Author’s response to reviews

Title: Automatically Identifying Social Isolation from Clinical Narratives for Patients with Prostate Cancer

Authors:

Vivienne Zhu (viviennezh@gmail.com)

Leslie Lenert (lenert@musc.edu)

Brian Bunnell (bunnellb@musc.edu)

Jihad Obeid (jobeid@musc.edu)

Melanie Jefferson (sweatma@musc.edu)

Chanita Hughes-Halbert (hughesha@musc.edu)

Version: 2 Date: 14 Feb 2019

Author’s response to reviews:

Response to reviewers: We appreciate the reviewers’ careful review of our manuscript and thank them for their thoughtful and helpful comments and suggestions. We have made several revisions to respond to these comments, and these changes enhance the quality of our report. Detailed information on the changes is provided below.

[Reviewer 1]

[Comment] The one response that still leaves some unanswered questions, in my opinion, is the issue about why this particular NLP approach was chosen, and not, say, an open source alternative such as cTAKES or GATE. From an NLP perspective, the methodology might need some additional motivation - if, say another research group would want to try to replicate this or make a comparison study but are not able to buy a commercial system, it would be good to add a brief discussion as to how this could be done or at least some more discussion on alternative NLP approaches.

[Response] We used highly adopted and evaluated NLP software to conduct the analyses in this study. Linguamatics I2E is a highly rated commercial NLP software that is used by hospitals, pharmaceutical companies, and other organizations involved in US health care. In our Discussion section, lines 12-16 on page 9, we added the following text: “In this study, we utilized I2E’s user-friendly interface tools to accelerate the development of a lexicon for social
isolation; the lexicon generated from this study is reusable for open source NLP software (e.g., cTak) as customized dictionaries; thus the lexicon establishes a foundation for rapid progress on NLP tasks and later dissemination across research communities.”

[Reviewer 2]

Background

[Comment] It would be a better argument in this paper if there is an explanation for why we cared about extracting social isolation information from the EHRs. For example, what kinds of knowledge could be studied or generated by using the NLP to extract the social isolation information from narrative clinical notes.

[Response] As described in the Background section, social isolation is important to clinical outcomes across a wide variety of decisions. For instance, social isolation is an important social determinant that has been associated with an increased risk for recurrence and mortality among breast cancer patients and has an impact on health outcomes that is equal to the effects of smoking, obesity and hypertension. Social isolation is among the social determinants that the National Academy of Medicine recommends should be documented in the electronic health record. Please refer to background section lines 2-34 on pages 2 for the additional rationale we provided to further support the extraction of social isolation from electronic health records.

[Comment] The authors compared the outcome of the study with elderly population in the method and discussion sections. I assumed that it could be due to the prostate cancer patients tended to be elder males but this information was not mentioned in the background section.

[Response] The purpose for this study is to develop and evaluate NLP approaches extracting social isolation from clinical notes for prostate cancer patients, and we did not study age specifically. However, the study population is older; therefore, a published paper about social isolation for the elderly is a good resource for comparison. In our Results section, lines 30-32, page 5, we added the following text: “The study cohort (prostate cancer patients) were older (nearly 70 years of age on average), and a majority identified as White and as Medicare/Medicaid beneficiaries.

Method

[Comment] Under "Development of the lexicon for social isolation, this statement, "Social isolation is not commonly documented in the EHR including in clinical notes", was stated but the rationale of this statement is not clear.
In our Methods section, lines 39-41, page 3, we revised the text as follows: “Generating a lexicon of social isolation is challenging because there are no documented standards, and data collection strategies for social isolation in the EHR are in an early stage of development.”

Validated instruments were used to develop the social isolation lexicon. Only "Loneliness Scale" was listed. It would be clear to list all the instruments used to develop the lexicon.

We have clarified that we used the Loneliness Scale to develop the social isolation lexicon (see page 3, lines 41-45).

Two domain experts generated five seed terms ("lack companionship", "feel left out", "isolated", "loneliness", and "lonely"). Comparing these seed terms to the Table 2, it seems like that "lack companionship", "feel left out", and "isolated" were not included in the "final" list; but there is no explanation about it. Is there any limitation by using the seed terms generated by behavioral science researchers?

When we started to develop the seed terms, we intended to include as many of the initial terms as possible. We utilized standard terminologies integrated in I2E, the literature, and available instruments, and we relied on the suggestions of our behavioral science researchers at MUSC. Two terms (“lack companionship” and “feel left out”) were not found in the clinical notes; “isolated” was found (“feels/feel isolated”, “socially isolated”), and “loneliness” and “lonely” were frequently found. Social isolation is a research area under behavioral science.

