Author’s response to reviews

Title: A computational method to quantitatively measure pediatric drug safety using Electronic Medical Records

Authors:

Gang Yu (yugbme@zju.edu.cn)
Xian Zeng (xian_zeng@zju.edu.cn)
Shaoqing Ni (chgcp@zju.edu.cn)
Zheng Jia (zjia@zju.edu.cn)
Weihong Chen (whhchen@sina.cn)
Xudong Lu (lvxd@zju.edu.cn)
Jiye An (an_jiye@zju.edu.cn)
Huilong Duan (duanhl@zju.edu.cn)
Qiang Shu (shuqiang@zju.edu.cn)
Haomin Li (hmli@zju.edu.cn)

Version: 2 Date: 09 Dec 2019

Author’s response to reviews:

Firstly, we would like to thank all the editors and reviewers for their thoughtful review of the manuscript and giving us this opportunity to improve this manuscript better. These comments are all valuable and helpful for improving our article. We agree with almost all their comments and have revised our manuscript accordingly. We respond below in detail to each of the comments and questions.

Reviewer reports:

Haitao Pan, Ph.D. (Reviewer 1): Please remove the tracking, then submit; otherwise, it is hard to read. Thanks!

Response: Thanks Dr. Haitao Pan. We provide the manuscript with track change in previous round of review based on the journal asked us to make sure all changes to the manuscript are indicated in the text by highlighting or using track changes. I also provide a clean version of
manuscript through Email when the editor asked for it. Hopefully they have contact you for this problem.

I will submit a version without tacking. The comments and questions of previous round of review were also listed below in case you have no records. We really appreciate your time and effort.

Haitao Pan, Ph.D. (Reviewer 1): The topic is very interesting and needy. However, I have several concerns as follows.

(1) The authors detected drug-diagnosis associations from two EMR datasets, one from a children hospital and one from a general hospital. I cannot see the point of why the authors want to involve the general hospital. Since the different datasets with different drug-diagnosis matrices, it will generate different clusters. How to make fair comparisons based on different clusters? And, what kind of conclusions can be inferred from these analyses?

Response: Thank you very much for this concern, because it is one of the key points in this study. As the workflow of the computational method shown in Figure 1, the data from general hospital were used for the comparing study. We know there are different prescription pattern between adult and pediatric population due to many factors such as different administration policies and different diseases incidence in two populations (as shown in Supplemental S1). The different disease spectrum between the children and adults is an important information to understand the difference needs of drug in pediatrics. Understanding the needs of drugs in pediatrics lay the foundation to understand the requirement of pediatric drug and evaluate the safety of pediatric drug. Therefore, a dataset that generated from a general hospital was compared with the dataset generated from a children’s hospital to show the difference prescription patterns and the treatment gap between children and adults. The Figure 2 and Figure 3 were used to show the difference.

Your concern is reasonable and thoughtful. In theoretical, different datasets generated different drug-diagnosis matrices. However, the drug-diagnosis matrices generated from clinical practice are not a randomly drug diagnosis pairs but imply knowledge of both disease and clinical drugs. Drug-diagnosis associations that used to compare all passed the hypergeometric test (Bonferroni-adjusted P<0.05). The statistically significant drug-diagnosis implied the current clinical knowledge that were used in the routine clinical practice. So basically, in a country with similar drug administration policy and drug supplies, the drug-diagnosis matrix will be similar for a targeted population. We used the significant drug-diagnosis pairs to compare the prescription patterns between children and adults in two ways. One is aligning these pair by diagnosis and to differentiate diseases incidence and their associated drugs in two population. The other is
aligning them by drug and to show the drug used volume and their associated diagnosis in two population.

