Protocol

Identifying Firearm Violence Exposure in Primary Care Clinical Notes: Protocol for Developing a National Language Processing Text Classifier

Natalie Carwright[1], PhD; Frances M Biel[2], MPH, MS; Megan Hoopes[2], MPH; Ali Al Bataineh[3], PhD; Pedro Rivera[2], MS; Kerry Bet[2], MPH; Nicole Cook[2], MPA, PhD

[1]Department of Mathematics, Norwich University, Northfield, VT, United States

[2]OCHIN, Portland, OR, United States

[3]David Crawford School of Engineering, Norwich University, Northfield, VT, United States

Corresponding Author:
Natalie Carwright, PhD
Department of Mathematics
Norwich University
158 Harmon Drive
Northfield, VT, 05663
United States
Phone: 1 802 485 2324
Email: ncartwri@norwich.edu

Abstract

Background: Structured data codes capture acute bodily injury from firearm violence but do not necessarily describe follow-up care from bodily injury and secondary exposure to firearm violence (eg, witnessing a shooting, being threatened by a firearm, or losing a loved one to gun violence and injury from firearms) even though such exposure is associated with many short- and long-term health impacts. Clinical notes from electronic health records (EHRs) often contain data not otherwise captured in structured data fields and can be categorized using natural language processing (NLP).

Objective: This study protocol outlines the steps being taken to develop an NLP text classifier for determination of exposure to firearm violence (both primary and secondary exposure) from ambulatory primary care and behavioral health EHR clinical notes for persons aged ≥5 years.

Methods: The study will use unstructured data from clinical notes taken between 2012 and 2022 from OCHIN, a multistate network of community health organizations using a single instance of Epic EHR. We describe the process of developing a labeled dataset for supervised NLP development that includes establishing a lexicon (words related to firearm violence) to identify potentially relevant notes, followed by a review of text extracted from a sample of these notes. We then describe the process of building, training, and evaluating candidate machine learning, neural network, and large language model NLP text classifiers. From this, a final NLP model is chosen then evaluated on a new set of randomly selected notes. An engaged stakeholder advisory committee will provide input and guidance on methods and results to identify and address potential biases in the NLP text classifiers.

Results: The study was funded in September 2023. Study activities have been ongoing through July 2025 and we are currently evaluating NLP text classifiers. We expect that the final model will be selected by August 2025 and we will publish results of NLP model development and the final model performance in 2026.

Conclusions: This work describes the development of a novel NLP text classifier to identify exposure to firearm violence in ambulatory primary care and behavioral health clinical notes. The NLP model developed in this study may lead to increased ascertainment of patients with exposure, laying the groundwork for understanding the long-term impacts and outcomes of firearm violence exposure and presenting opportunities for improved patient care.

International Registered Report Identifier (IRRID): DERR1-10.2196/76681

XSL•FO
RenderX

Introduction

Background

As injury and death by firearm continues to increase in the United States, so too does secondary exposure to firearm violence (witnessing a shooting, being threatened by a firearm, or the injury or death of a loved one or acquaintance by gun violence). It is estimated that 3 million US children are exposed to firearm violence each year and that 54% of all US adults have experienced firearm violence [1,2]. Primary and secondary exposure can lead to adverse social, emotional, physical, and behavioral health impacts [1,3-6]. Despite the high prevalence of exposure to firearm violence, structured data (eg, diagnostic codes) on firearm violence is largely limited to acute primary exposure (ie, direct injury or death) and there are no International Classification of Diseases (ICD) codes that specifically capture follow-up care from injury or exposure without injury. As a result, most firearm violence research is conducted using acute injury data from hospitals, emergency departments, or death data and is therefore limited to individuals who sustained acute physical injury [7].

Comprehensive data from ambulatory care settings (primary and behavioral health care) are needed to better understand the short- and long-term impacts associated with firearm violence exposure [7-9]. ICD-10 diagnosis codes now exist to document adverse socioeconomic and community circumstances, but these codes are used infrequently across health systems and there is no specific code documenting secondary exposure to gun violence [10,11]. Another opportunity to identify this exposure in clinical settings is through the development and use of custom data collection instruments, such as a trauma history "SmartForm" described in prior work [12]. While these documentation practices show promise, they remain inconsistently available and underused.

