Reviewer’s report

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

Version: 0 Date: 04 Oct 2018

Reviewer: Jia-Wen Guo

Reviewer's report:

Preventing social isolation is critical to the health and wellbeing of cancer patients. This manuscript presented the process of developing a NLP approach to extract social isolation information from the clinical notes.

Here are the reviewer's comments:

Background

There is lack of definition of social isolation, which is the key concept of this manuscript.

The sample of this study was prostate cancer patients. Although the authors mentioned that the study sample from a large study, there is still lack of explanation for why using this particular population to explore "social isolation". For example, is "social isolation" is a particular issue from this population or is there any gender or cancer diagnosis consideration regarding social isolation?

Social isolation could be generally categorized as "subjective" social isolation and "objective" social isolation. It was not clear whether the authors wanted to address both types of social isolation or the particular one in this study.

Method

Under "NLP software" section, it seemed like that the authors tried to use an exemplar to show the usefulness of 12 E. But please indicate that the concept of fall risk is very different from the concept of social isolation.

Under "Data source" section, the authors stated that "This study was conducted as part of a transdisciplinary center in precision medicine and minority men's health……". Please describe what "minority" men meant. Was any specific ethnicity or race included in the study?
Under "Development of the lexicon for social isolation" section, domain experts' knowledge was added to generate the social isolation terms. Please describe the background and knowledge expertise of those domain experts and why the authors considered that they were domain experts. "Live alone" was excluded in the lexicon; however, it could be part of objective social isolation. It was not clear what "WILL SEND CITATION" was indicated in the content. Based on the reference, Hughes et al's reference (titled "A short scale for measuring loneliness in large surveys: Results from two population-based studies"), the scales included in this article were measuring loneliness but not social isolation; but the manuscript content stated they were "Social Isolation Scale", which cannot be found in Hughes' article. Social isolation and loneliness are related but they are still different concepts.

Under "Development of NLP algorithm ....." section, one domain expert was used to chart review. Potentially, it could have a bias by using one domain expert which could be a limitation of this study design.

Under "NLP algorithm performance .....", two domain expects were used to validate the results. However, the report of the inter-rater/coder reliability (IRR) was not showed or used.

Results

It would be clear to give one or two examples to explain why the term "all family" is usually a pre-negation in clinical context.

"Nurse practitioner" was showed as one of the most common author types but not "nurse" who may also document patient's social isolation. I was wondering whether nursing notes were used in this study. Besides, please add the n and % for each author type.

"identified having social isolation by NLP were more likely to be White and Medicare/Medicaid patients" may be misleading description because the White and Medicare/Medicaid patients were the majority of the sample. Therefore, for the group which was identified as no social isolation, the patients were also likely be While and Medicare/Medicaid.

Under "NLP algorithm performance", two domain experts were used to review the notes but there was no IRR to show the consistency between these two experts.

Discussion

Based on Table 2, this statement, "African Americans have a slightly lower prevalence of social isolation compared to whites." may be a misleading statement because the number of African Americans was lower than white in the sample. Therefore, the audience also see the same thing in the group identified without social isolation.
There was no discussion regarding the finding of the study specifically for prostate cancer patients, which is the sample of the study.

The authors mentioned that the social exclusion and social outcast, two SNOMED-CT terms, were not found in clinical notes. However, there was no further discussion regarding this finding. It is suggested to the authors to discuss the possibility of these two terms are not commonly used in clinical settings or between/among providers. The clinical notes are used by the healthcare providers; if providers do not use the social exclusion and social outcast in their work or communicated with others, these two terms may not be showed in the clinical notes.

Social support related terms (e.g., lack of social support, limited social support) were included as part of lexicon of social isolation. However, social support and social isolation are different concepts; it could be the reason that social support is listed as another standard concept in SNOMED CT.

Due to the lexicon mixed with social support terms, the authors may need to further discuss it. By presenting that this term is suggested by the domain experts may not be sufficient also could bring an issue of whether the experts understand the difference between social support and social isolation.

Limitations

Limitations of this study were addressed properly. Another important limitation which is suggested to consider is that using clinical notes to detect or describe patients' social isolation has its limitation due to social isolation is subjective; therefore, the clinical notes which are used by the providers may not include all the social isolation information from the patients. The clinical notes capture the most information about providers' observations regarding the patient and some patients' chief complaints. Therefore, to understand whether a patient experiences social isolation, we still need to conduct a comprehensive assessment.

Moreover, the lexicon in this study may not be completed and also mix with "social support" terms which is different from social isolation.

Conclusion

Based on the concern related to including social support related terms in the lexicon of social isolation, there may be a question of whether the NLP algorithms in this study really can accurately identify the patient experienced social isolation.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

No

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests.

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal.