

Protocol

Improving the Reporting Quality of Studies on Information Extraction From Clinical Texts: Protocol for the Development of a Consensus-Based Reporting Guideline

Daniel Reichenpfader^{1,2}, MSc; Henning Müller^{3,4}, Prof Dr; Kerstin Denecke¹, Prof Dr

¹Institute for Patient-Centered Digital Health, Department of Engineering and Computer Science, Bern University of Applied Sciences, Biel, Switzerland

²PhD School of Life Sciences, Faculty of Medicine, University of Geneva, Geneva, Switzerland

³Department of Radiology and Medical Informatics, University of Geneva, Geneva, Switzerland

⁴Informatics Institute, HES-SO Valais-Wallis, Sierre, Switzerland

Corresponding Author:

Daniel Reichenpfader, MSc

Institute for Patient-Centered Digital Health, Department of Engineering and Computer Science

Bern University of Applied Sciences

Quellgasse 21

Biel 2502

Switzerland

Phone: 41 31 848 60 93

Email: daniel.reichenpfader@bfh.ch

Abstract

Background: Information extraction (IE) from clinical texts is increasingly important in health care; yet, reporting practices remain inconsistent. Existing guidelines do not fully address the unique challenges of IE studies. IE methods vary widely in their design, ranging from rule-based systems to advanced large language models, contributing to heterogeneity in reporting. While several reporting frameworks exist for applications of artificial intelligence in health care, they primarily focus on prediction modeling or clinical trials and associated protocols rather than text-based IE.

Objective: This study aims to develop the Clinical Information Extraction (CINEX) guideline, a consensus-based reporting guideline for studies on clinical IE.

Methods: The CINEX guideline is developed following an established guideline methodology, including a 3-round electronic Delphi (eDelphi) study with domain experts and a final in-person consensus meeting. The eDelphi process includes feedback loops and predefined consensus thresholds, with items rated on a 10-point scale for both relevance and maturity. The final consensus meeting is held as a hybrid workshop at the MEDINFO 2025 conference and focuses on finalizing the items that reached consensus.

Results: Our results will provide a validated reporting guideline for studies on clinical IE. A preliminary set of 28 reporting items was drafted from a scoping review and existing frameworks. The draft guidelines include 5 key dimensions: information model, architecture, data, annotation, and outcome. This draft guideline will be refined through the eDelphi process. It is designed to be technology-agnostic and applicable across diverse IE approaches, including not only large language models but also traditional machine learning methods and rule-based and hybrid systems.

Conclusions: The CINEX guideline provides structured, expert-validated guidance for reporting clinical IE studies, improving transparency, reproducibility, and comparability. The final guideline will be disseminated alongside an explanatory document to support adoption and implementation.

International Registered Report Identifier (IRRID): PRR1-10.2196/76776

JMIR Res Protoc 2025;14:e76776; doi: [10.2196/76776](https://doi.org/10.2196/76776)

Keywords: reporting guideline; information extraction; artificial intelligence; natural language processing; consensus-based approach; eDelphi study

Introduction

Information extraction (IE) refers to techniques that automatically identify and structure key information from unstructured text. These methods enable efficient reuse of clinical narratives, supporting decision-making, research, and automation in electronic health records. Specific use cases include phenotyping (eg, extraction of specific diseases), drug-related tasks (eg, dosage IE), and clinical workflow optimization (eg, adverse event detection) [1]. Furthermore, IE can reduce the burden of manual chart review, enable large-scale epidemiological studies, and support real-time decision support.

Recent advances in natural language processing (NLP), especially through large language models (LLMs), have improved IE capabilities. However, this rapid technical evolution has introduced fragmentation in methods, terminology, and evaluation. We identified this heterogeneity during a scoping review of studies describing IE specifically from radiology reports. Studies vary widely in defining the target information, annotating reference standards, evaluating system performance, and disclosing implementation details [2]. Without standardized reporting, it becomes difficult to interpret results, compare systems, replicate experiments, or translate the developed effective algorithms into clinical practice. This suboptimal reporting quality had already been identified in a systematic review conducted by Davidson et al [3] in 2021.

