A simple, consistent estimator of heritability for genome-wide association studies

Armin Schwartzman¹, Andrew J. Schork², Rong Zabolcki¹, and Wesley K. Thompson^{1,2}

¹Division of Biostatistics, University of California, San Diego, La Jolla, CA

²Institute of Biological Psychiatry, Mental Health Center St. Hans, Mental Health Services Copenhagen, Roskilde, Denmark

October 16, 2017

Abstract

Analysis of genome-wide association studies (GWAS) is characterized by a large number of univariate regressions where an outcome, a quantitative trait, is regressed on hundreds of thousands to millions of genomic markers, i.e. single-nucleotide polymorphism (SNP) counts, one marker at a time. Assuming a linear model linking the markers to the outcome, this article proposes an estimator of the heritability of the trait, defined here as the fraction of the variance of the trait explained by the genomic markers in the study. The estimator, called GWAS heritability (GWASH) estimator, is easy to compute, highly interpretable, and is consistent as the number of markers and the sample size increase. More importantly, it can be computed from summary statistics typically reported in GWAS, not requiring access to the original data. The estimator takes full account of the linkage disequilibrium (LD) or correlation between the SNPs in the study through moments of the LD matrix, estimable from auxiliary datasets. Unlike other proposed estimators in the literature, the precision of the estimate is obtainable analytically, allowing for power and sample size calculations for heritability estimates.

1 Introduction

Modern statistics is increasingly seeing situations where there is an interest in understanding how an outcome measured on a large number of subjects is related to a much larger number of predictors. Genome-wide association studies (GWAS) are an excellent example of this, where the outcome is an observed trait or phenotype and the predictors are single-nucleotide polymorphisms (SNPs), each captured by the number of copies of the reference allele (0, 1 or 2). The sample size in typical GWAS may be in the order of tens to hundreds of thousands, while the number of SNPs may be ten to one hundred times as large, in the order of millions.

Suppose the outcome is related to the predictors by a linear model. Since the number of coefficients in the model is larger than the sample size, this is an under-determined problem and it is impossible to estimate the coefficients without additional assumptions. Low dimensional summaries, however, are estimable. In particular, in this paper we focus on the heritability, defined as the proportion of variance of the outcome explained by the measured predictors. In the GWAS literature this is often referred to as chip heritability or SNP heritability [24]. Heritability is an important parameter in that it helps quantify the proportion of the observed outcome

that can be predicted from genetic factors. In other words, the estimation of heritability is often the first step in investigating the genetic basis of traits and diseases.

A prototypical GWAS aims to identify individually important genetic factors by regressing an outcome variable on each SNP, one at a time, selecting only the most stringently significant SNPs ($p < 5 \times 10^{-8}$) as discoveries. The basic GWAS approach has become a mainstay in the genetic analysis of complex traits. Thousands of studies have been performed and tens of thousands of candidate causal variants have been cataloged for all variety of trait and disease (www.ebi.ac.uk/gwas/, [16]). In part due to funding institution data sharing mandates, to increase transparency and to fuel post-hoc and secondary analysis of GWAS results [18], per-SNP univariate regression statistics (beta coefficients, t -statistics, p -values, standard errors, etc.) are now regularly published along with GWAS articles. While privacy concerns often prevent the sharing of subject level genotypes and phenotypes, these summary statistics are readily available for hundreds of individual GWAS studies (e.g., www.ebi.ac.uk/gwas/downloads/summary-statistics).

It is of interest, therefore, to develop an estimator of heritability that can provide accurate estimates using only summary statistics from GWAS. Computational efficiency is another desired property given the large size of the data. And, as with any estimation procedure, interpretability is also desired in order to gain further insights into the data. For example, we wish to understand how heritability is affected by the correlation between the predictor genomic markers, called linkage disequilibrium (LD) in the context of GWAS.

