

Research Letter

Engaging Stakeholders With Professional or Lived Experience to Improve Firearm Violence Lexicon Development

Nicole Cook¹, MPA, PhD; Frances M Biel¹, MPH; Kerry Ann Bet¹, MPH; Marion R Sills¹, MPH, MD; Ali Al Bataineh², PhD; Pedro Rivera¹, MS; Anna R Templeton¹, DNP; Natalie Cartwright², PhD

¹OCHIN, Portland, OR, United States

²Norwich University, Northfield, VT, United States

Corresponding Author:

Nicole Cook, MPA, PhD
OCHIN
PO Box 5426
Portland, OR, 97228
United States
Phone: 1 9546126511
Email: cookn@ochin.org

Abstract

Framing the public health burden of firearm violence should include people with secondary exposure to firearm violence beyond acute bodily injury, yet such data are limited. Electronic health record clinical notes, when leveraged through natural language processing (NLP), are a potential source of data on firearm exposure. As part of NLP lexicon development, diverse stakeholders were engaged to identify keywords, and our findings demonstrated that engaging diverse stakeholders adds valuable input to NLP development.

JMIR Form Res 2025;9:e68105; doi: [10.2196/68105](https://doi.org/10.2196/68105)

Keywords: firearm violence; firearm injury; stakeholder engagement; natural language processing; electronic health record

Introduction

Exposure to firearm violence includes witnessing a shooting, being threatened by a firearm, losing a loved one to gun violence, and sustaining injuries from firearms [1]. Such exposure is associated with adverse physical and behavioral impacts. To better understand health sequelae following firearm violence exposure, some researchers recommend including these exposures in ongoing surveillance and research [2,3]. Although structured data (eg, diagnostic codes) for exposure to firearm violence are largely unavailable, unstructured electronic health record (EHR) data represent a potential source of information collected in the clinical care process [4]. To explore the applicability of such data in surveillance and research, we developed a natural language processing (NLP) pipeline for using unstructured EHR fields, including clinical notes, to identify those with exposure to firearm violence and subsequently understand more about the health impacts of such exposure. A crucial part of formative work on NLP lexicon development is effectively engaging a broad representation of stakeholders to contribute lexicon

terms that might not otherwise be known to study teams [5]. In this research letter, we describe the process and outcomes of engaging stakeholders in NLP lexicon development.

Methods

Study Design

We assembled a stakeholder advisory committee (SAC) comprised of patients, advocates, clinicians, researchers, and others with lived and/or professional experience that includes exposure to firearm violence. The SAC structure was designed to include community advocates, patients with lived experience, practicing clinicians, clinical researchers, data scientists, and clinical informaticists. SAC members were recruited via prior collaboration with a study team member, word of mouth, and presentations to patients and clinicians at web-based workgroup meetings. The final SAC composition included 12 members—5 community advocates and/or patients with lived experience, 1 physician, 2 clinician researchers, 2 clinical informaticists, and 2 data scientists.

In April 2024, SAC members reviewed the lexicon of identified keywords that was developed by the study team [4] and informed by MacPhaul et al's [6] NLP study, which investigated firearm injury intent. Afterward, SAC members suggested additional terms from their lived or professional experience that may indicate exposure to firearm violence. The review of terms and suggestion of new terms were done via an asynchronous, interactive SAC meeting.

New terms identified by the SAC were searched across EHR clinical notes from 7,103,301 patients who were receiving primary or behavioral health care at community-based health centers in the OCHIN multistate network [7]. A maximum of 30 random notes per keyword were reviewed by a study team member to determine whether a note indicated exposure to firearm violence from a powder mechanism [8]. Descriptive statistics were used to summarize results in a table for review, discussion, and adjudication by the study team and SAC.

Ethical Considerations

The Norwich University institutional review board deemed this study exempt. Stakeholders agreed to guide this study and received compensation for participating in monthly meetings; they were not research participants. Stakeholders completed

a professional services agreement, which outlined roles and responsibilities, and a nondisclosure agreement prior to joining the SAC. The EHR study data used by the study team for NLP development were a limited dataset. SAC members were not provided access to the limited dataset.

Results

The initial lexicon of identified keywords provided by the study team to the SAC included 19 terms, of which 13 were used in pilot work and 6 were additionally identified by the study team [4]. SAC members identified 35 additional keywords that were not included in the first iteration of the lexicon. Of the 35 terms, 27 had at least one mention in at least one clinical note. Of the 585 clinical notes reviewed, 5 contained 4 SAC-identified terms—*bbs*, *buckshot*, *firing*, and *metal pole*—and possibly indicated exposure to firearm violence. The study team met to perform the final adjudication for including or excluding each term. Two terms (*buckshot* and *firing*) were determined to have sufficient contextual information to indicate true exposure (Table 1). Notes that included these terms were added to the NLP testing and training dataset (Figure 1).

