Author’s response to reviews

Title: A stratification method based on clustering for the minimization of data masking effect in signal detection

Authors:

Jianxiang Wei (jxwei@njupt.edu.cn)
Yue Ding (b14070501@njupt.edu.cn)
Ming Li (526540057@qq.com)
Jun Sun (574278909@qq.com)

Version: 2 Date: 15 Aug 2019

Author’s response to reviews:

Dear Editor,

Thank you very much for your attention and the reviewer’s evaluation and comments on our paper. We have revised the manuscript according to your kind advices and reviewer’s detailed suggestions again. Enclosed please find the responses to the reviewer. Please refer to the document named “Manuscript-revised version2”, in which modified parts are marked in red. We sincerely hope this manuscript will be finally acceptable to be published on BMC Medical Informatics and Decision Making. Thank you very much for all your help and looking forward to hearing from you soon.

Best regards,
Jianxiang Wei
Nanjing University of Posts and Telecommunications
Email: jxwei@njupt.edu.cn

Response to Prof. Jim Slattery (Reviewer 2):

1. The answer given to Prof Salvo's first comment in the initial review has raised a doubt in my mind that was not there before. The authors say ‘If some DECs were detected as suspected signals by disproportionality analysis but did not appear in the reference database, we thought these signals were masked.’ This is not clearly the case. A signal associated with a DEC that does not appear in the reference database is very likely to be a false positive. It might make sense to suggest that signals corresponding to DECS that do appear in the reference database but which did not signal prior to clustering might have been previously masked. But even this is an assumption, it is only a statistically significant positive shift in the overall balance of results that demonstrates useful correction of masking.

Response: Thanks for your comments and very good advice. We are so sorry to make this mistake. Revision: If some DECs were not detected as suspected signals by disproportionality analysis but did appear in the reference database, we thought these signals were masked. Among 37,193 DECs of our study dataset, 12,493 DECs appear in the reference database, denoted as known ADRs. The detailed revision can be found in Line 31-35, Page 4.

2. The answer to my own first point is also unclear but this may partly be a matter of terminology. I think rather than 'known signals' the authors mean 'known ADRs'. A signal is just the result of a test applied to the data that may or may not correspond to an ADR. However, even if this is a list of ADRs the precision-recall curve should certainly pass through the precision-recall point cited in the paper. This was (0.4882, 0.4039) with stratification and (0.5456, 0.2993) without stratification, both of which are well away from the curves. This suggests something has gone wrong with the
calculations and they should be checked. Response: Thanks for your comments. We are sorry for the confusion caused by the terminology 'known signals', which should be 'known ADRs'. We re-expertimented with all the total data instead of known ADRs in the study dataset. The two points, (0.4882, 0.4039) with stratification and (0.5456, 0.2993) without stratification, were both given in the Fig. 4. Revisions were made in the Results section and the Discussion section, respectively.(1) Results sectionRevision: A precision-recall curve was adopted to further illustrate the performance of signal detection. First, sorting the study dataset in descending order based on the IC025 values before and after stratification respectively, and then calculating precision and recall for each DEC gradually. The precision-recall curve was shown in Fig. 4. Fig. 4 Precision-recall curveThe detailed revision can be found in Line 4-9, Page 8. (2)Discussion sectionRevision: Fig. 4 showed that the recall of non-stratification method was slightly better than that of stratification method under the same precision at the beginning of signal detection, but with more and more true positive signals were detected subsequently, the recall of stratification method was significantly better than before. On the whole, under the same precision, the recall of stratification is better than that obtained without stratification, which shows that the performance of signal detection has improved after stratification. The detailed revision can be found in Line 9-17, Page 9.