To address reporting variability in artificial intelligence (AI) studies, several guidelines have emerged. The CONSORT-AI (Consolidated Standards of Reporting Trials–Artificial Intelligence) extension provides a checklist for reporting clinical trials that include AI-based interventions [4]. Its counterpart, SPIRIT-AI (Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence), focuses on protocol reporting for such trials, ensuring clarity and completeness before trial execution [5]. More recently, TRIPOD+AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis–Artificial Intelligence) [6] and TRIPOD-LLM (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis specifically tailored for large language models) [7] have extended the original TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis) [8] guidelines to include the application of regression and machine learning methods as well as LLM-based approaches for prediction model studies. These guidelines promote transparency and reproducibility for studies describing AI-based interventions and prediction models. However, none are applicable to the reporting needs of IE from clinical text, which involves distinct tasks, input-output structures, and evaluation methods as compared to diagnostic modeling or intervention and their associated trials.

This paper presents the protocol for the development of the Clinical Information Extraction (CINEX) guideline, a consensus-based reporting guideline for clinical IE studies.

This protocol outlines the methodology behind the CINEX guideline, including a 3-round modified electronic Delphi (eDelphi) process with domain experts. The development of this guideline follows best practices for health research reporting and includes steps for item refinement, a final in-person consensus meeting, and the publication of an explanation and elaboration (E&E) document. The CINEX guideline aims to close the reporting gap in clinical IE research, promoting rigor, transparency, and harmonization across diverse methodological paradigms.

Methods

Overview and Preliminary Guideline Development

The planned development process of the CINEX guideline was registered and publicly listed with the EQUATOR (Enhancing the Quality and Transparency Of Health Research) Network as a reporting guideline under development in August 2024 prior to its official start in October 2024 [9]. The EQUATOR Network is an international initiative that promotes transparent and accurate reporting of health research to improve the reliability and value of the published literature [10]. The CINEX guideline follows the methodological guidance outlined by Moher et al [11] for developing health research reporting guidelines. Throughout this protocol, references to Moher's checklist items are indicated in angle brackets (<>).

The need for this guideline was identified during a scoping review of studies on LLM-based IE from radiology reports <1> [2]. The review aimed to assess the current state of the art in terms of performance, training and modeling approaches, clinical use cases, datasets, annotation methods, and commonly reported challenges. Derived from the review's aims, a data extraction table was drafted and populated by one author. The scoping review revealed important barriers to study comparability, including untransparent performance metric calculations, lack of external validation, and limited availability of source code. These problems hinder transparency and reproducibility in the field. Based on the data extraction table as well as inspired by existing reporting guidelines, an initial set of 28 candidate reporting items was drafted and published <2> [12].

eDelphi Study

To refine the preliminary guideline, a 3-round eDelphi study was planned between May and July 2025 <5,6>. The Delphi method was chosen as a structured, iterative process to achieve expert consensus, in line with the recommendations of Moher et al [11] for reporting guideline development.

We designed the eDelphi process based on the methodological principles of Häder [13], Nasa et al [14], as well as Trevelyan and Robinson [15]. We aim to recruit between 20 and 30 participants <4>. Minimum response rates are set at 30% for the first round and 70% for the following rounds, consistent with Häder's [13] guidance.

Eligible participants include the authors of studies identified in the preceding scoping review, as well as domain experts with regard to clinical IE, recruited through the personal and professional networks of the executive committee. Interested individuals will be invited via email to join the study. The survey itself will be conducted using the open-source tool LimeSurvey [16].

The eDelphi study will comprise exactly 3 rounds [13]. Consensus for the inclusion of an item in the guideline will be defined as a mean rating of 8 or higher on a 10-point scale, with an SD of 2 or less. Exclusion will be defined as a mean rating of 3 or lower, also with an SD of ≤ 2 . Items falling outside these thresholds will be revisited during a final consensus meeting held after the eDelphi process. Response stability is recorded but not used to assess early termination. Only in the first round, panelists have the possibility to add additional items for each domain. Exclusion as well as finalization due to consensus of items is only conducted after rounds 2 and 3. Rephrasing of items and descriptions is conducted after each round based on the panelists' comments.

In the first round, participants will be presented with the draft reporting items, including any proposed value sets where applicable. They will be asked to assess the relevance and maturity of each item using a 10-point scale with labeled endpoints only ("not at all relevant" and "not at all mature" to "very relevant" and "very mature"). Each item additionally includes a default "No answer" option. To approximate interval-level data, a larger number of anchor points was selected instead of a classical 5-point scale based on methodological recommendations [17]. To exclude or include (finalize) an item, the abovementioned mean and SD thresholds must be met for both relevance and maturity.

Feedback will be proactively provided by the executive committee for each new round (active feedback loop): participants re-evaluate items that were not excluded, assisted by aggregated ratings of the whole panel (as histograms), moderated anonymized comments from the previous round, and their own previous ratings. This format enables reflection and informed re-evaluation.

