Author's response to reviews

Title: Genotype-Informed Calculation of Risk of Coronary Heart Disease Based on Genome-Wide Association Data Linked to the Electronic Medical Record

Authors:

Keyue Ding (ding.keyue@mayo.edu)
Kent R Bailey (baileyk@mayo.edu)
Iftikhar J Kullo (kullo.iftikhar@mayo.edu)

Version: 3 Date: 8 September 2011

Author's response to reviews: see over
Reviewer #2

The authors have responded adequately to most of the comments. It would have been helpful to note the changes in the response to the reviewers itself rather than having to look at the manuscript.

The following are my questions/suggestions

1. The authors state: “A significant proportion of controls underwent ECG testing and those with a study positive for ischemia were excluded...”. It would be useful to mention what they mean by a study positive for ischemia and by this criteria how many were excluded.

The stress ECGs are overread by an exercise physiologist and standard criteria are used to assign ‘positivity’ to the ECG [1]. Based on our prior data for the exercise ECG laboratory, ~2% of patients referred to the lab who do not have known coronary artery disease, end up with positive stress ECG.

2. Details on how the controls were chosen (which is important) are still lacking. Were the controls consecutive? How were they chosen? This should be described.

Over a period of three years (from October 2006 to May 2009), patients without known coronary heart disease referred to the stress ECG lab were approached for enrollment in the study. The recruitment rate was ~80%. Since we used only a single study coordinator, not all consecutive patients that were referred for stress ECG could be recruited. Nonetheless, we do not envisage a systematic bias in the sampling of patients for this study.

We have added this information in the revised manuscript (Page 3, paragraph 2).

Please clarify if there were 1262 controls (which I seem to have noted in the original review, perhaps my error) or 1243 individuals? If there was a change please clarify why?

The number of controls is 1243. In the revised manuscript, we used all patients in the eMERGE network from the five sites for quality control including a formal approach to define European ancestry based on genotype data (see page 4, paragraph 1). An additional 19 patients were excluded after quality control. The quality control in the original manuscript was based on the Mayo cohort and 99% patients were self-reported whites.

3. The first statement in the study participant section “In the Mayo electronic MEdical Records and GEOnomics (eMERGE) cohort, ascertained for a GWAS of peripheral arterial disease, coronary heart disease (CHD) was defined as the presence of the International Classification of Disease-9-Clinical Modification (ICD-9-CM) diagnosis codes for ischemic heart disease including 410.33-414.33, or a history of percutaneous coronary intervention or coronary artery bypass surgery (ICD-9-CM procedure codes 36.10-36.14)” does not read well. What is meant by “ascertained for PAD”?

The aim of the Mayo eMERGE project is to identify common variants associated with peripheral arterial disease, ie, the primary phenotype. We have deleted the term ‘ascertained for a GWAS of peripheral arterial disease’ in the revised manuscript to avoid confusion (Page 3, paragraph 2).

4. I think a separate limitations section will be useful rather than incorporating in a table and discussion text.

Based on the reviewer’s comments, we merged the paragraphs 2 and 3 from the Discussion section into a ‘limitation’ section in the revised manuscript (Page 10, paragraph 3; page 11, paragraph 1).

5. Similarly the advantage of having an EMR is not apparent and would be useful to highlight/discuss. A statement in the background talks of how EMR can be useful to defray phenotyping costs. Was this the case in the current analysis?
The use of already available phenotypes of interest in the EMR can substantially lower costs of phenotyping which remains a bottleneck for genetic studies. We have incorporated these statements into the revised manuscript (Page 2, last paragraph).

What was the reliability of the ICD 9 coding as this is presumably done by different clinicians who have different thresholds for coding disease?

We have previously assessed the accuracy of phenotyping algorithms [2] and refer to this study in the revised manuscript (Page 2, paragraph 2).

6. The authors in response to point # 7, reviewer 3 point out to the figure based on which it is apparent as to where the GRS RR/ OR is <1. However my question was more a general one of the value of a GRS where having <12 risk alleles (of a max possible of 22) was associated with less risk. The fact that most people have up to 12 at risk alleles/SNPs suggest that the GRS is not that powerful. A comment on this would be useful.

Figure 2 is used to illustrate the genotype effects (ie, combined odds ratios and combined relative risks) according to the number of risk alleles. As shown in Figure 2, the 25th and 75th percentile of the combined odds ratio is 0.774 and 1.261, corresponding to 11 and 14 risk alleles, respectively. We have added this statement in the revised manuscript (Page 7, paragraph 2).

References