Another promising emerging source of data is electronic health record (EHR) clinical notes [13]. Clinical notes include relevant medical care notes that are needed for comprehensive, longitudinal patient health care. Accessing unstructured clinical data using advanced natural language processing (NLP) methods has demonstrated utility in similar types of firearm and trauma research, yet the application of these methods to outpatient settings remains largely unexplored [13-17]. Early work using a keyword search of firearm violence terms identified that exposure to firearm violence is available in unstructured data in ambulatory care settings [18]. This finding provided rationale for developing an NLP text classifier to identify cases of firearm violence exposure documented in clinical notes.

This protocol describes the development of an NLP text classifier to identify firearm violence exposure in primary care and behavioral health care clinical notes. Ambulatory clinics providing primary and behavioral health care offer a novel source of data on firearm violence, particularly exposures that may not have resulted in acute emergency care, hospital encounters, or death. The rich contextual information (eg, patient social history and comorbidities) contained in primary care and behavioral health clinical notes presents an opportunity to expand our understanding of the short- and long-term adverse effects of firearm violence exposure [19-21]. The addition of data from this novel source, combined with learnings from more established data sources on firearm injury and death, could greatly expand our understanding of risks and outcomes associated with exposure [7,22].

Objective

This research protocol describes the development of a stakeholder-informed NLP model to identify patients with exposure to firearm violence using clinical notes from a large, multistate network of ambulatory health care centers by classifying text as indicative as exposure to firearm violence or not.

Methods

Study Aims and Design

This protocol describes the development of a stakeholder-informed NLP text classifier to identify patients with firearm violence exposure documented in unstructured clinical notes during a primary care or behavioral health care encounter. The current study builds on pilot work that confirmed the presence of unstructured data in clinical notes that included firearm violence keywords that had contextual meaning of exposure [18]. Study findings, when available, will be reported following guidance for reporting machine learning model development in biomedical research [23].

Setting

The data used in this study come from OCHIN, a multistate network of community health organizations using a single instance of Epic EHR [24]. OCHIN is a nonprofit health care innovation center that offers a fully hosted, highly customized instance of Epic practice management and EHR solutions at over 2000 health care delivery sites across 40 states. OCHIN member health organizations provide health care services to patients in low-resource settings, regardless of their ability to pay. We will access clinical notes from unstructured records in natural language serving as a record of treatment, diagnoses, and clinical progress from ambulatory primary care and behavioral health care encounters to patients aged ≥5 years at encounter from January 1, 2012, to December 31, 2022. These chart notes will be used to develop an NLP model that will classify the note text as representative of firearm violence exposure or not.

Ethics Approval

OCHIN community health organizations, as part of their contract with OCHIN, authorize OCHIN to create limited datasets of member information for certain research activities consistent with applicable law and aligning with best practices of research, including institutional review board oversight of study activities.

This study was reviewed and deemed exempt by the Norwich University Institutional Review Board. Only research staff who are OCHIN employees have access to clinical notes, and efforts were made to remove direct patient identifiers prior to review.

Data Source

Data for NLP model development will be identified by searching for keyword phrases in clinical notes for study-eligible patient encounters. We define study-eligible patient encounters as those occurring between January 1, 2012, and December 31, 2022, with patients who were aged ≥5 years on their date of encounter (4,024,926 patients). This study period aligns with the date range of EHR data availability at the time of study start. The age restriction was imposed to reduce the number of "false positive" firearm exposure measures resulting from screening for household gun ownership commonly assessed in early childhood encounters [25]. We further restricted eligible ambulatory encounters to those categorized as either primary or behavioral health, using a previously defined algorithm (31,801,530 encounters) [26].
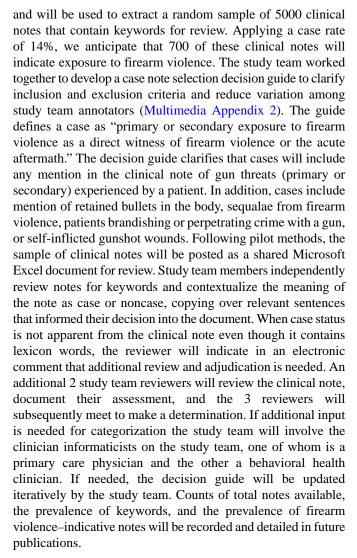
Stakeholder Advisory Committee

The NLP model development will be guided by an 8- to 10-person stakeholder advisory committee (SAC), a multidisciplinary group of individuals who will review methods, provide guidance on testing and evaluation, participate in interpreting and disseminating study findings, and evaluate the social and ethical merits for ongoing study and model deployment. SAC members include people with lived experience of firearm violence exposure, clinicians, researchers, clinical informaticists, and data scientists.

