

Protocol for a Consensus-based Reporting Guideline for Studies on Information Extraction from Clinical Texts (CINEX)

Daniel Reichenpfader, Henning Müller, Kerstin Denecke

Submitted to: JMIR Research Protocols
on: April 30, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	4
---------------------------------	----------

Preprint
JMIR Publications

Protocol for a Consensus-based Reporting Guideline for Studies on Information Extraction from Clinical Texts (CINEX)

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

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

²University of Geneva Faculty of Medicine PhD School of Life Sciences Geneva CH

³University of Geneva Department of Radiology and Medical Informatics Geneva CH

⁴HES-SO Valais-Wallis Informatics Institute Sierre CH

Corresponding Author:

Daniel Reichenpfader MSc
Bern University of Applied Sciences
Department Engineering and Computer Science
Institute for Patient-Centered Digital Health
Quellgasse 21
Biel
CH

Abstract

Background: Information extraction (IE) from clinical texts is increasingly important in healthcare, yet reporting practices remain inconsistent. Existing guidelines do not fully address the unique challenges of IE studies.

Objective: To develop CINEX, a consensus-based reporting guideline for studies on clinical IE.

Methods: CINEX is developed following established guideline methodology, including a three-round electronic Delphi (eDelphi) study with domain experts and a final in-person consensus meeting.

Results: A preliminary set of 28 items was drafted from a scoping review and existing frameworks. The draft guideline includes five key dimensions: information model, architecture, data, annotation, and outcome. This draft guideline will be refined through the eDelphi process starting May 2025.

Conclusions: CINEX will provide structured, expert-driven guidance for reporting clinical IE studies, improving transparency, reproducibility, and comparability.

(JMIR Preprints 30/04/2025:76776)

DOI: <https://doi.org/10.2196/preprints.76776>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/76776>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/76776>

Original Manuscript

Research protocol

Daniel Reichenpfader^{1,2}, Henning Müller^{3,4}, and Kerstin Denecke¹

¹Institute for Patient-Centered Digital Health, Bern University of Applied Sciences, Biel/Bienne, Switzerland

²Faculty 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, Quellgasse 21, 2502 Biel/Bienne, Switzerland. Phone: +41 31 848 60 93. Email: daniel.reichenpfader@bfh.ch.

Protocol for a Consensus-based Reporting Guideline for Studies on Information Extraction from Clinical Texts (CINEX)

Abstract

Background: Information extraction (IE) from clinical texts is increasingly important in healthcare, yet reporting practices remain inconsistent. Existing guidelines do not fully address the unique challenges of IE studies.

Objective: To develop CINEX, a consensus-based reporting guideline for studies on clinical IE.

Methods: CINEX is developed following established guideline methodology, including a three-round electronic Delphi (eDelphi) study with domain experts and a final in-person consensus meeting.

Results: A preliminary set of 28 items was drafted from a scoping review and existing frameworks. The draft guideline includes five key dimensions: information model, architecture, data, annotation, and outcome. This draft guideline will be refined through the eDelphi process starting May 2025.

Conclusions: CINEX will provide structured, expert-driven guidance for reporting clinical IE studies, improving transparency, reproducibility, and comparability.

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. In healthcare, this includes tasks like named entity recognition, relation extraction, template filling, and summarization. These methods enable efficient reuse of clinical narratives, supporting decision-making, research, and automation in electronic health records. Recent advances in natural language processing (NLP), especially through large language models (LLMs), have transformed IE capabilities. LLMs now support both traditional and generative approaches to extracting clinical information, with tasks ranging from identifying diagnoses to generating discharge summaries. However, this rapid technical evolution has introduced fragmentation in methods, terminology, and evaluation. Studies often vary in reported elements, task definitions, and benchmarking practices, making results difficult to interpret, compare, or reproduce. As a result, reporting guidelines have emerged to support the development and evaluation of these studies. However, while existing guidelines such as CONSORT-AI [1], SPIRIT-AI [2], and TRIPOD-ML [3] have improved reporting standards for certain types of AI studies, none directly address the specific challenges of information extraction from clinical text.