As shown in Figure 2, the clustered heatmap on the top section shows the clustered of 3205 diagnosis used in the children’s hospital. In the first row of heatmap, the red section contains 1137 unique diagnosis that only used in the children’s hospital. The other 2068 diagnoses shown in green are also used in the general hospital. Three other rows of heatmap show the distribution of associated drugs in two populations. The disease incidence and associated drugs from both hospitals were shown in mirror histograms and aligned to the diagnosis in heatmap. Seven groups were classified based on their features. From this figure, readers will understand there are many unique diagnosis (group A) in the children’s hospital do not have associated drugs. There are also some common diseases share by children and adults but with different prescription patterns (group C, D, E, F, G). The clustering heatmap in Figure 2B also explains the different among these diagnosis groups.

The Figure 3 using the drug-diagnosis pairs data but align the heatmap and histogram by drug. In the first row of heatmap, green part contains 513 drugs used in the children’s hospital. The other 672 drugs shown in red were only used in the general hospital. Four drug groups were classified based on this comparing study. From this visualization of the comparing analysis, reader will understand there many drugs (Group B) used in children’s hospital do not have associated diagnosis in the general hospital. The clustering heatmap shown in the Figure 3B also help readers to understand the composition of different drugs groups.

In short, the comparing analysis show the drug treatment gap between the children’s hospital and the general hospital and help readers and policy makers better understand the problem of medication for children. It is also the first quantitatively and visualization approach to show this gap. Furthermore, when we talk about the drug safety in pediatrics, we not only concern about drugs used in pediatrics but also drugs should but are not used in pediatrics due to the safety issue. Without the dataset from the general hospital we cannot show this part of information. We rewrite some sentences in abstract and main text to justify the comparing study.

I guess the thing you real concerned may be the stability of the drug clusters we discussed in the limitation. That drug clusters were generated by the clusterboot function of R (version 3.4.0) and used to discovery the clinical scenario with similar or related drugs. We use the Jaccard Index to assessment of cluster wise stability and identified the best cluster number from 100 resampling and runs. By the way, we do not use this drug clusters for the comparing study. These drug clusters were only used for the pediatric drug safety evaluation.

In this revision, we rewrite some of the sentence and figure legends to clarify this. Thank you very much to help us to improve this critical part of the paper.

(2) The paper focuses on developing a quantitative drug safety level for pediatric population under the detected clusters. It it novel and conveys informative messages, however, the authors should not just show which cluster has more safety issue, but should at least try to explain/reveal
the reason behind the scene; and also have some discussions about the possible approaches to improve the situations (for each cluster).

Response: Thank you very much for this comment. We have added more discussion in this revision regarding the clinical meaning of the drug clusters and the high priority of pediatric drugs requiring. As a method paper, we focused on providing a computational method to evaluation the situation. Improve the drug safety in pediatric is a complex and arduous project that relies on multiple forces such as government, industry and clinical researches. The method proposed in this study is only one step to improve the capability to evaluation this situation objectively. As more and more clinical data were electronically accumulated in daily practice, such method will have more and more application space. For example, US National Institute of Child Health and Human Development (NICHD) that in charge of updating annually priority list of pediatric drugs requiring could using the national EHR to generate such quantitative results to support their decision and evaluated the improvements. The expertise panel in such organization will proposed specific approaches to improve the situations for each drug clusters. Although we don’t have such expertise, we have published the results on an open accessed website (http://kb4md.org:4000/peddrugcluster) and it will help the experts to assess individual drug in different drug clusters under specialized clinical conditions.

We added one more discussion paragraph to discuss this in this revision. Thank you very much for your suggestions.

(3) The general hospital, Shanxi Dayi hospital, should also be approved by it local IRB.

Response: This study was approved by the Institutional Review Board/Ethics Committee of Children’s Hospital of Zhejiang University School of Medicine (Hangzhou, China). The signed Data Use Agreements (DUA) of the anonymized clinical data from Shanxi Dayi hospital were also obtained and several papers have been publicized based on this dataset. All research was performed in accordance with relevant guidelines and regulations.

Thank you very much for this remind. We have mentioned the DUA in this revision.