In this paper, we propose an estimator called GWAS heritability (GWASH) estimator. The estimator is based on the variance-fraction estimator in [6] and is astonishingly simple. For a GWAS with m predictors and n independent subjects, the estimator is

$$\hat{h}_{\text{GWASH}}^2 = \frac{\hat{m}_{\text{eff}}}{n} (s^2 - 1), \quad (1)$$

where s^2 is the empirical variance of the GWAS t -statistics (up to a small transformation) and $\hat{m}_{\text{eff}}$ is the estimated effective number of SNP predictors. The latter is equal to the number of predictors m divided by a factor $\hat{\mu}_2$, an estimate of the second spectral moment of the LD matrix that captures in a single number the effect of LD: the higher the LD, i.e. the higher the correlation between the predictors, the lower their effective number.

The GWASH estimator is not only easy to remember and compute as a simple formula. It also has an interpretation as being proportional to excess empirical variance of the univariate t -statistics with respect to the complete null hypothesis of independence between the outcome and the predictors, in which case the empirical variance is about 1. The empirical variance s^2 is in itself an intuitive quantity that summarizes the strength of the relationship between the predictors and the outcome, and has been used as a simple measure of signal to noise ratio in GWAS contexts [19]. Thus, the proposed estimator has the nice property that it increases linearly with enrichment, where the proportionality constant depends on LD.

Moreover, the formula dictates that LD affects the estimation of heritability as a scaling factor, yielding a definition of the effective number of SNPs involved. Computing the factor $\hat{\mu}_2$ to find the effective number of predictors is the only relatively difficult part of the estimation. The factor $\hat{\mu}_2$ estimates $\mu_2 = \text{tr}(\tilde{\Sigma}^2)/m$, where $\tilde{\Sigma}$ is the correlation matrix of the predictors, the LD matrix. As a first approximation, the patterns of correlations among SNPs can be taken as a feature of a given population and estimated from publicly available data resources such as the 1000 genomes project (1KGP) [10] (<http://www.internationalgenome.org/>). This approach has been reasonable when the reference sample plausibly represents the same population as the GWAS sample in contexts including imputation [13], heritability estimation (e.g., [2]), functional fine-mapping (e.g., [21]) and various post-hoc burden tests (e.g., [5]). We

propose an efficient way of calculating the factor $\hat{\mu}_2$ so that the entire LD matrix need not be computed.

From a theoretical point of view, following [6], we show that the GWASH estimator is consistent as m and n increase to a limiting fixed ratio, which could be greater than 1. We also provide a formula for estimating the asymptotic standard error. Simulations are used to show the statistical and computational performance of the estimator in non-asymptotic settings.

As an alternative method, Linkage Disequilibrium Score (LDSC) regression [2] has become the most popular approach for estimating heritability from summary statistics. LDSC estimates heritability by regressing squared z-scores (per SNP univariate regression t or Wald statistics) on corresponding “LD Scores,” defined as estimates of the sum of squared correlations for a given SNP and all others. While an effective and computationally efficient approach, LDSC was not derived from a well-specified generative model and relies on a number of heuristics, including binning LD scores, censoring outlying values, and empirical approximations to standard errors. These features are difficult to consider analytically and limit opportunities for power analyses. To compare with LDSC, we here consider a stylized version of that estimator without binning, bootstrap and other elements that allows for an easier comparison both analytically and in small scale simulations. Our simulations show that the GWASH estimator performs similarly to the LDSC regression and is closely related analytically under some conditions, suggesting avenues to further improve the theoretical foundation for both methods.

In the rest of the paper we provide the theoretical foundation for the GWASH estimator, proving its consistency and deriving its asymptotic standard error. We confirm these properties in simulations. We illustrate the feasibility of the GWASH estimator in several real GWAS datasets and show how the asymptotic standard error formula can be used for performing power analysis in prospective GWAS.