To address these gaps, we propose CINEX: a reporting guideline for clinical information extraction. CINEX aims to provide a structured yet flexible framework for reporting IE studies, agnostic to the underlying technology (e.g., rule-based, LLM-based, hybrid). It is tailored specifically to the needs of IE research in healthcare — enabling greater transparency, comparability, and reuse across studies. This protocol outlines the methodology behind CINEX, including a three-round modified electronic Delphi (eDelphi) process with domain experts. The guideline development follows best practices for health research reporting, and includes steps for item refinement, a final in-person consensus meeting, and the publication of an explanatory document (E&E).

Methods

Overview and Preliminary Guideline Development

The planned development process of CINEX was registered and publicly listed with the EQUATOR Network as a reporting guideline under development in August 2024 prior to its official start in October 2024 [4]. The EQUATOR (Enhancing the QUALity and Transparency Of health Research) Network is an international initiative that promotes transparent and accurate reporting of health research to improve the reliability and value of the published literature [5]. CINEX follows the methodological guidance outlined by Moher et al. for developing health research reporting guidelines [6]. 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 information extraction from radiology reports <1> [7]. Based on the findings of that review, as well as inspired by existing reporting guidelines, an initial set of candidate reporting items was drafted and published <2> [8].

eDelphi Study

To refine the preliminary guideline, a three-round eDelphi study is 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 Moher et al.'s recommendations for reporting guideline development [6].

We designed the eDelphi process based on the methodological principles of Häder [9], Nasa et al. [10], as well as Trevelyan and Robinson [11]. We aim to recruit between 20 and 30 participants to ensure both broad expertise and statistical validity <4>. Minimum response rates are set at 30 % for the first round and 70 % for the following rounds, consistent with Häder's guidance. If these thresholds are not met, additional experts may be recruited in later rounds.

Eligible participants include authors of studies identified in the preceding scoping review, as well as domain experts 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 [12].

The eDelphi study will comprise exactly three rounds [9]. Consensus for inclusion of an item in the guideline will be defined as a mean rating of 8 or higher on a 10-point Likert scale, with a standard deviation (SD) of 2 or less. Exclusion will be defined as a mean rating of 3 or lower, also with an SD ≤ 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, panellists have the possibility to add additional items for each domain.

Exclusion of items is only conducted after rounds two and three. Rephrasing of items and descriptions is conducted after each round.

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 Likert scale. To approximate interval-level data, this scale was selected instead of a classical 5-point scale based on methodological recommendations [13].

Feedback will be proactively provided by the executive committee for each new round (active feedback loop): Participants re-evaluate items that have not been excluded, augmented by aggregated ratings of the whole panel (as histograms), moderated anonymized comments from the previous round, as well as each participant's own previous ratings. This format enables reflection and informed re-evaluation. To minimize bias, items within each domain of the preliminary guideline will be presented in randomized order in each round.

Besides participating in the e panel, each expert completes a self-disclosure of their expertise in clinical information extraction. Consistent with Häder, 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 CINEX guideline publication and explanatory document.

A pretest of the first-round survey will be conducted among the executive committee [n=3, DR, HM, 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 this workshop, the results of the eDelphi study are presented and all items lacking consensus are discussed. The workshop results in the formal finalization of the CINEX guideline. After this workshop, the guidance statement is finalized <9>, complemented by an E&E <10>, and both are made available as publication <11>. The guideline development process will be reported in a structured way in accordance to the Accurate Consensus Reporting Document (ACCORD) reporting guideline [14]. Responsibilities for post-publication activities <12-18> are divided upon the contributing authors. This study was granted an exemption by the competent cantonal ethics committee 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 five dimensions (information model, architecture, data, annotation, and outcome) has been published in August 2024 [8]. These items will be presented in the first round of the eDelphi study.

Discussion

Principal Results

This protocol presents the development of CINEX, a reporting guideline specifically developed for clinical IE studies. CINEX addresses key challenges in the field — including inconsistent

terminology, reporting, and evaluation — by providing a structured, technology-agnostic framework tailored to IE tasks in healthcare. The guideline is developed through a three-round eDelphi study, supported by established methodology for health research reporting and finalized through a consensus meeting.

Limitations

Our approach has several limitations. 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, CINEX may require future updates to accommodate rapid developments in LLMs and multimodal approaches.

