Referee's comments:
On genetic correlation estimation with summary statistics
from genome-wide association studies

The paper investigates the factors that determine the bias in the cross-trait genetic correlation estimation. A bias-corrected estimator is proposed and tested via simulation/data analysis. In general:

1. I think the results are of general interest to both statistics and genetics communities. I like seeing the discussion of sample size ratio, heritability, screening in cross-trait PRS estimation; it is a useful contribution to the GWAS literature. The presentation, however, is made needlessly complicated by not fully explaining the notations at the first appearance. For example, in the introduction (line 41 of page 3), the authors stated the main results in terms of (n, p, h^2) without introducing n, h^2. Also confusing is that the proposal of new method was tangled with the technical depository (e.g. line 23 on page 11), but ultimately, I think they should be separated.

2. The simulation was a little underwhelming. Basically, you simulated polygenic traits from exactly the same models you fitted, and validated the derived bias. The results do not seem to bring anything new, except for telling the difference between finite-sample vs. asymptotical properties. A more useful experiment would be assessing the robustness of the estimator to model misspecification. The framework that authors proposed requires a series of assumptions (which I will discuss them below), and I would like to see how the PRS estimator performs in more practical settings. For example, population structure is a well-known issue in GWAS (Price et al. Nat Genet 2006, H. M. Kang et al, Nat Genet 2011). It would be interesting to assess the cross-trait PRS bias in the presence of population stratification, admixture, and cryptic relatedness etc.

3. Analysis framework. The claimed asymptotic is built on a number of conditions. I did not find the discussion of these conditions very illuminating. For example, Condition 1 requires $\frac{m}{n} \to \gamma_0 > 0$. It seems to imply that the number of causal SNPs m has to increase with sample size n of individuals? Condition 2 imposes joint distribution on (α_i, η_i) and on (α_j, β_j), but no assumption on (β_j, η_j)? At least, I would like to see more discussions on these conditions, which are necessary, which are only for technical convenience, and their implication in the context of GWAS.

So I think this is good. I cannot quite figure out whether it rises to the JASA level. I will let editors decide.

Minor comments:

1. Page 5, condition 1. The definitions of n, p are not introduced.
2. Page 6, Definition 1. Both the vectors $\boldsymbol{\alpha}, \boldsymbol{\eta}$ have non-zero means. So, by "correlation", the authors really mean the "inner product"?
3. Page 9, line 3. "$\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\beta}}$, $\ldots, \hat{h}^2_\eta$ are available." Any assumptions (consistency, efficiency) on

these estimators? Are they estimated from a separately independent sample or the discovery/target sample?

4. Figures 1 and 2 are simulated from $h^2_\alpha = h^2_\beta = \cdots = 1$. So, all these traits are perfectly heritable without error variance?

5. Page 20, line 22. "...which the SNP genotypes are independently sampled from $\{0, 1, 2\}$..". It would be interesting to simulate genotypes from samples with population structure.

6. The current simulation and data analysis focus more on the RPS estimation. The authors briefly commented on the testing/inference properties on top of page 12. I would like to see numerical experiment in that regard.

7. The supplementary software only has the code for plotting figures. It would be desirable to have a stand-alone package for RPS estimation/inference that makes new methods ready for use by practitioners.