Engaging stakeholders with professional or lived experience to improve firearm violence NLP lexicon

Nicole Cook, Frances M. Biel, Kerry Ann Bet, Marion R. Sills, Ali Al Bataineh, Pedro Rivera, Anna Rose Templeton, Natalie Cartwright

Submitted to: JMIR Formative Research
on: October 29, 2024

Table of Contents

Engaging stakeholders with professional or lived experience to improve firearm violence NLP lexicon

Nicole Cook[1] PhD, MPA; Frances M. Biel[1] MPH; Kerry Ann Bet[1] MPH; Marion R. Sills[1] MD, MPH; Ali Al Bataineh[2] PhD; Pedro Rivera[1] MS; Anna Rose Templeton[1] DNP; Natalie Cartwright[2] PhD

[1]OCHIN Portland US
[2]Norwich University Northfield US

Corresponding Author:
Nicole Cook PhD, MPA
OCHIN
PO Box 5426
Portland
US

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 is limited. Electronic Health Record (EHR) clinical notes leveraged through National Language Processing (NLP) is a potential data source on firearm exposure. As part of NLP lexicon development, diverse stakeholders were engaged to identify keywords. Findings demonstrated that engaging diverse stakeholders adds valuable input that will support NLP development and model performance.

(JMIR Preprints 29/10/2024:68105)
DOI: https://doi.org/10.2196/preprints.68105

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 v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Engaging stakeholders with professional or lived experience to improve firearm violence NLP lexicon

WORD COUNT: 578

ACKNOWLEDGEMENTS:
This work was supported by the AIM-AHEAD Coordinating Center, funded by the NIH.
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.
Research reported in this publication was supported by the Office of the Director, National Institutes of Health Common Fund under award number 1OT2OD032581-01. The work is solely the responsibility of the authors and does not necessarily represent the official view of AIM-AHEAD or the National Institutes of Health.
Award Number: 1OT2OD032581-02-347
OTA: 1OT2OD032581

Conflict of Interest Statement:
Nicole Cook, Phd, MPA[1], Frances M. Biel, MPH[1] , Kerry Ann Bet, MPH[1], Marion R. Sills, MD, MPH.[1], Ali Al Bataineh, Ph.D.[2] , Pedro Rivera, MS[1], Anna Rose Templeton, DNP.[1], Natalie Cartwright, Phd[2] have no conflict of interests related to this publication.

INTRODUCTION

As injury and death by firearm continues to increase in the United States, so too does exposure to firearm violence with or without acute bodily injury. Such exposure can lead to adverse physical and behavioral impacts.[1] To better frame the public health problem of firearm violence researchers recommend that firearm violence exposure be included in surveillance and research.[2 3] Identifying exposure to firearm violence from electronic health record (EHR) data is limited by the lack of International Classification of Disease codes for firearm exposure; it is also not a standard structured field collected in the process of clinical care. Thus, the AA-CONVENE study team is developing a Natural Language Processing (NLP) pipeline to use 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 this work is effectively engaging diverse stakeholders to guide the use of this unstructured data in NLP lexicon development. Recognizing the importance of fostering trusted relationships with diverse communities in Artificial Intelligence/Machine Learning (AI/ML) work[4], the study team works with 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. An early activity of the SAC was to review a list of keywords identified by the study team informed by MacPhaul et. al's NLP study investigating firearm injury intent.[5] These keywords are used to search for clinical notes that are indicative of exposure to firearm violence and then used in training the NLP model. In this brief, we describe the process and outcomes of engaging stakeholders in NLP lexicon development.

METHODS:

Designed to include multiple stakeholder groups, the SAC was established during the first three months of the AA-CONVENE study and includes 12 members: five community advocates and/or patients with lived experience; one physician; two clinician researchers; two clinical informaticists; and two data scientists with firearm violence experience. In April 2024, SAC members were asked to

review the lexicon of identified keywords and suggest additional terms from their lived or professional experience that may indicate exposure to firearm violence. This lexicon of new candidate keywords was used to search across EHR clinical notes from 7,103,301 patients receiving primary or behavioral healthcare at community-based health centers in the OCHIN multistate network. A maximum of thirty random notes per keyword were reviewed by a study team member to determine whether the note was indicative of exposure to firearm violence. Descriptive statistics were used to summarize results in a table for review by the study team and the SAC.

RESULTS:

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 that were reviewed, five notes from four terms (“bbs”, “buckshot”, “firing”, “metal pole”) had notes possibly indicating exposure to firearm violence. The study team met to perform final adjudication. Two terms (“buckshot” and “firing”) were determined to have sufficient contextual information to indicate true exposure.

DISCUSSION:

Including diverse stakeholder advisors in AI/ML research is an important strategy to reduce algorithmic bias and contribute to health equity.[6] In NLP lexicon development, keywords contributed by the SAC led to identification of novel clinical notes for input into training and testing datasets to improve NLP model performance. Stakeholders will continue to inform ongoing development of the NLP model.

REFERENCES

1. Mitchell KJ, Jones LM, Turner HA, Beseler CL, Hamby S, Wade R, Jr. Understanding the Impact of Seeing Gun Violence and Hearing Gunshots in Public Places: Findings From the Youth Firearm Risk and Safety Study. Journal of interpersonal violence 2021;36(17-18):8835-51 doi: 10.1177/0886260519853393[published Online First: Epub Date]|.
2. Kaufman EJ, Delgado MK. The Epidemiology of Firearm Injuries in the US: The Need for Comprehensive, Real-time, Actionable Data. JAMA 2022;328(12):1177-78 doi: 10.1001/jama.2022.16894 %J JAMA[published Online First: Epub Date]|.
3. Cook N, Sills M. Tracking All Injuries From Firearms in the US. JAMA 2023;329(6):514-14 doi: 10.1001/jama.2022.21997 %J JAMA[published Online First: Epub Date]|.
4. 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 digital health 2023;2(6):e0000288 doi: 10.1371/journal.pdig.0000288[published Online First: Epub Date]|.
5. MacPhaul E, Zhou L, Mooney SJ, et al. Classifying Firearm Injury Intent in Electronic Hospital Records Using Natural Language Processing. JAMA Network Open 2023;6(4):e235870-e70 doi: 10.1001/jamanetworkopen.2023.5870[published Online First: Epub Date]|.
6. Zou J, Schiebinger L. Ensuring that biomedical AI benefits diverse populations. EBioMedicine 2021;67:103358 doi: 10.1016/j.ebiom.2021.103358[published Online First: Epub Date]|.

Table 1: Manual review of Electronic Health Record clinical notes (N=585) containing stakeholder advisory committee identified firearm lexicon terms from 300 ambulatory health care clinics

SAC identified term	**Number of notes in EHR with keyword**	**Number of clinical notes reviewed[1,3]**	**# of notes with possible indication of exposure to firearm violence**	**Text portion reviewed by study team to determine exposure**	**Final determination that reviewed clinical notes indicated firearm violence exposure**
buckshot	1	1	1	*My dad had just been shot in the back, he had **buckshot** in him.*	yes
firing	13	13	2	*He reports there was a **firing** of fire arms at his apartment complex this last weekend.*	yes
				*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.*	
metal pole	2	2	1	*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
bbs	30	30	1	*Cl said he had a gun and he was not going to jail. (He had a **BB** gun, broken, threw it in the woods)*	no

1. All notes for terms that had less than 30 notes were reviewed. For terms with >30 notes, a random sample of notes were reviewed.
2. Not indicative of firearm violence exposure
3. Additional 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, ammunition

Figure 1: Diagram of SAC keyword term review for firearm violence exposure