

## Reviewer Report

**Title:** Desiderata for the development of next-generation electronic health record phenotype libraries

**Version:** Original Submission    **Date:** 6/27/2021

**Reviewer name:** Wei-Qi Wei

### Reviewer Comments to Author:

High-quality phenotype definitions are desirable for clinical research. A phenotype library of portable, reproducible and validated phenotyping definitions will be valuable for the research community. The authors examined the work phenotyping models, implementation and validation, and summarized several desiderata for best practices in this review. Some points mentioned in the paper were similar to the previous report cited (<https://academic.oup.com/jamia/article/22/6/1220/2357938>). My primary concern regarding this piece of work is the phenotyping scope. The discussion and thoughts fit well for most rule-based phenotype definitions. However, more and more phenotyping research moves forward to either machine-learning-based or high-throughput approaches (e.g., PheMAP and PheNorm). Therefore, it is necessary to add discussions on these approaches. In addition, NLP algorithms could be vastly complicated. Therefore, it is essential to add more discussions regarding the complexity beyond NLP languages and packages.

### Methods

Are the methods appropriate to the aims of the study, are they well described, and are necessary controls included? Choose an item.

### Conclusions

Are the conclusions adequately supported by the data shown? Choose an item.

### Reporting Standards

Does the manuscript adhere to the journal's guidelines on [minimum standards of reporting?](#) Choose an item.

Choose an item.

### Statistics

Are you able to assess all statistics in the manuscript, including the appropriateness of statistical tests used? Choose an item.

### Quality of Written English

Please indicate the quality of language in the manuscript: Choose an item.

## Declaration of Competing Interests

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

- 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?
- 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?
- Do you hold or are you currently applying for any patents relating to the content of the manuscript?
- 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?
- Do you have any other financial competing interests?
- 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.

no

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.

Choose an item.

To further support our reviewers, we have joined with Publons, where you can gain additional credit to further highlight your hard work (see: <https://publons.com/journal/530/gigascience>). On publication of this paper, your review will be automatically added to Publons, you can then choose whether or not to claim your Publons credit. I understand this statement.

Yes Choose an item.