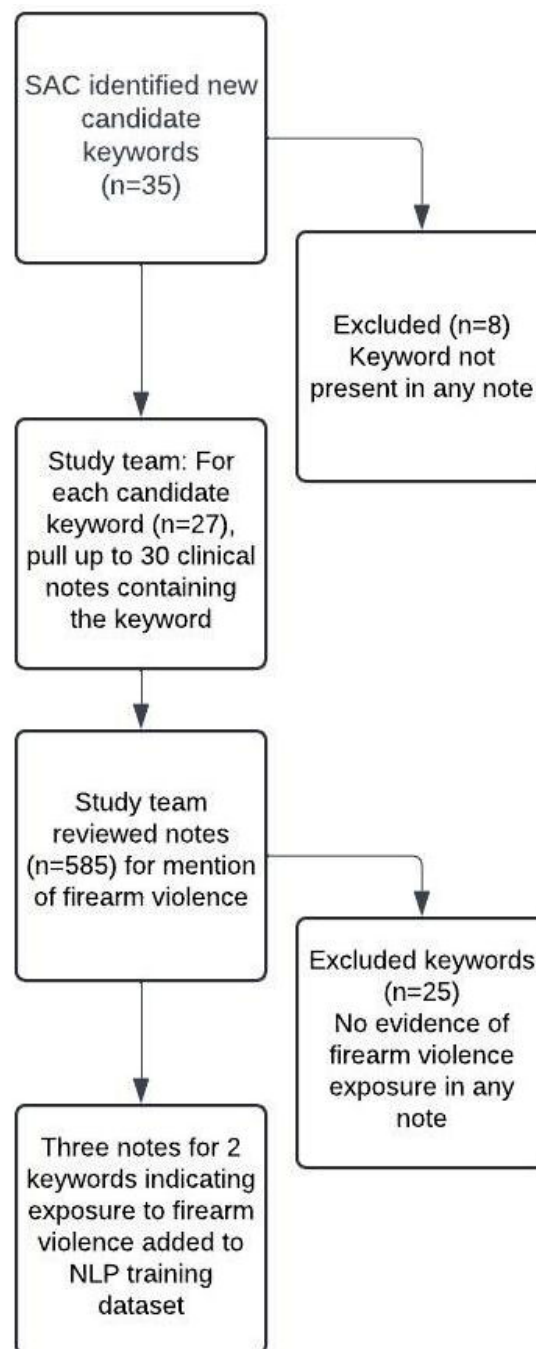
Table 1. Manual review of electronic health record (EHR) clinical notes (N=585) containing stakeholder advisory committee (SAC)-identified firearm lexicon terms from 300 ambulatory health care clinics.

SAC-identified term	Notes in EHR with keyword, n	Clinical notes reviewed ^{a,b} , n	Notes with possible indication of exposure to firearm violence, n	Text portion reviewed by study team to determine exposure	Final determination that reviewed clinical notes indicated firearm violence exposure
<i>buckshot</i>	1	1	1	<ul style="list-style-type: none"> “My dad had just been shot in the back, he had buckshot in him.” 	Yes
<i>firing</i>	13	13	2	<ul style="list-style-type: none"> “He reports there was a firing of fire arms at his apartment complex this last weekend.” “A few days later, the same brother was apparently responsible for firing a gun into the house; the bullet traversed two rooms and came to rest very close to the patient’s son.” 	Yes
<i>metal pole</i>	2	2	1	<ul style="list-style-type: none"> “Pt believes she is being tracked to be murdered, pt sees others are potential threats, as people who are trying to kill her- ... reports pt threatening another wielding a metal pole” 	No
<i>bbs</i>	30	30	1	<ul style="list-style-type: none"> “CI said he had a gun and he was not going to jail. (He had a BB gun, broken, threw it in the woods)” 	No

^aAll notes for terms that had less than 30 notes were reviewed. For terms with >30 notes, a random sample of notes was reviewed.

^bAdditional stakeholder advisory committee-identified terms with clinical notes that did not indicate firearm violence exposure include *Banger*, *draco*, *semiautomatic*, *gat*, *toaster*, *drill*, *stick*, *strap*, *trigger*, *heat*, *heater*, *iron*, *metal*, *nina*, *nine*, *piece*, *pole*, *rod*, *burner*, *cap*, *carrying*, *69*, and *ammunition*.

Figure 1. Diagram of SAC keyword term review for firearm violence exposure. NLP: natural language processing; SAC: stakeholder advisory committee.



Discussion

Effectively engaging varied stakeholders in NLP lexicon development led to the identification of 2 additional terms (*buckshot* and *firing*) that were previously not considered by the study team. Novel clinical notes with these keywords

were input into the training and testing dataset to enhance NLP model performance. Although most words suggested by the SAC did not indicate true exposure, the exercise demonstrated that periodic engagement of different stakeholder advisors in artificial intelligence and machine learning research is an important strategy to reduce data bias [9,10].

Acknowledgments

The authors are grateful to Angela Williams, Patricia Devine Wilder, Jennifer Obermeyer, Jelena MacLeod, and the rest of the stakeholder advisory committee for their work driving efforts toward trauma-informed care that supports collection and identification of firearm violence exposure data in patient electronic health records.