Development of the SAC will be conducted during the first 3 months of the study. SAC members will meet monthly, alternating between synchronous and asynchronous meeting formats, to guide study activities and will be compensated for their time (the mean compensation will be $185 per hour). Due to the dynamic nature of the research activities and the novel application of a SAC for NLP development around firearm violence, the ongoing contribution of SAC members after lexicon development and review (see below) will continue to evolve as study activities progress. We anticipate that SAC members will participate in interpreting research findings and provide ongoing input into the applicability of the model in clinical practice.

Dataset Development

During pilot work, a list of keywords informed by MacPhaul et al [13] was applied to a set of notes for patients with known exposure to firearm violence. A random sample of 30 notes from each of 3 text string search categories (broad, gun-only, and shooting) were reviewed to contextualize how keywords are used in clinical notes and to identify lexicon refinements (eg, remove preset template EHR phrases and add keywords[18]). Among the 90 notes reviewed, 13 (14%) were determined to be true "cases," while all other notes did not indicate exposure to or injury from firearm violence.

In this study, the lexicon (Multimedia Appendix 1) was expanded to include additional terms used by Chew et al [27]

and will be used to extract a random sample of 5000 clinical notes that contain keywords for review. Applying a case rate of 14%, we anticipate that 700 of these clinical notes will indicate exposure to firearm violence. The study team worked together to develop a case note selection decision guide to clarify inclusion and exclusion criteria and reduce variation among study team annotators (Multimedia Appendix 2). The guide defines a case as "primary or secondary exposure to firearm violence as a direct witness of firearm violence or the acute aftermath." The decision guide clarifies that cases will include any mention in the clinical note of gun threats (primary or secondary) experienced by a patient. In addition, cases include mention of retained bullets in the body, sequalae from firearm violence, patients brandishing or perpetrating crime with a gun, or self-inflicted gunshot wounds. Following pilot methods, the sample of clinical notes will be posted as a shared Microsoft Excel document for review. Study team members independently review notes for keywords and contextualize the meaning of the note as case or noncase, copying over relevant sentences that informed their decision into the document. When case status is not apparent from the clinical note even though it contains lexicon words, the reviewer will indicate in an electronic comment that additional review and adjudication is needed. An additional 2 study team reviewers will review the clinical note, document their assessment, and the 3 reviewers will subsequently meet to make a determination. If additional input is needed for categorization the study team will involve the clinician informaticists on the study team, one of whom is a primary care physician and the other a behavioral health clinician. If needed, the decision guide will be updated iteratively by the study team. Counts of total notes available, the prevalence of keywords, and the prevalence of firearm violence–indicative notes will be recorded and detailed in future publications.

The study team will also engage the SAC to review keywords and suggest additional terms. These terms will be evaluated for inclusion into the lexicon by performing a keyword search to identify potentially related notes, followed by a review of a sample (up to 30 per term, if available in notes) of these notes. Terms that identify at least one note indicative of firearm violence exposure will be added to the lexicon. Engaging stakeholders is an important approach for incorporating broad community perspectives in developing and applying NLP and other artificial intelligence in health care, reducing bias and improving applicability [28-30].

NLP Text Classifier Development

The adjudicated notes will become the labeled dataset for training and testing our NLP text classifier. The study team will use Python (Python Software Foundation) to build the NLP text classifier. The output of our classifier will result in a new structured data variable that identifies exposure to firearm violence from behavioral health or primary care encounters in ambulatory settings.

The NLP text classifier will be trained as follows:

1. The dataset, containing chart notes and the adjudicated label, will be randomly separated into training and test sets using an 80% to 20% train-test split, stratified by label.

2. The study team will use the training set to train several baseline machine learning (ML) and neural network (NN) classifiers available in the *scikit-learn* Python library [31] and Tensorflow platform [32] with Keras application programming interface [33], respectively, implemented with their default values.

- Baseline ML models will consist of a pipeline with a vectorizer and classifier. The vectorizer will be a term frequency inverse document frequency [34] implemented via *scikit-learn's* TfidfVectorizer. ML classifiers will include random forest, multinomial naïve Bayes, gradient boosting, and LightGBM, implemented using *scikit-learn* and the LightGBM library [35].

