THE UNIVERSITY OF WARWICK

A Method for Fast Whole-brain Aggregate Heritability Estimation

Xu Chen[1], Matthew F. Glasser[2], David C. Van Essen[2], Thomas E. Nichols[1,3], Steve Smith[3]

[1] Department of Statistics, University of Warwick, Coventry, UK,

[2] Department of Anatomy & Neurobiology, Washington University at St. Louis, St. Louis, MO, USA,

[3] FMRIB Centre, Oxford University, Oxford, UK

email: xu.chen@warwick.ac.uk, t.e.nichols@warwick.ac.uk

Introduction

Heritability (the proportion of variability attributable to genetic sources) is a vital quantitative genetic measure. Non-zero heritability is needed to certify a trait as a "phenotype". Heritability can also be used as a general measure of biological validity, e.g. ranking different pre-processing techniques by heritability of the resulting phenotype. While such comparisons can be done element-wise over the phenotypes (e.g. by voxels or surface elements), a whole-brain summary of heritability can simplify the comparisons.

We propose a simple measure of aggregate heritability that is easy to compute and involves no ACE model fitting. We derive analytical results that show this aggregate measure is closely related to the average of element-wise heritability. We validate the analytical results with simulations and illustrate the method on 22 different phenotypes based on the data of 196 subjects from the publicly released Human Connectome Project (HCP) [1], comparing the ranking of this fast aggregate method to the slower traditional ACE-based estimates of average heritability (see also applications in [2][3]).

Methods

We arrange data from an imaging twin study into a subject-by-element data matrix, with one row for each subject and one column for each data element (e.g. voxel, surface element, etc). Conventional heritability analyses work on a single univariate phenotype, here a single column. Our method proceeds by computing the correlation coefficient between rows, over phenotypic elements. That is, each subject pair generates one correlation coefficient r. If $\bar{r}_{\text{MZ}}$ and $\bar{r}_{\text{DZ}}$ are the average of the correlations for monozygotic (MZ) and dizygotic (DZ) twin pairs respectively, we propose AggHe $= 2\times(\bar{r}_{\text{MZ}}-\bar{r}_{\text{DZ}})$ as the aggregate estimate of heritability, analogous to Falconer's heritability estimate $h^2 = 2\times(\rho_{\text{MZ}}-\rho_{\text{DZ}})$.

Box 1(A) provides expressions for the expected values of $\bar{r}_{\text{MZ}}$, $\bar{r}_{\text{DZ}}$ and $\bar{r}_{\text{UN}}$ (unrelated individuals), which are exact up to a routine approximation ($\mathbb{E}[\text{A}/\text{B}] \approx \mathbb{E}[\text{A}]/\mathbb{E}[\text{B}]$ for random variables A and B). Unrelated individuals have non-zero mean correlation, dependent on the variability of the true mean at each element. Box 1(B) shows that if data is mean-centered (columns demeaned), then $\bar{r}_{\text{UN}} = 0$. Box 1(C) shows that if data is mean-centered & variance-normalised, $\bar{r}_{\text{MZ}}$ and $\bar{r}_{\text{DZ}}$ differ by half the average heritability over phenotypic elements, though shifted and scaled by $\overline{\text{ERV}}/2$ and $1-\overline{\rho^{\text{P}}}$ (see box caption for the definitions). The validity of these analytical results for AggHe is assessed with simulation.

We demonstrate the ability of AggHe to rank phenotypes. Using analyses described in [4][5], the HCP data of brain structure, function and connectivity were pre-processed in different ways; see [3] for more details. The proposed aggregate heritability method is used to rank the heritability of 22 phenotypes. As a comparison, we use an ACE model fit [6] on each element of each of these phenotypes, and then compute the variance-weighted (with weights of $\sigma_i^2/\overline{\sigma^2}$, for element i in Box 1) and unweighted (with equal weights) mean measures of heritability. We use permutation and bootstrap inference to get p-values and CI's for aggregate and mean heritability measures.

