Reviewer’s report

Title: Dissecting genetic factors affecting phenylephrine infusion rates during anesthesia: a genome-wide association study employing EHR data

Version: 0 Date: 21 May 2019

Reviewer: Matthew Zawistowski

Reviewer’s report:

The paper provides a nice example of electronic health record data used to address an interesting and important clinical question. The study is limited by low samples size (as is often the case with rare traits in EHR) but the authors do provide preliminary statistical and biological evidence for a potentially novel genetic variant influencing effectiveness of phenylephrine during anesthesia. However, none of the statistical associations reported reach the standard levels of genome-wide significance and, importantly, the authors point out that replication studies are necessary to confirm the reported associations.

Major Concerns:

1. The authors analyze their cohort using two distinct analytic strategies. First, the cohort is divided into discovery and replication groups, the results of which are then meta-analyzed (which should be virtually equivalent to analyzing all the samples together). The authors present p-values from each of these analyses (discovery, replication, meta-analysis), and it is difficult at times to understand which analysis they are referring to and how significance should be assessed. It appears the discovery/replication approach did not provide convincing evidence (understandably due to sample size) and most of their discussion of top associations is based on p-values from the meta-analysis. In this case, I do not see how the discovery/replication analysis adds to the paper. I would simply drop the discovery/replication analysis and present results from a single analysis of all samples together. Perhaps include Table S2 as a main paper table to summarize this analysis and provide your follow up text on the three SNPs with suggestive evidence (I imagine they will still have p<1e-7).

Was there a specific rationale for splitting the cohort into discovery and replication groups when sample size was already low? There was little chance of obtaining true genome-wide significance in the full cohort, let alone in the smaller discovery group. Did you perhaps gain access to the replication samples after the analysis of the discovery group?

2. The authors use k-means clustering to identify potential phenotypic subtypes for phenylephrine response. Based on only three features (average infusion rate, mean and standard deviation of SBP), the authors claims they can identify three distinct subtypes.
and name them "Sensitive", "Intermediate", and "Resistant." I have several concerns about this analysis:

i. Line 96: "Empirical observations discern three types of response to phenylephrine IV infusion" Is this statement based on the k-means clustering results from this paper or prior clinical observations? At the point this statement is made in the introduction, it leads the reader to believe that these subtypes are already well established. It needs to be made clear if this statement is a claim of this paper or cite relevant work if it already established.

ii. There is certainly heterogeneity in patient response to phenylephrine, but is there prior biological or clinical indication for distinct subphenotypes? I do not find the three clusters created by k-means clustering to be convincingly different profiles to believe they are distinct. Instead Figure 2A shows a spectrum of phenylephrine rates and mean SBP, with the proposed sub-phenotype definitions being rather arbitrary. Running k-means with k=3 will always create three clusters, but it does not ensure the resulting clusters represent meaningful differences. In fact, the authors directly state "There are no clear boundaries between clusters… indicating the response is not a discrete trait." (Line 231-3). Therefore, declaring patients as "well-classified" (line 225) is contradictory to the presented data.

iii. The k-means clustering is based on a rather limited number of clinical features (infusion rate, mean and SD of SBP). Did the authors try a broader set of features that might actually create more definitive and distinct clusters?

iv. The authors dropped 97 patients from the analysis because they had inconsistent cluster assignments for patients with two or more anesthesia episodes (Line 229). This inconsistency is important because it speaks to whether phenylephrine response is a property of the surgical event or a phenotype of the patient. That is, is patient response to phenylephrine consistent across independent surgical anesthesia encounters?

Having patients with multiple anesthesia episodes presents a really nice opportunity to formally test the consistency of the proposed phenylephrine response sub-phenotypes. How many patients in total had more than one anesthesia episode with phenylephrine? Please provide a complete breakdown on the consistency of clustering (numbers of concordant and discordant cluster assignments for the same patient).

v. The authors claim differences in comorbidity rates among the phenotype subgroups. Please provide a statistical assessment of these differences. Are the differences in prevalences statistically significant given the number of traits you compared across subgroups?

Minor comments:
1. Provide some summary on the number of SBP measurements per patient, and perhaps give a brief description of how/when these measurements are taken during the course of phenylephrine infusion (so someone not familiar can better understand the process and resulting data).

2. An alternative phenotype would seem to be the change in SBP over the course of the infusion as that more directly quantifies the biological response to phenylephrine. Was this measure feasible? Did you consider it?

3. The authors mention that there are patients with multiple surgeries/phenylephrine infusions. How do they deal with these non-independent measurements during the association analysis?

4. Please include the number of SNPs tested as part of the GWAS

5. Line 283: Authors state "We failed to replicate the previously reported association between Thr164Ile and Gly16Arg polymorphisms in the ADRB2 and phenylephrine response." Please provide the association results for these SNPs (rs#, p-values, effect sizes, direction of effect, etc) which might still be valuable to others.

6. Line 266: What is meant by "full" linear regression model? Does this include main effects for age, sex, body weight and their interaction terms with the SNP?

7. As a follow-up to the previous item, I find Table 3 confusing, particularly the SNPxWeight interaction term. Were interaction terms included in the original discovery/replication GWAS of average infusion rate? If yes, how did you assess association of the SNP when it has a main effect and interaction term?

8. The authors include an analysis in which they claim to investigate the predictability of three lead SNPs to differentiate their phenotypic subtypes. The authors performed regression analyses to test for associations between the SNPs and the subtypes, they did not formally assess the ability of the SNPs to predict subtype. The authors should properly describe this work as an "association" analysis or include a formal prediction analysis (e.g. ROC curves, AUC).

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.

Yes

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

Statement on potential review bias
Please complete a statement on potential review bias, considering the following questions:

1. Did you co-author any publication with an author of this manuscript in the last 5 years?
2. Are you currently or recently affiliated at the same institution as an author of this manuscript?

If you can answer no to all of the above, write 'I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria'. If your reply is yes to any, please give details below.

I declare that I did not publish with these authors in the last 5 years and also meet the affiliation criteria

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