Comparison with Prior Work

CINEX builds on the methods used in guidelines like CONSORT-AI, SPIRIT-AI, and TRIPOD-LLM, but is the first to focus specifically on clinical IE. Our study design incorporated randomized item presentation, detailed feedback loops, and a 10- instead of 5-point Likert scale 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 CINEX's distinct contribution.

Conclusions

CINEX 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.

Acknowledgements

Reichenpfader: Methodology, Writing – Original Draft, Writing – Review & Editing

Müller: Supervision, Writing – Review & Editing

Denecke: Conceptualization, Supervision, Writing – Review & Editing

The CINEX reporting guideline is part of a PhD project at the University of Geneva. No dedicated funding was obtained.

Conflicts of Interest

None declared

Abbreviations

AI: Artificial Intelligence

EQUATOR: Enhancing the QUALity and Transparency Of health Research

IE: Information Extraction

JMIR: Journal of Medical Internet Research

LLM: Large Language Model

NLP: Natural Language Processing

References

1. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial

- reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* Nature Publishing Group; 2020 Sep;26(9):1364–1374. doi: 10.1038/s41591-020-1034-x
2. Cruz Rivera S, Liu X, Chan A-W, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med* Nature Publishing Group; 2020 Sep;26(9):1351–1363. doi: 10.1038/s41591-020-1037-7
 3. Gallifant J, Afshar M, Ameen S, Aphinyanaphongs Y, Chen S, Cacciamani G, Demner-Fushman D, Dligach D, Daneshjou R, Fernandes C, Hansen LH, Landman A, Lehmann L, McCoy LG, Miller T, Moreno A, Munch N, Restrepo D, Savova G, Umeton R, Gichoya JW, Collins GS, Moons KGM, Celi LA, Bitterman DS. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat Med* Nature Publishing Group; 2025 Jan;31(1):60–69. doi: 10.1038/s41591-024-03425-5
 4. The EQUATOR Network. EQUATOR network. Reporting guidelines under development for other study designs. CINEX. 2024. Available from: <https://www.equator-network.org/library/reporting-guidelines-under-development/reporting-guidelines-under-development-for-other-study-designs/#CINEX>
 5. The EQUATOR Network. About us | EQUATOR Network. 2025. Available from: <https://www.equator-network.org/about-us/> [accessed Mar 30, 2025]
 6. Moher D, Schulz KF, Simera I, Altman DG. Guidance for Developers of Health Research Reporting Guidelines. *PLoS Med* 2010 Feb 16;7(2):e1000217. PMID:20169112
 7. Reichenpfader D, Müller H, Denecke K. A scoping review of large language model based approaches for information extraction from radiology reports. *npj Digit Med* 2024 Aug 24;7(1):222. doi: 10.1038/s41746-024-01219-0
 8. 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. *Studies in Health Technology and Informatics* IOS Press; 2024. doi: 10.3233/SHTI240744 ISBN:978-1-64368-533-5
 9. Häder M. *Delphi-Befragungen: Ein Arbeitsbuch*. Wiesbaden: Springer Fachmedien; 2014. doi: 10.1007/978-3-658-01928-0 ISBN:978-3-658-01927-3
 10. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: How to decide its appropriateness. *World J Methodol* 2021 Jul 20;11(4):116–129. PMID:34322364
 11. Trevelyan EG, Robinson PN. Delphi methodology in health research: how to do it? *European Journal of Integrative Medicine* 2015 Aug 1;7(4):423–428. doi: 10.1016/j.eujim.2015.07.002
 12. Limesurvey GmbH. LimeSurvey: An Open Source survey tool Hamburg, Germany. URL. Hamburg: Limesurvey GmbH; 2024. Available from: <http://www.limesurvey.org>
 13. Wu H, Leung S-O. Can Likert Scales be Treated as Interval Scales?—A Simulation Study. *Journal of Social Service Research* Routledge; 2017 Aug 8;43(4):527–532. doi: 10.1080/01488376.2017.1329775

14. Gattrell WT, Hungin AP, Price A, Winchester CC, Tovey D, Hughes EL, van Zuuren EJ, Goldman K, Logullo P, Matheis R, Harrison N. ACCORD guideline for reporting consensus-based methods in biomedical research and clinical practice: a study protocol. Research Integrity and Peer Review 2022 Jun 7;7(1):3. doi: 10.1186/s41073-022-00122-0

