Use of a Natural Language Processing-Based Approach to Extract Suicide Ideation and Behavior from Clinical Notes to Support Depression Research

Noa Palmon, BS, Safiyy Momen, BS, Michelle B. Leavy, MPH, Gary Curhan, MD, ScD, Costas Boussios, PhD, Richard E. Gliklich, MD | OM1, Inc., Boston, MA, USA

Background

- Suicide, as the tenth most common cause of death in the United States (US) in 2019, is a major public health concern [1]
    - In 2019 the age-adjusted suicide rate was 13.93 per 100,000 [2] and affected approximately 47,500 individuals [1]
- Suicide ideation and behavior (SI) is not recorded consistently in electronic health records (EHRs) using structured fields, limiting the robustness and utility of these data for research and clinical care [3]
- Natural language processing (NLP) can be used to extract relevant information from clinical notes and could improve availability of information on suicide ideation and behavior in real-world data sources to support research and clinical care

Objective

The objective of this study was to determine the feasibility of extracting SI from clinical notes.

Methods

Data Source

Data were drawn from the OM1 Real-World Data Cloud (OM1, Inc, Boston, MA, USA)

- A continuously updated dataset of US patients derived from deterministically linked, de-identified, individual-level health care claims, EHR, and other data covering 2013 to present day

Study Population

Patients with at least one clinical note that mentioned suicide were included in the study cohort

Approach

- An NLP-based approach was used to identify linguistic patterns used to record SI or denial thereof in clinical notes
- A language model was then constructed and optimized for identification of SI and behavior based on identified linguistic patterns

Methods (continued)

Approach (continued)

- The language model extracted the following concepts:
    - Presence of suicidal ideation
    - Frequency of suicidal ideation in the past 2 weeks, where available (once, several days, more than half the days, nearly every day)
    - Date of suicidal ideation
    - Negation of suicidal ideation
    - Documentation of a suicide attempt
- Notes were categorized as one of the following (Table 1)
    - Negations of SI (e.g., 'patient denies thoughts of suicide')
    - Affirmations (e.g., 'patient reports having thoughts of suicide')
    - Neither negations nor affirmations (e.g., notes that include medication label text about SI, family history)
    - Unknown (e.g., notes that mention suicide but do not indicate current SI)

Table 1: Example Categorizations of SI Concepts

Category	Definition	Example Note Excerpt
Negation	Does not report suicide ideation	… Has been on Zoloft for depression and he seems to be doing well on it, denies any suicidal ideations or thoughts…
Affirmation	Reports suicide ideation	…Stream of thoughts are linear. Thought content is appropriate. She has suicidal thoughts without intent. She denies auditory or visual hallucinations…
Neither	Notes that include medication label text about SI, family history, information about past suicide attempts	Past Psych History: outpt: outpt therapy past few months; inpt: none; past suicide attempts: none Uncle committed suicide in 2007[CR][LF]has cousins with D/A abuse issues
Unknown	No relevant information	[CR][LF]Hallucinations & delusions: + visual hallucinations Suicidality & homicidality: [CR][LF]Insight: poor

Validation

To ensure reliability of the patterns, subject matter experts validated the language model

Results

- Application of the inclusion criteria yielded more than 3.8 million patient notes
- Using this approach with validation of specific phrases by subject matter experts, we identified 2,088,144 clinical notes with a negation of SI and 25,255 clinical notes with an affirmation. The remaining notes were classified as neither negations nor affirmations or unknown (Table 2)
    - Of the 2,113,399 notes with either negation or affirmation of SI, 1.2% were affirmations and 98.8% were negations

Table 2: Results of Note Categorization

Category	Number of Notes Matched	Percent of All Notes
Negation	2,088,144	54.16%
Affirmation	25,255	0.66%
Neither	164,797	4.27%
Unknown	1,580,318	40.99%
Totals	3,858,514	

- The precision of the model was 0.83 for identifying the presence of suicide ideation and behavior and 0.93 for identifying the negation of suicide ideation and behavior (Table 3)
    - Recall was 0.71 and 0.70 for presence and absence of suicide ideation, respectively

Table 3: Preliminary Precision and Recall Using Custom NLP Model

Concept	Precision	Recall
Presence of suicide ideation and behavior	0.83	0.71
Negation of suicide ideation and behavior	0.93	0.70

Discussion

Strengths

- The custom NLP model created was based on linguistic patterns from millions of clinical notes across US clinical practices and clinical specialties
- The custom approach had high precision for both positive and negative reports of SI
- The model was validated by subject matter experts

Limitations

- With customized NLP, generalizability of the SI model to alternative data sources is not known, as clinical documentation patterns may vary across care settings
- The model does not differentiate between active and passive SI

Future Directions

- Future efforts should include the assessment of the reproducibility of this approach in other data sources
- An understanding of the feasibility of classifying SI as passive or active using data contained within the clinical notes is also needed
- Further work should improve upon capabilities for distinguishing SI from homicidal ideation in clinical notes

Conclusions

- Extraction of SI information from clinical notes using a customized NLP language model is feasible
- An NLP-based approach for identifying SI from clinical notes may be used to improve data availability for depression and SI research and quality improvement efforts

Conference

ISPOR Virtual May 2021

Employment

Noa Palmon, Safiyy Momen, Michelle B. Leavy, Gary Curhan, Costas Boussios, Rich E. Gliklich are all employees of OM1, Inc., Boston, MA, USA

Funding

This work is supported by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund under Interagency Agreement 18-596R-18. The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the United States (U.S.) Department of Health and Human Services.

References

1. Centers for Disease Control and Prevention. Web-based Injury Statistics Query and Reporting System (WISQARS). Leading Causes of Death Reports 1981 - 2019. Atlanta, GA. U.S Department of Health and Human Services. https://www.cdc.gov/injury/wisqars/fatal.html (Accessed on April 23, 2021).
2. Centers for Disease Control and Prevention. Web-based Injury Statistics Query and Reporting System (WISQARS). Fatal Injury Reports, National, Regional and State, 1981 - 2019. Atlanta, GA. U.S Department of Health and Human Services. https://www.cdc.gov/injury/wisqars/fatal.html (Accessed on April 23, 2021).
3. Anderson HD, Pace WD, Brandt E, et al. Monitoring Suicidal Patients in Primary Care Using Electronic Health Records. The Journal of the American Board of Family Medicine. 2015;28:65-71.