(A)
$$\begin{cases} \bar{\text{r}}_{\text{MZ}} = \dfrac{\frac{\sigma_\mu^2}{\sigma^2}+\overline{\text{wh}^2}+\frac{\text{J}}{\text{J}-1}\overline{\text{wc}^2}-\overline{\text{wERV}}}{\frac{\sigma_\mu^2}{\sigma^2}+1-\overline{\text{w}\rho^{\text{P}}}} \\ \bar{\text{r}}_{\text{DZ}} = \dfrac{\frac{\sigma_\mu^2}{\sigma^2}+\frac{1}{2}\overline{\text{wh}^2}+\frac{\text{J}}{\text{J}-1}\overline{\text{wc}^2}-\frac{1}{2}\overline{\text{wERV}}}{\frac{\sigma_\mu^2}{\sigma^2}+1-\overline{\text{w}\rho^{\text{P}}}} \\ \bar{\text{r}}_{\text{UN}} = \dfrac{\frac{\sigma_\mu^2}{\sigma^2}}{\frac{\sigma_\mu^2}{\sigma^2}+1-\overline{\text{w}\rho^{\text{P}}}} \end{cases}$$ for the original data,

(B)
$$\begin{cases} \bar{\text{r}}_{\text{MZ}} = \dfrac{\overline{\text{wh}^2}+\frac{\text{J}}{\text{J}-1}\overline{\text{wc}^2}-\overline{\text{wERV}}}{1-\overline{\text{w}\rho^{\text{P}}}} \\ \bar{\text{r}}_{\text{DZ}} = \dfrac{\frac{1}{2}\overline{\text{wh}^2}+\frac{\text{J}}{\text{J}-1}\overline{\text{wc}^2}-\frac{1}{2}\overline{\text{wERV}}}{1-\overline{\text{w}\rho^{\text{P}}}} \\ \bar{\text{r}}_{\text{UN}} = 0 \end{cases}$$ for mean centering, and

(C)
$$\begin{cases} \bar{\text{r}}_{\text{MZ}} = \dfrac{\overline{\text{h}^2}+\frac{\text{J}}{\text{J}-1}\overline{\text{c}^2}-\overline{\text{ERV}}}{1-\overline{\rho^{\text{P}}}} \\ \bar{\text{r}}_{\text{DZ}} = \dfrac{\frac{1}{2}\overline{\text{h}^2}+\frac{\text{J}}{\text{J}-1}\overline{\text{c}^2}-\frac{1}{2}\overline{\text{ERV}}}{1-\overline{\rho^{\text{P}}}} \\ \bar{\text{r}}_{\text{UN}} = 0 \end{cases}$$ for mean centering & variance-normalisation.

Box 1: The expected values of correlation coefficient for MZ twins ($\bar{r}_{\text{MZ}}$), DZ twins ($\bar{r}_{\text{DZ}}$) and unrelated individuals ($\bar{r}_{\text{UN}}$). Here, σ_μ^2 is the variance of element-wise averages, σ^2 is the element-wise variance (σ_i^2 for element i), h^2 is the heritability, c^2 is the common environmental factor, ERV is the endophenotype ranking value (the product of each trait's square-root heritability and the genetic correlation), and ρ^{P} is the phenotypic correlation. The notations $\overline{\text{w}*}$ and $\overline{*}$ represent the variance-weighted and unweighted means across all phenotypic elements respectively.

Results

Fig. 1 shows simulation results, plotting bias of AggHe relative to the variance-weighted (top) and unweighted (middle) mean summaries, and the standard deviation for AggHe (bottom). While demeaning-only experiences some bias, the results of original data and demeaning & variance-normalisation have comparatively good bias with small variance.

In Fig. 2, real data results show AggHe compared to summary measures of heritability using 22 phenotypes. There is a monotonic relationship found between AggHe and these summaries for both estimates (top) and p-values (bottom). As expected we found a closer relationship between AggHe and the variance-weighted mean than with the unweighted mean.