- Baseline NN models will use a tokenizer with a 1000-word vocabulary fit, and transformed sequences will be padded to 95% of the maximum sequence length. A sequential model consisting of an embedding layer with dimension 100, a model layer, and a dense layer with sigmoid activation will be compiled using an Adam optimizer, binary cross entropy loss, 50 epochs, a batch size of 64, and a validation split of 0.20. Model layers will be either recurrent neural network, long short-term memory, or gated recurrent unit with 64 units.

3. Performance of baseline models will be measured by evaluation on the held-out test set using metrics such as precision, recall, specificity, and the F_1-score (in future work, we would instead use cross-validation on the training set to assess baseline model performance to avoid data leakage).

4. The ML and NN baseline models with the higher evaluation metrics on the held-out test set will then be tuned for model hyperparameters, architecture (NN), learning parameters (NN), and threshold.

- Tuning of the ML model will use BayesSearchCV from the *scikit-optimize* library [31] using the vectorizer-classifier pipeline with cross validation and 60 iterations and scored using average precision. The model with the best parameters will be fit on the training and validation data and then the optimal threshold will be determined using sckikit-learn's TunedThresholdClassifierCV with cross validation and the F_1-score.

- Tuning of the NN model will use a text vectorization layer with tunable max_tokens and output_sequence_length; an embedding layer with tunable output_dim; a tunable number of NN layers each followed by a dropout layer with a tunable number of units, activation function, and dropout rate; and lastly, a dense layer with sigmoid activation. The model will be compiled using an Adam optimizer and binary cross entropy loss. The model will be fit with a tunable batch size. The training set will be split into training and validation sets and the BayesianOptimization tuner will be used to tune hyperparameters using 60 trials with 3 executions per trial, an objective of maximizing the average precision score, 100 epochs, and early

stopping with patience=2. The model with the best parameters will again be fit on all the training data for 100 epochs, and the epoch with the maximum average precision score on the validation split will be assigned as the optimal number of epochs. The model with the best parameters will be fit on the training and validation data using this optimal number of epochs. The optimal threshold will be determined using cross validation and the F_1-score.

5. An open-source large language model, clinicalBERT [36], will be fine-tuned using the training and validation sets and the Transformers [37] and Datasets [38] libraries from HuggingFace to produce another NLP text classifier. The learning rate, batch size, and number of epochs will be tuned manually, using initial values of 0.00005, 16, and 5, respectively.

6. Evaluation of the tuned models and the clinicalBERT model will be done by assessing evaluation metrics of model predictions on the held-out test set; assessing evaluation metrics of model predictions on the held-out test set across subpopulations defined by age, sex, and race or ethnicity, and the intersection of these subpopulations; reviewing and discussing the performance of tuned and clinicalBERT models with the SAC.

7. Documentation of model performance and suggestions for any preprocessing, inprocessing, or postprocessing methods that should be pursued to mitigate any observed biases in the predictions of tuned or clinicalBERT models on the test set will be carried out.

8. One of the tuned models or the clinicalBERT model will be selected as our final NLP model, using the criteria in step 6.

9. Evidence of the validity of the final NLP model will be collected. We will do this by applying a keyword search to extract text excerpts from 1,000,000 randomly selected novel notes not previously adjudicated or used. From preliminary data pulls, we estimate that approximately 3% (n=30,000) of notes will have a keyword. Our final NLP model will be applied to these text excerpts, and a random sample of these (n=1000) will be adjudicated to give evidence as to the validity of the model. Adjudication will follow the same process described in the Dataset Development section above.

Addressing Ethics, Equity, and Bias in Model Development

The study team is approaching development of the NLP text classifier with a lens toward minimizing biases that may emerge [39-41]. Our primary strategy to address bias in model development is through the engagement of a SAC throughout the study. SAC members will participate in discussions about the use of an NLP-based classifier in a clinical context. Discussions with SAC members on the implications of false positives and false negatives at both the patient level and research findings level will help guide how potential NLP models will be evaluated (ie, which metrics are of higher importance). The SAC will be asked to consider bias based on patient characteristics and provider practices as well as

algorithmic bias. Inclusion of SAC-suggested terms indicative of firearm violence and considerations of model performance across subpopulations are key evaluative characteristics for considering, identifying, and then mitigating potential biases in a universal firearm violence NLP algorithm. For additional diversity in input, our research team will meet with OCHIN's primary care and behavioral health clinical workgroups, attended by clinical leaders from member health organizations during the study start up (Year 1, quarter 1).

