Reviewer's report

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

Version: 2 Date: 29 August 2011

Reviewer: Vijay Nambi

Reviewer's report:

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.

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. 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?

3. The first statement in the study participant section “In the Mayo electronic MEdical Records and GEnomics (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”?

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

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

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.
**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

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