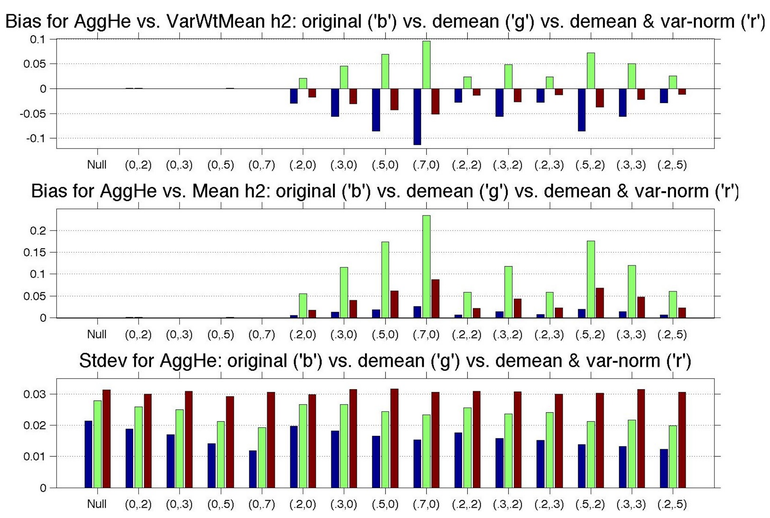

Figure 1: The results of applying demeaning-only and demeaning & variance-normalisation are compared with the result from the original data for AggHe. 1000 simulations for 15 ACE settings with the sample size $n = 58$ ($n_{\text{MZ}} = 32,\ n_{\text{DZ}} = 26$) and 1000 phenotypic elements were used to evaluate AggHe and compare it with the mean summaries. The phenotypic variance σ^2 varied by the element index, ρ^{P} varied over all element pairs and ERV $= 0$.

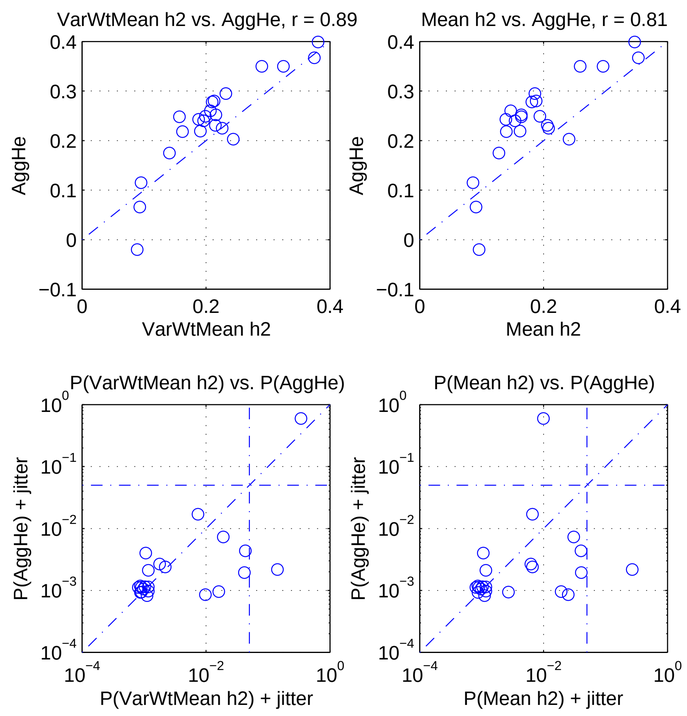

Figure 2: Comparison between AggHe measure and variance-weighted and unweighted mean summaries. The uniformly distributed jitter ($\mathbb{U}(0.8, 1.2)$) was used to allow the visualisation of the many p-values near 10^{-3}.

Conclusions

Our simulations indicate that the analytical results were relatively accurate for aggregate heritability with original data and after demeaning & variance-normalisation. Using real data, the extremely fast aggregate heritability is highly similar to the traditional (more computationally intensive) mean heritability summaries obtained by fitting an ACE model.

Acknowledgement

Supported by the Human Connectome Project (NIH Neuroscience Blueprint grant U54MH091657 and the McDonnell Center for Systems Neuroscience).

References

[1] Van Essen, DC, et al. (2013), *NeuroImage*, 80: 62-79. [2] Smith, SM, et al. (2014), *OHBM*, Poster 1719. [3] Van Essen, DC, et al. (2014), *OHBM*, Poster 3402. [4] Glasser, MF, et al. (2013), *NeuroImage*, 80:105-124. [5] Smith, SM, et al. (2014), *OHBM*, Poster 1718. [6] Chen, X, et al. (2014), *OHBM*, Poster 3410.