Besides participating in the eDelphi panel, each expert completes a self-disclosure of their expertise in clinical IE. Consistent with Häder [13], no demographic data will be collected beyond gender and country of affiliation. Anonymity among participants will be preserved throughout the eDelphi panel. At the end of the first round, participants will be invited to indicate whether they wish to participate in the final consensus meeting taking place after the completion of the third round and to contribute to the final publication of the CINEX guideline and the explanatory document.

A pretest of the first-round survey will be conducted among the executive committee (comprising authors DR, HM, and KD) to ensure clarity and functionality. These individuals will not participate in the actual eDelphi rounds.

Consensus Meeting, Finalization, and Outlook

For the final step of the guideline development process, a face-to-face hybrid consensus meeting is conducted, organized as a workshop at the MEDINFO conference in August 2025 <6,7,8>. During the workshop, the results of the Delphi process are presented, and items that achieved consensus are reviewed for resolving minor ambiguities and final phrasing. Items without prior agreement are discussed and, if needed, resolved by open voting. Consensus is defined as $\geq 80\%$ agreement. The meeting is considered quorate if $\geq 50\%$ of the panelists of the first round of the eDelphi panel are present. Persistent disagreements will be documented and reported transparently in the publication of the final guideline. The workshop shall result in the formal finalization of the CINEX guideline; thereafter, the guidance statement is finalized <9>, complemented by an E&E document <10>, and both are made available as publication <11>. The guideline development process will be reported in a structured way in accordance with the Accurate Consensus Reporting Document (ACCORD) reporting guideline [18].

Responsibilities for postpublication activities <12-18> are divided among the contributing authors: the CINEX guideline will be hosted on an open access online platform for ongoing feedback <12,14,16>, will be pilot-tested by panelists, and its impact on reporting quality will be evaluated. To ensure durability, an executive group will review the CINEX guideline at regular 3-year intervals, evaluating its impact <15> and issuing updates when new evidence or methods make a revision necessary <18>. We seek endorsement from journals and societies <13>. Translation is not planned for the initial release <17>. By following these postpublication activities, we aim at establishing a long-term relevance of the CINEX guideline.

Ethical Considerations

This study was granted an exemption by the competent ethics committee of the canton of Bern on April 29, 2025 (ID: Req-2025-00587). All eDelphi study participants gave electronic informed consent prior to participation.

Results

A preliminary guideline comprising 28 items and 5 dimensions (information model, architecture, data, annotation, and outcome) was published in August 2024 [12]. These items will be presented in the first round of the eDelphi study.

Discussion

Principal Results

This protocol presents the development of the CINEX guideline, a reporting guideline specifically developed for clinical IE studies. The CINEX guideline addresses key challenges in the field—including inconsistent terminology, reporting, and evaluation—by providing a structured,

technology-agnostic framework tailored to IE tasks in health care. By design, the CINEX guideline accommodates both rule-based and data-driven approaches, including LLM-based methods by focusing on essential reporting dimensions (eg, data sources, annotation strategies, information model, and evaluation metrics) that apply across paradigms. This ensures that differences in implementation, such as rule-based extraction pipelines or end-to-end architectures, can be transparently described and compared within a shared reporting structure. To facilitate adoption, we will pursue parallel endorsement of the CINEX guideline by multiple journals in biomedical informatics, promote its use by providing an online website (inspired by TRIPOD-LLM [7]), and invite panelists to apply the CINEX guideline in their own research.

Limitations

Our approach has several limitations. First, the development process for the CINEX guideline is based on guidance for reporting guideline development in health research as outlined by Moher et al [11]. While the CINEX guideline addresses health-related research, it also intersects significantly with the computer science domain. Currently, however, there are no methodological frameworks for reporting guideline development in computer science. Given this gap, we adopted the framework of Moher et al [11], which is well-established and has informed the development of most reporting guidelines endorsed by the EQUATOR Network, including guidelines with a technical focus (eg, CONSORT-AI). Nonetheless, we acknowledge that this choice may not fully capture disciplinary nuances outside the health sciences. The CINEX guideline will therefore be developed with input from the computer sciences as well as clinical disciplines and be tested for clarity and relevance in both contexts.

Participant recruitment relies partly on professional networks, which may limit diversity. While consensus thresholds and item rating methods are clearly defined a priori

with this protocol, they remain subjective to a certain degree. The final in-person consensus meeting, although valuable for discussion, may introduce social influence and new opinions from participants that have not participated in the eDelphi panel. Additionally, the CINEX may require future updates to accommodate rapid developments in LLMs and multimodal approaches.

