Reviewer's report

Title: Adjusting heterogeneous ascertainment bias for genetic association analysis with extended families

Version: 2
Date: 2 February 2015
Reviewer: Alarcon Flora

Reviewer's report:

In the context of association analysis with family-based samples, the authors propose a statistic that takes into account the heterogeneity of ascertainment bias among family members in order to increase the power to identify associated variants.

The authors refer to the method developed by Won and Lange in order to develop a test statistic based on the probability for each family member to be affected calculated with a latent continuous liability. While Won and Lange introduce random effects in their model to account for a familial effect, this paper introduces the liability to take into account an heterogeneous effect of ascertainment bias and then, two test statistics are derived.

Then, the authors compare their two statistics with the WL statistic from Won and Lange in terms of empirical power through a simulation study.

Results show that empirical power estimates with FQLS1 and FQLS2 are close to those with the WL statistic even if the FQLS statistics are more efficient than WL. The performance of FQLS statistics increases with the heritability and the size of the family.

Finally the proposed method is applied to the genome-wide association analysis of Alzheimer's disease.

Major Compulsory revisions:

One of the mean contribution of this work relies on the derivation of the efficiency increasing of the proposed test statistic. This is shown in the supplementary material (Text S1). However, in its present form, the derivation is not correct:

- Several minor mistakes occur that makes the derivation difficult to understand. For instance, line17, a symbol $\mathbb{E}$ is written instead of vertical dots. Line19, $-2$ has to be $+2$. Line21, last equality is wrong it should be written as follow:

$$
\mathbb{E}\left[ \mathbb{E}\left( (y_{ij}-\mathbb{E}(y_{ij}|Y_{-(ij)})) |Y_{-(ij)} \right) \left(\mathbb{E}(y_{ij}|Y_{-(ij)})-\mu\right)^t \right]$$

- More importantly, in this text S1, the mean issue is to prove the inequality on page2 which is not rigorously done in this version. Indeed, the authors have to prove that the last term of the right-hand side of equation line19 is positive definite. This is not proved in the next equation that allows only to conclude that this term is null. Moreover, the equations line21 are not justified: why the authors
use this approximation which is generally not true?

These points have to be fixed before considering possible publication.

Furthermore, ascertainment process was simulated for a half of the sample and an other ascertainment process was simulated for the other half, describing an heterogeneous ascertainment process. It would be interesting to simulate other ascertainment processes in order to assess the robustness of the proposed method. I am not sure to understand the interest of the way of simulating the heterogeneity: why do not simulate part of the sample with at least 1 individual (for example) and the other part with at least 4 affected individuals.

Additionally, the statistical model is not detailed in the paper. This would allow to discuss the differences between this model and the one of Won and Lange. Consequently $\beta$ parameter equal to zero under the null hypothesis, is introduced line 8-page8 without prior definition.

Moreover the authors have written that the power loss attributable to the misspecified prevalence is not substantial. However Table 8 suggests a significant loss of power. Could the authors explain their conclusion? And what happen if $n_{miss}$ is bigger than 3?

Finally I am surprised that $FQLS_2$ give better results even when the probands are well designed. Could the authors discuss this results.

Minor Essential revisions:

- typo for the probability density function line10-page9
- missing "s" line15-page10
- line18-page14: It is Figure 4 in place of Figure 2
- line23-page 14: Figure 6 in place of Figure 4

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

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interest