2 GWAS

2.1 The classic polygenic model

Suppose that a continuous outcome variable or phenotype is measured together with a panel of genotype markers at m loci for each of n independent subjects. Let y_i and $\vec{x}_i = (x_{i1}, \dots, x_{im})$ denote the outcome and genomic panel for subject $i = 1, \dots, n$. According to the classic polygenic model [9, 15], the outcome is generated according to the linear model

$$y_i = \vec{x}_i \beta + \varepsilon_i, \quad i = 1, \dots, n, \quad (2)$$

where the error terms ε_i are independent with mean 0 and variance σ^2 . This model may also be written in matrix form as

$$\mathbf{y} = \mathbf{X} \beta + \boldsymbol{\varepsilon}, \quad (3)$$

where $\mathbf{y} = (y_1, \dots, y_n)^T$, $\beta = (\beta_1, \dots, \beta_m)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ and $\mathbf{X}$ is the regression matrix with rows $\vec{x}_i$, $i = 1, \dots, n$. It is also useful to write the regression matrix in terms of its columns as $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$.

True to the sampling scheme, we shall consider the genomic panels $\vec{x}_i$ to be randomly drawn from the population together with the associated phenotypes. Let $\Sigma = \text{Cov}(\vec{x}_i)$ denote the $m \times m$ covariance matrix between genomic markers in the underlying population. The corresponding correlation matrix, which we shall denote $\tilde{\Sigma} = \text{Cor}(\vec{x}_i)$, is often called the linkage disequilibrium (LD) matrix. For simplicity, our model does not explicitly include other fixed covariates (e.g. age, gender, ethnicity factors, etc.) but rather we shall assume that these other

covariates have been regressed away. The interpretation of the coefficients and the heritability shall be conditional on having accounted for those other covariates and is the same as if those covariates had been included in the full model.

Similarly, rather than including an intercept term, we may equivalently assume that the vector $\mathbf{y}$ and the columns $\mathbf{x}_1, \dots, \mathbf{x}_m$ of $\mathbf{X}$ have been centered by subtracting the vector average, so that $\mathbf{1}^T \mathbf{y} = 0$ and $\mathbf{1}^T \mathbf{x}_j = 0$ for $j = 1, \dots, m$. A nice consequence of centering is that, for the centered data, $E(\mathbf{y}) = 0$ and $E(\mathbf{X}) = 0$, where the expectation is taken with respect to the population distribution. Hence, the model (2) or (3) have no intercept term.

2.2 Heritability

The heritability h^2 is defined as the variance explained by the predictors in model (3). Specifically, model (3) has the variance decomposition

$$\text{Var}(y_i|\boldsymbol{\beta}) = E(\boldsymbol{\beta}^T \vec{\mathbf{x}}_i \vec{\mathbf{x}}_i \boldsymbol{\beta}|\boldsymbol{\beta}) + E(\varepsilon_i^2) = \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta} + \sigma^2. \quad (4)$$

since $E(\vec{\mathbf{x}}_i) = 0$. Let $\tau^2 = \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}$. The heritability is the quantity [8, 15]

$$h^2 = \frac{\tau^2}{\tau^2 + \sigma^2}. \quad (5)$$

Note that in this model the vector $\boldsymbol{\beta}$ is fixed and arbitrary, with no pre-specified distribution. The model places no restrictions on the distribution of model coefficients as long as they yield the proper heritability. Thus, as opposed to other methods such as [24, 2, 25], no distributional assumptions are required on $\boldsymbol{\beta}$ in order to estimate heritability, an important point given recent debate in the literature [22].