(4) All figures are challenging to read due to low resolution.

Response: we are sorry about low resolution figures in the emerged pdf file. There are many detail information and data embedded in figures to help reader to understand our method. We will use 300*300 dpi resolution for all the figures in this revision. We also regenerated all the figures to make it clearer. Thank you very much for this comment. The figures that shown in this paper is relative complicated and hard to read. We also revised the figure legends and try our best to make it clear received by our readers.

Thank you very much for your time to review this manuscript and provide all these insights. Your comments helped clarify and improve our paper. We really appreciate it. Regarding the
language issue, we also asked professional English language editing services from American Journal Experts to polish this manuscript.

Darko Krnic (Reviewer 2): Please include all comments for the authors in this box rather than uploading your report as an attachment. Please only upload as attachments annotated versions of manuscripts, graphs, supporting materials or other aspects of your report which cannot be included in a text format.

Please overwrite this text when adding your comments to the authors.

Response: Thank you very much for your encouragement and we really appreciate it. Regarding the concerns about the language, we asked the professional English language editing services from American Journal Experts to improve the language of this manuscript.

Edita Runjic (Reviewer 3): Thank you for asking me to review this paper.

It emphasizes the important question about drug safety in pediatric population and proposes novel method for measuring the drug safety levels. Authors did a good job on presenting their method that could be helpful in addressing the problems of insufficient drug safety data.

Here are some points for the authors to consider:

I had trouble understanding rational behind drug clusters and associated diagnoses shown in table S1. Some diagnoses repeated through different clusters and there were inconsistencies with diagnoses in the same cluster (example: cluster NO 5 should be associated with pneumonia and there is Myocardial injury diagnosis associated with that group; cluster NO 7 should be kidney group and there is Acute upper respiratory infection diagnosis associated with that cluster, etc.). Further explanation about process of obtaining drug clusters and list of drugs for each cluster could be useful.

There are some grammatical and language errors throughout paper.

Response: Thank you very much for your time to review this manuscript.

The associated kernel diagnoses of drug clusters shown in Table S1 were generated from a statistical analysis that based on real-world clinical data. The inconsistencies reflect the real-world scenarios, for example, in cluster NO 5, the Myocardial injury is one of the major and severe complications of Pneumonia. Up to 30% of patient admitted to hospital for Community-acquired Pneumonia develop cardiovascular complications [1]. Myocardial injury is also very common and typical in children with Pneumonia [2]. In cluster NO 7, it’s a kidney group especially focus on Nephrotic Syndrome (NS). Changes in immunity are complex in nephrotic
syndrome, affecting the cellular and humoral balance of specific immunity which may impair response to infection, currently recognized that at least half of relapses are triggered by infection. Recently it is proved that at least 50% of activities in pediatric onset NS are stimulated by a viral upper respiratory tract infection; this may be due to non-specific host response to infection more than to virus itself or their antibody response. [3,4,5]. As the associations were generated from real-world clinical data and passing the Bonferroni-adjusted hypergeometric test. These superficial inconsistencies reflect the real-world clinical scenarios, and it also show exactly the advantage of this computational methods. All the clinical scenarios can be explained by the clinicians and all the drug clusters were reasonable, objective, and do not depend on experts.

We added a paragraph to explain the inconsistencies in clinical scenario in this revision. The source code for how to obtain drug clusters were added in the supplemental source code 1. A supplemental Table S2 were also added to list all drugs for each cluster. A website (http://kb4md.org:4000/peddrugcluster) is also provided to help reader to explore these drugs and associations.

Regarding the concerns about the language, we have asked the professional English language editing services from American Journal Experts to improve the langue of this manuscript for two rounds. We also do our best to check the grammar in this revision.

Reference:


We have two rounds professional editing from American Journal Experts. The certificate verification code from AJE is 54FD-53A9-0D7C-5888-5B8C. Hopefully they provide a service that can meet the requirements of journal publishing.