Reviewer's report

Title: Disease Associations Depend on Visit Type: Results from a Visit-Wide Association Study

Version: 0 Date: 26 Apr 2019

Reviewer: Weiye Wang

Reviewer's report:

This study investigated the association between outpatient diagnosis codes and certain visit types (Ob/Gyn clinics and cancer clinics), as well as compared the difference of "pain-related" conditions between these two clinics. A new concept of VisitWAS was established based on Phenome-Wide Association Studies (PheWAS). The manuscript is quite straightforward with clear results and conclusions. I have several concerns and suggestions about the contents.

1. From the introduction part, I firstly learned that PheWAS analyze the association between phenotypes and disease (line 9). According to my understanding, PheWAS examine the effects of many phenotypes on a single disease/genetic variant. However, in the following sentences, PheWAS became a study that examines the "gene-disease associations" (line 13-19), which sounded the same as the well-known GWAS. And later, the author wrote that PheWAS investigated genes associated with "structure disease codes" (line 19-21).

I thought the description here needs to be clarified: what is the exact definition of PheWAS (e.g.: Does it investigate the association of multi-phenotypes with a single trait such as gene/disease? Or the association of multi-genes with a single disease/phenotype?) Current words are confusing. It may also make the readers feel confused when they learn about the definition of VisitWAS: What is the full name of "VisitWAS"? If it were "Visit-Wide Association Studies", will it examine the relationship between many visit types and a single disease trait? Or will it examine the relationship between many disease traits and a single visit type?

2. Line 51 in the introduction part, it will be helpful to provide reference on why and how to define the "high pain" groups. Additionally, totally how many clinics were in your data (what are they)? Why were only Ob/Gyn and cancer clinics chosen? Why not other clinics (some may be highly concerned such as heart disease clinic)? Some more rationale should be addressed.
3. Line 11-14 in the 2.1 Dataset. More background details of data source are required. For instance, from which year to which year were the out-patient data obtained? How were these out-patient data obtained by the Hospital of UPenn (self-report? Lab examination? Follow-up call?)? Where is this hospital and patients in which area was covered?

4. Table 1, I am curious about the age distribution of participants. Please provide std together with mean or IQR together with median of their age. If possible, stratifying age into categories (e.g.: <20 yrs, 20-50 yrs, 50+ yrs) and providing the number of participants in each category will be helpful. So that the readers can have a rough sense of how many women are at teenage, reproductive age, and post-menopause.

5. I can understand that the author used Fisher's exact test because of the small sample size. But in order to provide a clear logic chain on why Fisher's was chosen, it may be better to show some power calculation and prove that the sample size isn't big enough to give acceptable power for a chi-square test or something else.

6. Figure 1 is kind of weird. I cannot find N=81, N=8 and N=882 from any part of the main text. The numbers are quite different from what Table 1 provided. Figure 1 or Table 1, one of them must be wrong. Also, the pictures are confusing. It will be much better to explain totally how many clinics were included in "all clinics" and what are they.

**Level of interest**

Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

**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

Do you want to get recognition for reviewing this manuscript? Add a record of this review to Publons to track and showcase your reviewing expertise across the world’s journals. Signing up is quick, easy and free!

Yes