2.3 GWAS univariate regressions

In GWAS, the vector of SNP effects $\boldsymbol{\beta}$ is estimated by univariate regression coefficients. Since $\mathbf{y}$ and the columns $\mathbf{x}_j$ are assumed centered, there is no need to fit an intercept term and the slope parameters β_j for each SNP $j = 1, \dots, m$ are estimated via

$$\hat{\beta}_j = (\mathbf{x}_j^T \mathbf{x}_j)^{-1} \mathbf{x}_j^T \mathbf{y} = \|\mathbf{x}_j\|^{-2} \mathbf{x}_j^T \mathbf{y}. \quad (6)$$

The univariate regression estimates are typically converted into t-scores by dividing by an estimate of standard error at each SNP. For each $j = 1, \dots, m$, the residual variance is

$$\hat{\sigma}_j^2 = \frac{1}{n-2} \|\mathbf{y} - \mathbf{x}_j \hat{\beta}_j\|^2. \quad (7)$$

yielding the t-score

$$t_j = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}_j^2 (\mathbf{x}_j^T \mathbf{x}_j)^{-1}}} = \frac{\|\mathbf{x}_j\| \hat{\beta}_j}{\hat{\sigma}_j}. \quad (8)$$

The goal is to produce an estimator of heritability that relies on the above so-called summary statistics $\hat{\beta}_j$, $\hat{\sigma}_j^2$ and t_j , $j = 1, \dots, m$. We describe the heritability estimator in general in Section 3 and return to the summary statistics in Section 5.1.

3 The Dicker estimator

To better describe the derivation of the GWASH estimator, we first discuss the estimator proposed by [6]. Addressing the high-dimensional case where m is greater than n , [6] proposes an estimator of the fraction of variance explained by $\mathbf{X}$ in model (3) when the vector of coefficients β is fixed. While not called heritability there, this fraction is the same as the heritability defined in (5). Since ordinary least squares methods fail when $m > n$, Dicker’s estimator is based instead on a clever use of the method of moments. Dicker proposes two forms of the estimator depending on whether the covariance matrix Σ , typically unknown, is estimable or not.

3.1 The Dicker estimator for estimable covariance

An estimable covariance matrix Σ presumes the existence of a norm-consistent positive definite estimator $\hat{\Sigma}$, despite the dimension m being larger than the sample size n . An example of an estimable covariance matrix is a diagonal Σ , so that the columns of $\mathbf{X}$ are uncorrelated but have different variances. Other examples include matrices where the correlation structure is captured by a fixed number of parameters, such as autoregressive (AR) and exchangeable correlation models.

Proposition 1. *Let $\hat{\Sigma}$ denote the estimator of Σ . The Dicker estimator of h^2 for estimable covariance [6, Sec. 4.1], denoted here $\hat{h}_I^2$, can be written (in our notation) as*

$$\hat{h}_I^2 = \frac{m}{n} \left(\frac{\|\hat{\Sigma}^{-1/2} \mathbf{X}^T \mathbf{y}\|^2}{m \|\mathbf{y}\|^2} - 1 \right), \quad (9)$$

where n is replaced by $n - 1$, owing to the centering of $\mathbf{y}$ and the columns of $\mathbf{X}$ [6, Sec. 1].

The estimator (9) requires a consistent estimator $\hat{\Sigma}$ of the covariance matrix Σ , which is not available without further assumptions. The sample covariance matrix

$$\mathbf{S} = \frac{1}{n - 1} \mathbf{X}^T \mathbf{X}, \quad (10)$$

whose entries are the sample covariances of the columns of $\mathbf{X}$, is an unbiased estimator of Σ , satisfying $E(\mathbf{S}) = \Sigma$. It is, however, not norm-consistent because the dimension m is larger than the sample size n .

Assuming that the true correlation is nonzero only close to the diagonal, as it is often assumed in GWAS, consistent estimators may be obtained, for example, by banding the sample covariance matrix [1, 3]. Even so, the estimator (9) requires computation of the inverse square root $\hat{\Sigma}^{-1/2}$. This is computationally taxing for the typical large matrix size m in GWAS in the order of magnitude of a million. Dicker’s estimator for unestimable covariance avoids this problem.

3.2 The Dicker estimator for unestimable covariance

When a model for Σ is not sufficiently specified to be estimable, [6] offers another form of the estimator that replaces estimation of Σ by estimation of its first few moments.