Comparison With Prior Work

The CINEX guideline builds on the methods used in guidelines like CONSORT-AI, SPIRIT-AI, and TRIPOD-LLM and is the first to focus specifically on clinical IE. Our study design incorporated detailed feedback loops and a 10-point scale instead of a 5-point scale aiming to improve rating precision. While existing AI guidelines address broader study types, none are tailored to the unique needs of clinical text-based IE, underscoring the CINEX guideline's distinct contribution. Furthermore, we seek to address the challenges faced by prior reporting guidelines: first, adoption and enforcement by journals have often been inconsistent, thus limiting their impact; to mitigate this, we will accompany the CINEX guideline with an online template checklist for authors and actively engage journals to encourage endorsement. Second, ambiguity in checklist items has been a barrier to compliance; therefore, the CINEX guideline will be supported by a detailed E&E document. Third, the CINEX guideline will include example items tailored for computer science as well as medical audiences, supporting broader adoption across disciplines.

Conclusions

The CINEX guideline fills a critical gap in the reporting of clinical IE studies by offering a dedicated, consensus-driven framework. It aims to improve reproducibility, comparability, and transparency in this growing area of clinical NLP. Continued community involvement and iterative updates will be essential to ensure its ongoing relevance.

Acknowledgments

The CINEX (Clinical Information Extraction) guideline is part of a PhD project at the University of Geneva. No dedicated funding was obtained. Generative AI (ChatGPT [GPT-4o and GPT-5] OpenAI) was used to assist with language editing of the manuscript draft and improve clarity and formatting of responses to peer review comments. No artificial intelligence tool was used to generate original scientific content. All content was reviewed and approved by the authors.

Authors' Contributions

Conceptualization: DR and KD

Methodology: DR

Supervision: HM and KD

Writing—original draft: DR

Writing—review and editing: DR, HM, and KD

Conflicts of Interest

None declared.

References

1. Wang Y, Wang L, Rastegar-Mojarad M, et al. Clinical information extraction applications: A literature review. *J Biomed Inform.* Jan 2018;77:34-49. [doi: [10.1016/j.jbi.2017.11.011](https://doi.org/10.1016/j.jbi.2017.11.011)] [Medline: [29162496](https://pubmed.ncbi.nlm.nih.gov/29162496/)]