The research reported in this work was powered by PCORnet. PCORnet has been developed with funding from the Patient-Centered Outcomes Research Institute (PCORI) and conducted with the Accelerating Data Value Across a National

Community Health Center Network (ADVANCE) Clinical Research Network (CRN). ADVANCE is a clinical research network in PCORnet led by OCHIN in partnership with Health Choice Network, Fenway Health, University of Washington, and Oregon Health & Science University. ADVANCE's participation in PCORnet is funded through the PCORI Award RI-OCHIN-01-MC.

This work was supported by the AIM-AHEAD Coordinating Center, funded by the National Institutes of Health (NIH). The research reported in this publication was supported by the Office of the Director, NIH Common Fund, under award number 1OT2OD032581, OTA: 1OT2OD032581. The work is solely the responsibility of the authors and does not necessarily represent the official view of AIM-AHEAD or the NIH.

Conflicts of Interest

None declared.

References

1. Schumacher S, Kirzinger A, Presiado M, Valdes I, Mollyann B. Americans' experiences with gun-related violence, injuries, and deaths. KFF. Apr 11, 2023. URL: <https://www.kff.org/other/poll-finding/americans-experiences-with-gun-related-violence-injuries-and-deaths/> [Accessed 2025-04-07]
2. 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](https://doi.org/10.1001/jama.2022.16894)] [Medline: [36166012](https://pubmed.ncbi.nlm.nih.gov/36166012/)]
3. 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](https://doi.org/10.1001/jama.2022.21997)] [Medline: [36786795](https://pubmed.ncbi.nlm.nih.gov/36786795/)]
4. Cook N, Biel FM, 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. Nov 4, 2024;7(4):ooae120. [doi: [10.1093/jamiaopen/ooae120](https://doi.org/10.1093/jamiaopen/ooae120)] [Medline: [39498385](https://pubmed.ncbi.nlm.nih.gov/39498385/)]
5. Wan R, Kim J, Kang D. Everyone's voice matters: quantifying annotation disagreement using demographic information. In: Proceedings of the 37th AAAI Conference on Artificial Intelligence. Association for the Advancement of Artificial Intelligence; 2023:14523-14530. [doi: [10.1609/aaai.v37i12.26698](https://doi.org/10.1609/aaai.v37i12.26698)]
6. MacPhaul E, Zhou L, Mooney SJ, et al. Classifying firearm injury intent in electronic hospital records using natural language processing. JAMA Netw Open. Apr 3, 2023;6(4):e235870. [doi: [10.1001/jamanetworkopen.2023.5870](https://doi.org/10.1001/jamanetworkopen.2023.5870)] [Medline: [37022685](https://pubmed.ncbi.nlm.nih.gov/37022685/)]
7. Health IT and EHR solutions - community health care. OCHIN. URL: <https://ochin.org/> [Accessed 2025-04-07]
8. Pulcini CD, Goyal MK, De Souza HG, et al. A firearm violence research methodologic pitfall to avoid. Acad Emerg Med. Sep 2022;29(9):1140-1145. [doi: [10.1111/acem.14491](https://doi.org/10.1111/acem.14491)] [Medline: [35332976](https://pubmed.ncbi.nlm.nih.gov/35332976/)]
9. Zou J, Schiebinger L. Ensuring that biomedical AI benefits diverse populations. EBioMedicine. May 2021;67:103358. [doi: [10.1016/j.ebiom.2021.103358](https://doi.org/10.1016/j.ebiom.2021.103358)] [Medline: [33962897](https://pubmed.ncbi.nlm.nih.gov/33962897/)]
10. 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 30, 2023;2(6):e0000288. [doi: [10.1371/journal.pdig.0000288](https://doi.org/10.1371/journal.pdig.0000288)] [Medline: [37390116](https://pubmed.ncbi.nlm.nih.gov/37390116/)]

Abbreviations

EHR: electronic health record

NLP: natural language processing

SAC: stakeholder advisory committee

Edited by Amaryllis Mavragani; peer-reviewed by Mahmoud Elbattah, Richard Khoury; submitted 29.10.2024; final revised version received 18.03.2025; accepted 19.03.2025; published 21.04.2025

Please cite as:

Cook N, Biel FM, Bet KA, Sills MR, Al Bataineh A, Rivera P, Templeton AR, Cartwright N

Engaging Stakeholders With Professional or Lived Experience to Improve Firearm Violence Lexicon Development

JMIR Form Res 2025;9:e68105

URL: <https://formative.jmir.org/2025/1/e68105>

doi: [10.2196/68105](https://doi.org/10.2196/68105)

© Nicole Cook, Frances M Biel, Kerry Ann Bet, Marion R Sills, Ali Al Bataineh, Pedro Rivera, Anna R Templeton, Natalie Cartwright. Originally published in JMIR Formative Research (<https://formative.jmir.org/>), 21.04.2025. 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 Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.