Domain experts were used in the chart review (under Development of NLP algorithm to identify social isolation section”) but it was not clear about who these two domain experts were and why they were qualified for the chart review. Were they the same experts to generate the seed terms?

The two domain experts are not the experts who generated the seed terms. One reviewer is an Internist, and another one is a Psychiatrist. To avoid confusion regarding “domain expert” for our NLP performance evaluation, we refer to each of those experts as a “reviewer” and added their initials (BB and JO) in the text (Methods section, line 38, page 4).

The concept of social isolation is abstract. Without an adequate operational definition of social isolation for the domain experts could cause discrepancies during the manual
chart reviews. However, the training for the domain experts to make sure they used the consistent operational definition of social isolation was not mentioned in the manuscript. Although the authors mentioned the rater agreement is 97.4%; however, it is based on the "notes" but not based on the mentions. By the way, the percent agreement does not account for chance agreement.

[Response] In our Methods section, lines 39-40, page 4, we added the following text: “The reviewers were trained regarding the operational definition of social isolation and the methods of NLP assisted chart review”. In clinical notes, copying and pasting are very common; thus, accounting for sentence or mention level would inflate the results. Also, measuring by document level is a valid approach.

[Comment] "Two domain experts" were mentioned multiple times in the method sections; were they the same experts who are behavioral science researchers?

[Response] The two domain experts are not the experts who generated the seed terms for the lexicon. One reviewer is an Internist and another one is a Psychiatrist. To avoid confusion regarding “domain expert” for our NLP performance evaluation, we refer to each of those experts as a “reviewer” and added their initials (BB and JO) in the text (Methods section, line 38, page 4).

Results

[Comment] Table 1 shows the note type in the training dataset. Was the distribution of the note types in the testing dataset similar to those in the training dataset?

[Response] The distribution of the note types in the test dataset was similar. We added text as follows: “The distribution of note types in the test dataset was similar (data not shown).” (Results section, line 30, page 5)

[Comment] Figure 1 was hard to read due to the very small font size and the truncated words.

[Response] Figure 1 is the I2E built-in virtualization tool for output. We enlarged Figure 1, so the font is bigger. However, we cannot correct the truncation because it is a I2E built-in output display.
[Comment] In Table 2, the first word of some terms was capitalized. Do they have any meaning?

[Response] We intended to report the morphologic and synonymous variations to demonstrate the characteristics of clinical notes. These variations are different from standard clinical terminology.

[Comment] Based on the Table 3, this statement, "...more likely to be white and Medicare/Medicaid patients". Please also mentioned that the distribution of the race and insurance type of the "Positive" was similar to those presented in the "Negative".

[Response] In our Results section, lines 36-38, page 5, we revised the sentence as follows: “However, the statistical significance of these observation is uncertain due to the small number of NLP identified social isolation positives and the similar distribution of race and insurance type in the NLP identified negatives.”

[Comment] In Table 4, it would be more informative if the note types of the false positive findings were provided.

[Response] We did not specifically study false positive findings in relation to note type.

Discussion

[Comment] The authors mentioned "...the lexicon generated from the current study combines standard concepts and domain expert knowledge, our approach offers a more complete data extraction method". However, it was not very convinced. The domain experts only generated 5 terms and only 2 of them were listed in the final list. Moreover, what are those "standard concepts"?

[Response] The domain expert knowledge includes the suggestions of MUSC behavioral science researchers we consulted, as well as knowledge from the literature. I2E integrates standard terminologies, such as SNOMED, MESH, Rxnorm, etc. We included relevant concepts from SNOMED and MESH to generate the lexicon.
Limitations

[Comment] The authors developed a NLP approach to extract the social isolation information but only 24 terms were included. It is possible that many terms used to describe the social isolation were not included since the definition or description of social isolation can be vary depending on the population.

[Response] In our Limitations section, lines 14-18, page 10, we added the following text: “Our lexicon and NLP algorithms were developed to reflect the definition of social isolation for prostate cancer patients and may miss other terms suitable for other populations; therefore, the lexicon and algorithms may not be generalizable to other populations with different diseases or to other institutions without customization and evaluation.”

Conclusion

[Comment] Although the NLP algorithm in this study showed “a high accurate performance”, it could be due to a limited number of the relevant terms used in the algorithm.

[Response] In our Conclusions section, lines 41-43, page 10, we added the following text: “However, the lexicon was specifically developed for prostate cancer patients. Thus, customization and evaluation are needed for studying social isolation NLP extraction for other populations.”