Reviewer's report

Title: EXTracion of EMR Numerical Data: An Efficient and Generalizable Tool to EXTEND Clinical Research

Version: 0 Date: 05 Aug 2019

Reviewer: Jeffrey Klann

Reviewer's report:

This is a well-written description and evaluation of a simple regex NLP approach for extracting numerical values in clinical notes. Particular strengths of the paper are the strong evaluation section and the proposed availability of the tool freely on GitHub.

However, at the time of this review, the GitHub repository did not exist. This needs to be rectified to be able to perform a comprehensive review.

Also, the manuscript presents a very simple layer on top of NTLK, and it is certainly not the only manuscript to describe using regexes for data extraction from clinical notes. Therefore, although the utility of the tool is strong, it does not seem particularly novel. To rectify this, please: provide more background and citations of related approaches, to contextualize and differentiate this study; highlight in more detail what is scientifically novel - the evaluation? the dictionary? The methodology? The software is probably not the novel component, because from the description it seems to be doing some simple array processing. (Though the missing source code makes this difficult to evaluate.)

A minor point in the evaluation: it is counterintuitive that the performance scores (e.g., PPV of EF) are higher when evaluating values than when only identifying existence. A careful read indicates that the existence-level evaluation treated each note as a single data point. Please add to the discussion an explanation of the impact of measuring each evaluation at different levels. You might also consider adding the average count of false positives and false negatives found per note for the value-level evaluation.

Finally, it seems curious that reference [10], which the manuscript states uses EXTEND, was published three years ago. I would assume EXTEND would need to be finished prior to that study being completed. Please explain.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

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.

Yes

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