Potential biases in our training data include missing patients where firearm violence exposure is not present in a clinical note, possibly due to care received outside the health center, variation in provider care, or documentation practices, and patients who do not self-report exposure. These biases are inherent to all EHR-based research, and generalizability should be interpreted in light of the convenience-based, albeit large, nature of the dataset used [42,43] Our data come from primarily low-resource populations but it is not necessarily representative of all subgroups or individuals most likely to experience exposure to firearm violence. Likewise, we do not have reliable data on many variables in which model bias may exist (eg, gender).

There is the possibility of annotation bias during adjudication of text excerpts from clinical notes, even with our experienced team members. To minimize this bias, the study team developed a case note selection decision guide to support review by different study team members. We did not assess inter-rater reliability but would recommend this be performed in future projects.

We will attempt to identify potential model bias by calculating model performance across subpopulations of sex, age, race and ethnicity, and the intersections of these subpopulations (ie, group fairness) as these are protected attributes that are reliably documented in our data. We will provide this information to the SAC for further consideration along with discussions of stereotypical bias that can occur by evaluating models over such groups, under-representation bias due to low numbers in some subgroups, and possible mitigation strategies.

There may be other types of biases present in an NLP model that require more in-depth analysis to identify and mitigate. Such analyses include counterfactual evaluation (eg, replacing pronouns), adversarial testing, checking for bias in word embeddings, and evaluating model bias on an external dataset. However, due to the short duration of this study that spans lexicon development to the evaluation of NLP models, and without the intent to put this model in production at the end of this study, we leave these advanced and necessary bias identification methods to future work. If we do identify any biases in this study, we will document and provide transparency to these biases. We do not foresee time to effectively mitigate all potential biases within this study period. If there are identified biases and time left in our study period, we will consider and use appropriate mitigation strategies, which will most likely involve postprocessing mitigation techniques such as calibration and threshold adjustment given time constraints.

Results

The AIM-AHEAD Impact of Gun Violence Exposure on Health was funded on October 1, 2023, for a 2-year period.

NLP Model Development

Data acquisition started in February 2024 and was completed in April 2024, with >85 million primary care and behavioral health notes from >7 million patients. The study team extracted text from a random sample of 5000 notes with at least 1 firearm violence term for review and adjudication between March 2024 and May 2024. The first NLP template for baseline models was run in July 2024. Model development will continue throughout the study period to identify the best performing model for potential deployment. We expect that the final model will be selected by August 2025, and we will publish results of NLP model development and the final model performance in 2026.

Stakeholder Advisory Committee

The SAC was established during the first 3 months of the study in 2023 and includes 12 members: 5 community advocates or patients with lived experience; 1 physician; 2 clinician researchers; 2 clinical informaticists; and 2 data scientists with firearm violence data and research experience. The SAC will continue to meet monthly for the duration of the study. SAC contribution includes providing guidance on lexicon development, review of NLP performance, and application of NLP model in clinical practice [30].

Discussion

While exposure to firearm violence is common, identification of this experience is rare in structured ambulatory EHR data, particularly for secondary exposure that does not result in acute bodily injury. This study will be the first to develop an NLP text classifier to identify firearm violence exposure in clinical notes. We expect the NLP model developed in this study will be able to increase ascertainment of primary care and behavioral health patients with exposure to firearm violence; these findings will lay the groundwork for understanding the long-term impacts and outcomes of firearm violence exposure and present opportunities for improved patient care. Engagement with OCHIN network providers throughout the study will facilitate integration of results with clinical care and documentation practices. NLP has been used in other contexts to identify processes poorly captured by ICD codes; we expect that our work will also be able to identify patterns of exposure not otherwise captured in patient EHRs [44-46].

Because of the novelty of applying NLP to clinical notes for this purpose, we intentionally kept our approach open-ended and did not define a precise taxonomy or operational definition for firearm violence exposure, nor will we attempt to delineate between different types of exposure (eg, witnessing a community event versus direct personal threat). This is a limitation and presents important opportunities for future study and refinement. On the other hand, clinical notes are unlikely to represent a comprehensive picture of firearm violence exposure, and documentation is likely biased by provider practices and patient characteristics [43]. The value of identifying patients with

exposure to firearm violence is to alert providers to the potential health care and support needs of such patients, to expand our knowledge of health sequalae following firearm violence, and to more fully frame the public health burden of firearm violence. Future work would benefit from collating clinical note–based NLP models with other sources of firearm violence exposure information—such as ICD-10 codes, personal and family history, and trauma screening—documented within and across clinical settings.

Additional limitations include the exclusion of clinical notes from children younger than 5 years, which limits generalizability, and the lack of more recent data, which may yield differential documentation patterns in structured or unstructured data. These limitations may be addressed in future studies.

