Reviewer’s report

Title: A novel data-driven workflow combining literature and electronic health records to estimate comorbidities burden for a specific disease: a case study on autoimmune comorbidities in patients with celiac disease

Version: 2 Date: 13 Jun 2017

Reviewer: Adam Dunn

Reviewer's report:

Thank you for inviting me to review the updated version of the manuscript. The authors describe the application of what is broadly a literature discovery and EHR co-morbidity mining pipeline to auto-immune conditions associated with celiac disease. The authors have addressed the suggestions and comments that I made.

1. The abstract is still unnecessarily long and could be made shorter and more precise.

2. The confidence intervals in the abstract are presented in an unusual way and should match the standard for the journal.

3. The conclusion in the abstract should draw the conclusion about the case study and not the general case - it would require many more case studies to prove the general case.

4. I still don't see the evidence for the statement "Today in all major hospitals Clinical Data Warehouses (CDW) gather information on hundred thousands of patients", which may not be true of *all* hospitals. Consider rephrasing.

5. There is probably no need to include the formula for an equation that has been cited and used in a standard way.
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.

Unable to assess

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