Proposition 2. *The Dicker estimator of h^2 for unestimable covariance [6, Sec. 4.2], denoted here $\hat{h}_{II}^2$, can be written (in our notation) as*

$$\hat{h}_{II}^2 = \frac{m\hat{m}_1^2}{n\hat{m}_2} \left(\frac{\|\mathbf{X}^T \mathbf{y}\|^2}{m\hat{m}_1 \|\mathbf{y}\|^2} - 1 \right) \quad (11)$$

where

$$\hat{m}_1 = \frac{1}{m} \text{tr}(\mathbf{S}), \quad \hat{m}_2 = \frac{1}{m} \text{tr}(\mathbf{S}^2) - \frac{m}{n-1} \hat{m}_1^2, \quad (12)$$

and $\mathbf{S}$ is the sample covariance matrix (10).

Proposition 2 of [6] states that if the entries of $\mathbf{X}$ and $\boldsymbol{\varepsilon}$ are Gaussian and $\boldsymbol{\Sigma}$ is not too far from the identity matrix (technical details omitted here), then $\hat{h}_{II}^2$ is approximately Gaussian with mean h^2 and variance

$$\frac{\psi_{II}^2}{n} = \frac{2}{n} \left(\frac{mm_1^2}{nm_2} + 2 \frac{m_1 m_3}{m_2^2} h^2 - h^4 \right), \quad (13)$$

for large m and n such that m/n is bounded, where

$$m_1 = \frac{1}{m} \text{tr}(\boldsymbol{\Sigma}), \quad m_2 = \frac{1}{m} \text{tr}(\boldsymbol{\Sigma}^2), \quad m_3 = \frac{1}{m} \text{tr}(\boldsymbol{\Sigma}^3). \quad (14)$$

By the commutative property of the trace, it can be shown that the quantities in (14) correspond to the first, second and third moments of the eigenvalues of $\boldsymbol{\Sigma}$. In that sense they can be called *spectral moments*.

An estimate of standard error for $\hat{h}_{II}^2$ can be obtained as the square root of (13) by plugging in the estimate of h^2 and those of m_1 and m_2 given by (12). As an estimate of m_3 , Dicker [6, Remark 12] suggests

$$\hat{m}_3 = \frac{1}{m} \text{tr}(\mathbf{S}^3) - \frac{3m}{n-1} \hat{m}_1 \hat{m}_2 - \frac{m^2}{(n-1)^2} \hat{m}_1^3. \quad (15)$$

4 The GWASH estimator

In GWAS, it is feasible to implement Dicker's estimator (11) if the entire dataset composed of $\mathbf{X}$ and $\mathbf{y}$ is available. However, often only GWAS summary statistics are available. The GWASH estimator is essentially a modification of the Dicker estimator where the columns of $\mathbf{X}$ are standardized. This standardization allows writing the estimator in terms of the correlation scores defined next, which easily translate into summary statistics.

4.1 Correlation scores

Let $\tilde{\mathbf{y}} = \sqrt{n-1} \mathbf{y} / \|\mathbf{y}\|$ be the standardized vector $\mathbf{y}$ so that $\|\tilde{\mathbf{y}}\|^2 / (n-1) = 1$. Similarly, let $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_m)$ be the result of standardizing the matrix $\mathbf{X}$ by columns, so that the standardized columns $\tilde{\mathbf{x}}_j = \sqrt{n-1} \mathbf{x}_j / \|\mathbf{x}_j\|$ satisfy $\|\tilde{\mathbf{x}}_j\|^2 / (n-1) = 1$, for $j = 1, \dots, m$. Because of the original centering, $\mathbf{1}^T \tilde{\mathbf{y}} = 0$ and $\mathbf{1}^T \tilde{\mathbf{x}}_j = 0$.