Ultimately, addressing the epidemic of gun violence in the US and understanding the long-term impacts of exposure will require more comprehensive data. The NLP model developed through this work represents one aspect of this unmet need [7-10].

Data Availability

Because chart notes contain protected health information, raw data for this study are not accessible to researchers outside of OCHIN. Model natural language processing code is available upon request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Firearm keywords.
[DOCX File , 20 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Case note selection decision guide.
[DOCX File , 21 KB-Multimedia Appendix 2]

References

1. Finkelhor D, Turner HA, Shattuck A, Hamby SL. Prevalence of childhood exposure to violence, crime, and abuse: results from the National Survey of Children's Exposure to Violence. JAMA Pediatr. Aug 2015;169(8):746-754. [doi: 10.1001/jamapediatrics.2015.0676] [Medline: 26121291]
2. Schumacher S, Kirzinger A, Presiado M, Valdes I, Brodie M. Americans' experiences with gun-related violence, injuries, and deaths. KFF. 2023. URL: https://www.kff.org/other/poll-finding/americans-experiences-with-gun-related-violence-injuries-and-deaths/ [accessed 2025-08-15]
3. Mitchell KJ, Jones LM, Turner HA, Beseler CL, Hamby S, Wade R. Understanding the impact of seeing gun violence and hearing gunshots in public places: findings from the Youth Firearm Risk and Safety Study. J Interpers Violence. Sep 2021;36(17-18):8835-8851. [doi: 10.1177/0886260519853393] [Medline: 31179801]
4. Turner HA, Mitchell KJ, Jones LM, Hamby S, Wade R, Beseler CL. Gun violence exposure and posttraumatic symptoms among children and youth. J Trauma Stress. Dec 2019;32(6):881-889. [doi: 10.1002/jts.22466] [Medline: 31833114]

5. Smith ME, Sharpe TL, Richardson J, Pahwa R, Smith D, DeVylder J. The impact of exposure to gun violence fatality on mental health outcomes in four urban U.S. settings. Soc Sci Med. Feb 2020;246:112587. [doi: 10.1016/j.socscimed.2019.112587] [Medline: 31958617]

6. Ranney M, Karb R, Ehrlich P, Bromwich K, Cunningham R, Beidas RS, et al. FACTS Consortium. What are the long-term consequences of youth exposure to firearm injury, and how do we prevent them? A scoping review. J Behav Med. Aug 2019;42(4):724-740. [FREE Full text] [doi: 10.1007/s10865-019-00035-2] [Medline: 31367937]

7. Cook N, Sills M. Tracking all injuries from firearms in the US. JAMA. Feb 14, 2023;329(6):514. [doi: 10.1001/jama.2022.21997] [Medline: 36786795]

8. Kaufman EJ, Delgado MK. The epidemiology of firearm injuries in the US: the need for comprehensive, real-time, actionable data. JAMA. Sep 27, 2022;328(12):1177-1178. [doi: 10.1001/jama.2022.16894] [Medline: 36166012]

9. Kaufman EJ, Delgado MK. Tracking all injuries from firearms in the US-reply. JAMA. Feb 14, 2023;329(6):514-515. [doi: 10.1001/jama.2022.22000] [Medline: 36786792]

10. Llamocca EN, Ahmedani BK, Lockhart E, Beck AL, Lynch FL, Negriff SL, et al. Use of codes for adverse social determinants of health across health systems. Psychiatr Serv. Jan 01, 2025;76(1):22-29. [doi: 10.1176/appi.ps.20240148] [Medline: 39308169]

11. ICD-coding of firearm injuries. National Center for Health Statistics. URL: https://www.cdc.gov/nchs/injury/ice/amsterdam1998/amsterdam1998_guncodes.htm [accessed 2025-08-15]

12. Cook N, Hoopes M, Biel FM, Cartwright N, Gordon M, Sills M. Early results of an initiative to assess exposure to firearm violence in ambulatory care: descriptive analysis of electronic health record data. JMIR Public Health Surveill. Feb 05, 2024;10:e47444. [FREE Full text] [doi: 10.2196/47444] [Medline: 38315521]

13. MacPhaul E, Zhou L, Mooney SJ, Azrael D, Bowen A, Rowhani-Rahbar A, et al. Classifying firearm injury intent in electronic hospital records using natural language processing. JAMA Netw Open. Apr 03, 2023;6(4):e235870. [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.5870] [Medline: 37022685]