2. Reichenpfader D, Müller H, Denecke K. A scoping review of large language model based approaches for information extraction from radiology reports. *NPJ Digit Med*. Aug 24, 2024;7(1):222. [doi: [10.1038/s41746-024-01219-0](https://doi.org/10.1038/s41746-024-01219-0)] [Medline: [39182008](https://pubmed.ncbi.nlm.nih.gov/39182008/)]
3. Davidson EM, Poon MTC, Casey A, et al. The reporting quality of natural language processing studies: systematic review of studies of radiology reports. *BMC Med Imaging*. Oct 2, 2021;21(1):142. [doi: [10.1186/s12880-021-00671-8](https://doi.org/10.1186/s12880-021-00671-8)] [Medline: [34600486](https://pubmed.ncbi.nlm.nih.gov/34600486/)]
4. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. Sep 2020;26(9):1364-1374. [doi: [10.1038/s41591-020-1034-x](https://doi.org/10.1038/s41591-020-1034-x)] [Medline: [32908283](https://pubmed.ncbi.nlm.nih.gov/32908283/)]
5. Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*. Sep 2020;26(9):1351-1363. [doi: [10.1038/s41591-020-1037-7](https://doi.org/10.1038/s41591-020-1037-7)] [Medline: [32908284](https://pubmed.ncbi.nlm.nih.gov/32908284/)]
6. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. Apr 16, 2024;385:e078378. [doi: [10.1136/bmj-2023-078378](https://doi.org/10.1136/bmj-2023-078378)] [Medline: [38626948](https://pubmed.ncbi.nlm.nih.gov/38626948/)]
7. Gallifant J, Afshar M, Ameen S, et al. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat Med*. Jan 2025;31(1):60-69. [doi: [10.1038/s41591-024-03425-5](https://doi.org/10.1038/s41591-024-03425-5)] [Medline: [39779929](https://pubmed.ncbi.nlm.nih.gov/39779929/)]
8. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med*. Jan 6, 2015;13(1):1. [doi: [10.1186/s12916-014-0241-z](https://doi.org/10.1186/s12916-014-0241-z)] [Medline: [25563062](https://pubmed.ncbi.nlm.nih.gov/25563062/)]
9. Reporting guidelines under development for other study designs. The EQUATOR Network. 2024. URL: <https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-other-study-designs/#CINEX> [Accessed 2025-09-18]
10. About us. The EQUATOR Network. 2025. URL: <https://www.equator-network.org/about-us/> [Accessed 2025-09-18]
11. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. Feb 16, 2010;7(2):e1000217. [doi: [10.1371/journal.pmed.1000217](https://doi.org/10.1371/journal.pmed.1000217)] [Medline: [20169112](https://pubmed.ncbi.nlm.nih.gov/20169112/)]
12. Reichenpfader D, Denecke K. Towards a reporting guideline for studies on information extraction from clinical texts. In: Mantas J, Hasman A, Demiris G, Saranto K, Marschollek M, Arvanitis TN, Ognjanović I, Benis A, Gallos P, Zoulias E, Andrikopoulou E, editors. *Stud Health Technol Inform*. IOS Press; 2024. [doi: [10.3233/SHTI240744](https://doi.org/10.3233/SHTI240744)] ISBN: 978-1-64368-533-5
13. Häder M. *Delphi-Befragungen: Ein Arbeitsbuch*. Springer Fachmedien; 2014. [doi: [10.1007/978-3-658-01928-0](https://doi.org/10.1007/978-3-658-01928-0)] ISBN: 978-3-658-01927-3
14. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: How to decide its appropriateness. *World J Methodol*. Jul 20, 2021;11(4):116-129. [doi: [10.5662/wjm.v11.i4.116](https://doi.org/10.5662/wjm.v11.i4.116)] [Medline: [34322364](https://pubmed.ncbi.nlm.nih.gov/34322364/)]
15. Trevelyan EG, Robinson PN. Delphi methodology in health research: how to do it? *Eur J Integr Med*. Aug 2015;7(4):423-428. [doi: [10.1016/j.eujim.2015.07.002](https://doi.org/10.1016/j.eujim.2015.07.002)]
16. LimeSurvey: an open source survey tool hamburg. LimeSurvey GmbH. 2024. URL: <http://www.limesurvey.org> [Accessed 2025-09-18]
17. Wu H, Leung SO. Can Likert scales be treated as interval scales?—A simulation study. *J Soc Serv Res*. Aug 8, 2017;43(4):527-532. [doi: [10.1080/01488376.2017.1329775](https://doi.org/10.1080/01488376.2017.1329775)]
18. Gattrell WT, Logullo P, van Zuuren EJ, et al. ACCORD (ACcurate COnsensus Reporting Document): A reporting guideline for consensus methods in biomedicine developed via a modified Delphi. *PLOS Med*. Jan 23, 2024;21(1):e1004326. [doi: [10.1371/journal.pmed.1004326](https://doi.org/10.1371/journal.pmed.1004326)]

Abbreviations

ACCORD: Accurate Consensus Reporting Document
AI: artificial intelligence
CINEX: Clinical Information Extraction
CONSORT-AI: Consolidated Standards of Reporting Trials–Artificial Intelligence
E&E: explanation and elaboration
eDelphi: electronic Delphi
EQUATOR: Enhancing the Quality and Transparency Of Health Research
IE: information extraction
LLM: large language model
NLP: natural language processing
SPIRIT-AI: Standard Protocol Items: Recommendations for Interventional Trials–Artificial Intelligence

TRIPOD+AI: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis–Artificial Intelligence

TRPIOD-LLM: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis Or Diagnosis specifically tailored for large language models

Edited by Javad Sarvestan; peer-reviewed by Jiaping Zheng, Paul Blazey; submitted 30.04.2025; final revised version received 10.09.2025; accepted 15.09.2025; published 24.09.2025

Please cite as:

Reichenpfader D, Müller H, Denecke K

Improving the Reporting Quality of Studies on Information Extraction From Clinical Texts: Protocol for the Development of a Consensus-Based Reporting Guideline

JMIR Res Protoc 2025;14:e76776

URL: <https://www.researchprotocols.org/2025/1/e76776>

doi: [10.2196/76776](https://doi.org/10.2196/76776)

© Daniel Reichenpfader, Henning Müller, Kerstin Denecke. Originally published in JMIR Research Protocols (<https://www.researchprotocols.org>), 24.09.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.researchprotocols.org>, as well as this copyright and license information must be included.