The main idea of the GWASH estimator is to replace $\mathbf{X}$ and $\mathbf{y}$ in (11) by their standardized versions $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$. Because (11) depends on the summary statistic $\|\mathbf{X}^T \mathbf{y}\|^2$, to become $\|\tilde{\mathbf{X}}^T \tilde{\mathbf{y}}\|^2$, it is convenient here to define what we call the *correlation scores*

$$u_j = \frac{1}{\sqrt{n-1}} \tilde{\mathbf{x}}_j^T \tilde{\mathbf{y}} = \sqrt{n-1} \frac{\mathbf{x}_j^T \mathbf{y}}{\|\mathbf{x}_j\| \|\mathbf{y}\|}, \quad j = 1, \dots, m, \quad (16)$$

or in vector form,

$$\mathbf{u} = \frac{1}{\sqrt{n-1}} \tilde{\mathbf{X}}^T \tilde{\mathbf{y}}.$$

The score u_j may be called correlation score because it is equal to $\sqrt{n-1}$ times the sample correlation between $\mathbf{x}_j$ and $\mathbf{y}$. The reason for the multiplying factor is that, under the null hypothesis of no heritability ($h^2 = 0$) so that $\mathbf{x}_j$ and $\mathbf{y}$ are independent, the score (16) is asymptotically normal with mean zero and variance one. In this sense it plays the role of a z-score.

4.2 The LD matrix

By standardization, we may define the sample covariance matrix of the columns of $\tilde{\mathbf{X}}$,

$$\tilde{\mathbf{S}} = \frac{1}{n-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}. \quad (17)$$

By definition, this is the sample correlation matrix with ones on the diagonal and can be referred to as the sample LD matrix.

Let $\tilde{\Sigma} = \text{Cor}(\tilde{\mathbf{x}}_i)$ be the population correlation matrix corresponding to the covariance matrix Σ . Analogous to (14), we can define the first three spectral moments of $\tilde{\Sigma}$ by

$$\mu_1 = \frac{1}{m} \text{tr}(\tilde{\Sigma}) = 1, \quad \mu_2 = \frac{1}{m} \text{tr}(\tilde{\Sigma}^2), \quad \mu_3 = \frac{1}{m} \text{tr}(\tilde{\Sigma}^3). \quad (18)$$

These quantities capture the total effect of LD between the genomic markers. If the genomic markers are independent with $\tilde{\Sigma} = \mathbf{I}$, then $\mu_2 = \mu_3 = 1$; otherwise both moments are greater than 1.

4.3 The GWASH estimator from subject-level data

The GWASH estimator is defined as a modification of the Dicker estimator where: 1) $\mathbf{X}$ and $\mathbf{y}$ in (11) are replaced by their standardized versions $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$; 2) the moment estimators $\hat{m}_1$, $\hat{m}_2$ and $\hat{m}_3$ in (11), (12) and (15) are replaced by moment estimators based on the correlation matrix (17) instead. Denote the latter estimators by $\hat{\mu}_1 = 1$, $\hat{\mu}_2$ and $\hat{\mu}_3$; details on these are given in section 4.4 and 5.3 below.

Performing the replacements outlined above yields the expression

$$\hat{h}_{\text{GWASH}}^2 = \frac{m}{n\hat{\mu}_2} \left(\frac{\|\tilde{\mathbf{X}}^T \tilde{\mathbf{y}}\|^2}{m(n-1)} - 1 \right). \quad (19)$$

However, this expression can be written succinctly in terms of the correlation scores. We may now define our estimator.

Definition 1. The GWAS heritability (GWASH) estimator is given by

$$\hat{h}_{\text{GWASH}}^2 = \frac{m}{n\hat{\mu}_2} (s^2 - 1), \quad (20)$$

where s^2 is the empirical second moment of the correlation scores:

$$s^2 = \frac{1}{m} \|\mathbf{u}\|^2 = \frac{1}{m} \sum_{j=1}^m u_j^2, \quad (21)$$

and $\hat{\mu}_2$ is an estimator of μ_2 in (18).