14. Goldstein EV, Mooney SJ, Takagi-Stewart J, Agnew BF, Morgan ER, Haviland MJ, et al. Characterizing female firearm suicide circumstances: a natural language processing and machine learning approach. Am J Prev Med. Aug 2023;65(2):278-285. [doi: 10.1016/j.amepre.2023.01.030] [Medline: 36931986]

15. Zafari H, Kosowan L, Zulkernine F, Signer A. Diagnosing post-traumatic stress disorder using electronic medical record data. Health Informatics J. 2021;27(4):14604582211053259. [FREE Full text] [doi: 10.1177/14604582211053259] [Medline: 34818936]

16. Brandt C, Workman TE, Farmer M, Akgün KM, Abel E, Skanderson M, et al. Documentation of screening for firearm access by healthcare providers in the veterans healthcare system: a retrospective study. West J Emerg Med. May 19, 2021;22(3):525-532. [FREE Full text] [doi: 10.5811/westjem.2021.4.51203] [Medline: 34125022]

17. Trujeque J, Dudley R, Mesfin N, Ingraham NE, Ortiz I, Bangerter A, et al. Comparison of six natural language processing approaches to assessing firearm access in Veterans Health Administration electronic health records. J Am Med Inform Assoc. Jan 01, 2025;32(1):113-118. [doi: 10.1093/jamia/ocae169] [Medline: 39530748]

18. Cook N, Biel F, Cartwright N, Hoopes M, Al Bataineh A, Rivera P. Assessing the use of unstructured electronic health record data to identify exposure to firearm violence. JAMIA Open. Dec 2024;7(4):ooae120. [doi: 10.1093/jamiaopen/ooae120] [Medline: 39498385]

19. DeVoe JE, Gold R, Cottrell E, Bauer V, Brickman A, Puro J, et al. The ADVANCE network: accelerating data value across a national community health center network. J Am Med Inform Assoc. 2014;21(4):591-595. [FREE Full text] [doi: 10.1136/amiajnl-2014-002744] [Medline: 24821740]

20. Gold R, Kaufmann J, Cottrell E, Bunce A, Sheppler CR, Hoopes M, et al. Implementation support for a social risk Screening and referral process in community health centers. NEJM Catal Innov Care Deliv. Apr 2023;4(4):10.1056/CAT.23.0034. [FREE Full text] [doi: 10.1056/CAT.23.0034] [Medline: 37153938]

21. Gruß I, Bunce A, Davis J, Dambrun K, Cottrell E, Gold R. Initiating and implementing social determinants of health data collection in community health centers. Popul Health Manag. Feb 2021;24(1):52-58. [FREE Full text] [doi: 10.1089/pop.2019.0205] [Medline: 32119804]

22. Fontanarosa PB, Bibbins-Domingo K. The unrelenting epidemic of firearm violence. JAMA. Sep 27, 2022;328(12):1201-1203. [doi: 10.1001/jama.2022.17293] [Medline: 36166046]

23. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Internet Res. Dec 16, 2016;18(12):e323. [FREE Full text] [doi: 10.2196/jmir.5870] [Medline: 27986644]

24. Ochin. URL: https://ochin.org/members [accessed 2025-08-15]

25. Dowd M, Sege R, Council on Injury, Violence,Poison Prevention Executive Committee, American Academy of Pediatrics. Firearm-related injuries affecting the pediatric population. Pediatrics. Nov 2012;130(5):e1416-e1423. [doi: 10.1542/peds.2012-2481] [Medline: 23080412]

26. Cook N, McGrath BM, Navale SM, Koroukian SM, Templeton AR, Crocker LC, et al. Care delivery in community health centers before, during, and after the COVID-19 pandemic (2019-2022). J Am Board Fam Med. Jan 05, 2024;36(6):916-926. [FREE Full text] [doi: 10.3122/jabfm.2023.230081R1] [Medline: 37857445]

27. Chew RF, Weitzel KJ, Baumgartner P, Oppenheimer WC, Liu S, Miller AB, et al. Improving text classification with Boolean retrieval for rare categories: a case study identifying firearm violence conversations in the Crisis Text Line database. RTI Press. URL: https://www.rti.org/rti-press-publication/improving-text-classification-boolean-retrieval-rare-categories-case-study-identifying-firearm-viole [accessed 2025-08-15]

