Reviewer's report

Title: A Richly Interactive Data Visualization Tool for Cohort Studies Using Electronic Medical Records

Version: 3 Date: 27 May 2015

Reviewer: Hong Kang

Reviewer's report:

In this article, using CKD as a model disease, the authors proposed a visual mining system which support exploratory data analysis of EMR data. The knowledge of relevant comorbidities that CKD patients develop over time are revealed by Sankey diagram, which makes sense of the EMR data. Overall, the authors have done extensive and solid investigation and well answered the question that how to make sense of the data. There are a few comments.

Major Compulsory revisions:

It is interesting that the system allow the user to define a set of factors by selecting independent codes or aggregating correlated ones based on their knowledge. The advantage is if the user wants to study the clinical trajectories of other disease, they can define a list of relevant factors and apply the same process to set up time stages, cluster patients, as well as explore the structure of the cohort trajectories. However, my concern is that, there would be a very high requirement for user’s experience of handling such complicated model independently, since most of them maybe healthcare providers without enough machine learning skills. I also concern about the stability of the system when using on other EMR data besides CKD, I suggest the authors to apply the system on another disease in order to support your point mentioned in the conclusion part that the analysis process is generalizable to any other disease that a user wishes to follow over time and can work with different clustering and filtering algorithms.

Minor Essential Revisions:

1. The figures present by Sankey diagram are very important for readers to understand the value of the system. Please make more detailed legend for these figures.

2. As the system is a web-based application, I suggest the authors to provide its link in the paper.

3. There’re several grammatical mistakes in the article, for instance, line 178, “Hence, our system also the user to rank and filter the associations based on their statistical importance.” Moreover, please reduce the exactly same sentences occurred in the parts of abstract, introduction and conclusion.

Level of interest: An article of importance in its field
Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests