: https://www.cdc.gov/hiv/pdf/library/reports/surveillance/cdc-hiv-surveillance-supplemental-report-vol-21-4.pdf [accessed 2019-03-21] [WebCite Cache ID 772ttRgTT]

3. Gardner EM, McLees MP, Steiner JF, Del RC, Burman WJ. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. Clin Infect Dis 2011 Mar 15;52(6):793-800 [FREE Full text] [doi: 10.1093/cid/ciq243] [Medline: 21367734]

4. The White House Office of National AIDS Policy. 2010. National HIV/AIDS Strategy for the United States URL: https://obamawhitehouse.archives.gov/sites/default/files/uploads/NHAS.pdf [accessed 2019-03-21] [WebCite Cache ID 772uxsIXz]

5. National HIV/AIDS Strategy for the United States: Updated to 2020. Washington, DC: The White House Office of National AIDS Policy; 2015. URL: https://files.hiv.gov/s3fs-public/nhas-update.pdf [accessed 2019-03-21] [WebCite Cache ID 772v7u272]

6. Chen M, Rhodes PH, Hall IH, Kilmarx PH, Branson BM, Valleroy LA, Centers for Disease Control and Prevention. Prevalence of undiagnosed HIV infection among persons aged ≥13 years--National HIV Surveillance System, United States, 2005-2008. MMWR Morb Mortal Wkly Rep 2012 Jun 15;61(Suppl):57-64 [FREE Full text] [Medline: 22695465]

7. Gagliardo C, Murray M, Saiman L, Neu N. Initiation of antiretroviral therapy in youth with HIV: a US-based provider survey. AIDS Patient Care STDS 2013 Sep;27(9):498-502 [FREE Full text] [doi: 10.1089/apc.2013.0110] [Medline: 23937549]

8. Wheeler WH, Ziebell RA, Zabina H, Pieniazek D, Prejean J, Bodnar UR, Variant, Atypical, and Resistant HIV Surveillance Group. Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, US-2006. AIDS 2010 May 15;24(8):1203-1212. [doi: 10.1097/QAD.0b013e3283388742] [Medline: 20395786]

9. Zanoni BC, Mayer KH. The adolescent and young adult HIV cascade of care in the United States: exaggerated health disparities. AIDS Patient Care STDS 2014 Mar;28(3):128-135 [FREE Full text] [doi: 10.1089/apc.2013.0345] [Medline: 24601734]

10. MacPherson P, Munthali C, Ferguson J, Armstrong A, Kranzer K, Ferrand RA, et al. Service delivery interventions to improve adolescents' linkage, retention and adherence to antiretroviral therapy and HIV care. Trop Med Int Health 2015 Aug;20(8):1015-1032 [FREE Full text] [doi: 10.1111/tmi.12517] [Medline: 25877007]

11. National Institute of Nursing Research. Self-Management: Improving Quality of Life for Individuals with Chronic Illness URL: http://www.ninr.nih.gov/newsandinformation/iq/self-management-workshop [WebCite Cache ID 6ypNYd5fN]

XSL•FO

RenderX

12. Naar S, Parsons JT, Stanton BF. Adolescent Trials Network for HIV-AIDS Scale It Up Program: rational and protocol overviews. JMIR Res Protoc 2019 Feb 1;8(2):e11204 [FREE Full text] [doi: 10.2196/11204] [Medline: 30707102]

13. Berg RC, Ross MW, Tikkanen R. The effectiveness of MI4MSM: how useful is motivational interviewing as an HIV risk prevention program for men who have sex with men? A systematic review. AIDS Educ Prev 2011 Dec;23(6):533-549. [doi: 10.1521/aeap.2011.23.6.533] [Medline: 22201237]

14. Jensen CD, Cushing CC, Aylward BS, Craig JT, Sorell DM, Steele RG. Effectiveness of motivational interviewing interventions for adolescent substance use behavior change: a meta-analytic review. J Consult Clin Psychol 2011 Aug;79(4):433-440. [doi: 10.1037/a0023992] [Medline: 21728400]

15. Miller WR, Rose GS. Toward a theory of motivational interviewing. Am Psychol 2009 Sep;64(6):527-537 [FREE Full text] [doi: 10.1037/a0016830] [Medline: 19739882]

16. Naar-King S, Outlaw A, Green-Jones M, Wright K, Parsons JT. Motivational interviewing by peer outreach workers: a pilot randomized clinical trial to retain adolescents and young adults in HIV care. AIDS Care 2009 Jul;21(7):868-873. [doi: 10.1080/09540120802612824] [Medline: 20024744]

17. Outlaw AY, Naar-King S, Parsons JT, Green-Jones M, Janisse H, Secord E. Using motivational interviewing in HIV field outreach with young African American men who have sex with men: a randomized clinical trial. Am J Public Health 2010 Apr 1;100(Suppl 1):S146-S151. [doi: 10.2105/AJPH.2009.166991] [Medline: 20147689]

18. Parsons JT, Golub SA, Rosof E, Holder C. Motivational interviewing and cognitive-behavioral intervention to improve HIV medication adherence among hazardous drinkers: a randomized controlled trial. J Acquir Immune Defic Syndr 2007 Dec 1;46(4):443-450 [FREE Full text] [doi: 10.1097/QAI.0b013e318158a461] [Medline: 18077833]

19. Carcone AI, Coyle K, Gurung S, Cain D, Dilones R, Jadwin-Cakmak L, et al. Implementation science research examining the integration of evidence-based practices into HIV prevention and clinical care: protocol for a mixed-methods study using the Exploration, Preparation, Implementation, and Sustainment (EPIS) Model. JMIR Res Protoc 2019 (forthcoming). [doi: 10.2196/11202]

20. Naar S, MacDonnell K, Chapman JE, Todd L, Gurung S, Cain D, et al. Testing a motivational interviewing implementation intervention in adolescent HIV clinics: protocol for a type 3, hybrid implementation-effectiveness trial. JMIR Res Protoc 2019 (forthcoming). [doi: 10.2196/11200]

Abbreviations

AC: analytic core
ART: antiretroviral therapy
CM: cascade monitoring
CPT: Current Procedural Terminology
EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act
ICD-10: International Statistical Classification of Diseases and Related Health Problems-10th revision
MI: motivational interviewing
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis
REC: Recruitment and Enrollment Center
SIU: Scale It Up
SRV: subject recruitment venue
SSL: Secure Sockets Layer
STI: sexually transmitted infection
YLH: youth living with HIV

XSL•FO
RenderX