28. Vishwanatha JK, Christian A, Sambamoorthi U, Thompson EL, Stinson K, Syed TA. Community perspectives on AI/ML and health equity: AIM-AHEAD nationwide stakeholder listening sessions. PLOS Digit Health. Jun 2023;2(6):e0000288. [FREE Full text] [doi: 10.1371/journal.pdig.0000288] [Medline: 37390116]

29. Wan R, Kim J, Kang D. Everyone's voice matters: quantifying annotation disagreement using demographic information. Proc AAAI Conf Artif. Jun 26, 2023;37(12):14523-14530. [doi: 10.1609/aaai.v37i12.26698]

30. Cook N, Biel FM, Bet KA, Sills MR, Al Bataineh A, Rivera P, et al. Engaging stakeholders with professional or lived experience to improve firearm violence lexicon development. JMIR Form Res. Apr 21, 2025;9:e68105. [FREE Full text] [doi: 10.2196/68105] [Medline: 40258170]

31. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. 2011. URL: http://jmlr.org/papers/v12/pedregosa11a.html [accessed 2025-08-15]

32. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. Zheng X. TensorFlow: large-scale machine learning on heterogeneous distributed systems. ArXiv. Preprint posted online on March 16, 2016. [FREE Full text]

33. Keras. GitHub. 2015. URL: https://github.com/fchollet/keras [accessed 2025-08-15]

34. Sparck Jones K. A statistical interpretation of term specificity and its application in retrieval. J Doc. Jan 1972;28(1):11-21. [doi: 10.1108/eb026526]

35. LightGBM R-package. lightgmb. URL: https://lightgbm.readthedocs.io/en/latest/R/index.html [accessed 2025-08-15]

36. Huang K, Altosaar J, Ranganath R. Clinicalbert: modeling clinical notes and predicting hospital readmission. ArXiv. Preprint posted online on April 10, 2019. [FREE Full text]

37. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: state-of-the-art natural language processing. 2020. Presented at: Conference on Empirical Methods in Natural Language Processing; November 16-20, 2020; Online.

38. Lhoest Q, Villanova del Moral A, Jernite Y, Thakur A, van Platen P, Patil S, et al. Datasets: a community library for natural language processing. 2021. Presented at: Conference on Empirical Methods in Natural Language Processing; November 7-11, 2021; Punta Cana, Dominican Republic. [doi: 10.4135/9781483368924.n365]

39. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - addressing ethical challenges. N Engl J Med. Mar 15, 2018;378(11):981-983. [FREE Full text] [doi: 10.1056/NEJMp1714229] [Medline: 29539284]

40. Chen Y, Clayton EW, Novak LL, Anders S, Malin B. Human-centered design to address biases in artificial intelligence. J Med Internet Res. Mar 24, 2023;25:e43251. [FREE Full text] [doi: 10.2196/43251] [Medline: 36961506]

41. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med. Sep 2019;25(9):1337-1340. [doi: 10.1038/s41591-019-0548-6] [Medline: 31427808]

42. Beach MC, Saha S, Park J, Taylor J, Drew P, Plank E, et al. Testimonial injustice: linguistic bias in the medical records of Black patients and women. J Gen Intern Med. Mar 22, 2021;36(6):1708-1714. [doi: 10.1007/S11606-021-06682-Z]

43. Perets O, Stagno E, Yehuda E, McNichol M, Anthony Celi L, Rappoport N, et al. Inherent bias in electronic health records: a scoping review of sources of bias. MedRXiv. Preprint posted online on April 12, 2024. [FREE Full text] [doi: 10.1101/2024.04.09.24305594] [Medline: 38680842]

44. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform. Apr 27, 2019;7(2):e12239. [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]

45. Ridgway JP, Uvin A, Schmitt J, Oliwa T, Almirol E, Devlin S, et al. Natural language processing of clinical notes to identify mental illness and substance use among people living with HIV: retrospective cohort study. JMIR Med Inform. Mar 10, 2021;9(3):e23456. [FREE Full text] [doi: 10.2196/23456] [Medline: 33688848]

46. Chae S, Song J, Ojo M, Topaz M. Identifying heart failure symptoms and poor self-management in home healthcare: a natural language processing study. Stud Health Technol Inform. Dec 15, 2021;284:15-19. [doi: 10.3233/SHTI210653] [Medline: 34920459]

Abbreviations

ADVANCE: Accelerating Data Value Across a National Community Health Center
EHR: electronic health record
ICD: International Classification of Diseases
ML: machine learning
NLP: natural language processing
NN: neural network
SAC: stakeholder advisory committee

XSL•FO
RenderX