Reviewer's report

Title: Rare Disease Knowledge Enrichment through a Data-Driven Approach

Version: 0 Date: 17 Sep 2018

Reviewer: Mei Liu

Reviewer's report:

The manuscript described a data-driven approach to enrich rare disease knowledge resource and demonstrated its value in providing decision support for rare disease diagnosis. The following are some concerns.

1. The study utilized clinical notes from EMR. What note types were included? E.g. HP, nursing flowsheet, discharge … or all EMR note types?

2. What constitute a phenotype in this study? All terms in HPO? How many phenotypes?

3. Typically, support and confidence are used together in association rule mining to determine significance of the rules learned. It usually requires user-specified minimum support and minimum confidence at the same time. A minimum support threshold is to find all frequent itemset and a minimum confidence is to apply to these frequent itemset to form rules. Using them as separate measures for discovering interesting rules is strange.

4. Phenotype-disease associations in SemMedDB has rankings? How were the rankings in SemMedDB determined?

5. Using disease co-occurrences in EMR as a gold-standard for validating correctness of suggested disease by Phenomizer and D3N may not be a good approach as: (1) disease co-occurrence in a note does not guarantee a positive association; (2) D3N is EMR-enriched so the evaluation between phenomizer and D3N against the disease co-occurrences in EMR is biased toward D3N.
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:

Needs some language corrections before being published

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