The GWASH estimator depends on the data only through two summary statistics, s^2 and $\hat{\mu}_2$. Under the null hypothesis of no heritability ($h^2 = 0$), $s^2 \rightarrow 1$ for large m by the law of large numbers. Thus, (20) expresses the estimate of heritability as proportional to the excess variance of the scores with respect to the null variance 1.

The quantity $\hat{\mu}_2$ contains all the necessary information about the correlation between the predictors. From (18), if the predictors are independent, $\mu_2 = 1$. Otherwise, $\mu_2 > 1$. This implies that ignoring LD causes overestimation of the heritability. Taking account of LD is equivalent to using a smaller number of predictors. In this sense, we may define $\hat{m}_{\text{eff}} = m/\hat{\mu}_2$ as the estimated effective number of markers and rewrite (20) as (1).

4.4 Estimation of the LD second spectral moment μ_2

From (12), an appropriate estimate of μ_2 is

$$\hat{\mu}_2 = \frac{1}{m} \text{tr}(\tilde{\mathbf{S}}^2) - \frac{m-1}{n-1} = 1 + \frac{1}{m} \sum_{i \neq j} \left(\tilde{S}_{ij}^2 - \frac{1}{n-1} \right). \quad (22)$$

The expression on the left, in comparison to (12), uses $\tilde{\mathbf{S}}$ instead of $\mathbf{S}$ and uses $\hat{\mu}_1 = 1$ instead of $\hat{m}_1$. The replacement of $m-1$ instead of m is more clearly understood in the expression on the right, obtained by replacing $\text{tr}(\tilde{\mathbf{S}}^2) = m + \sum_{i \neq j} \tilde{S}_{ij}^2$. Here we can see that $\hat{\mu}_2$ is equal to 1 (the value of μ_2 under no correlation) plus $1/m$ times the total squared correlation observed in the sample LD matrix $\tilde{\mathbf{S}}$, after subtracting from each term a bias correction of $1/(n-1)$. The extra term $1/(n-1)$ is the approximate second moment, for large n , of the empirical correlation $\tilde{S}_{ij}$ when the true underlying correlation $\tilde{S}_{ij}$ is zero. It is pervasive in the LD matrix and we may refer to it as a “correlation floor”.

The following lemma, whose proof is in the Appendix, states that $\hat{\mu}_2$ is a consistent estimator of μ_2 .

Lemma 1. *Assume the spectral moments $m_k = \text{tr}(\mathbf{\Sigma}^k)/m$, $k = 1, \dots, 4$, are bounded as m gets large. Then, as m and n get large such that m/n converges to a constant,*

$$\hat{\mu}_2 = \mu_2 + O_P\left(\frac{1}{n}\right). \quad (23)$$

4.5 Asymptotic properties of the GWASH estimator

By its construction, the GWASH estimator has similar asymptotic properties to that of the Dicker estimator (11), namely consistency and asymptotic normality. Theorem 1 below shows this formally and gives the theoretical justification for using the GWASH estimator in GWAS.

Assumption 1. Suppose that the assumptions of Proposition 2 of [6] hold, namely:

- The variance components σ^2 and τ^2 , as well as the spectral moments $m_k = \text{tr}(\mathbf{\Sigma}^k)/m$, $k = 1, \dots, 4$, are bounded.
- Let $\tau_k^2 = \beta^T \mathbf{\Sigma}^k \beta$, $\Delta_k = \sum_{\ell=1}^k |\tau_\ell^2 - \tau_0^2 m_\ell|$, and suppose that $\Delta_2 = o(1)$ and $\Delta_3 = o(1/\sqrt{n})$.

Theorem 1. *Under Assumption 1, as m and n get large such that m/n converges to a constant, the GWASH estimator (20) is asymptotically Gaussian with mean h^2 and variance*

$$\frac{\psi^2}{n} = \frac{2}{n} \left(\frac{m}{n\mu_2} + 2\frac{\mu_3}{\mu_2^2} h^2 - h^4 \right). \quad (24)$$

Theorem 1 implies consistency of the estimator for large m and n . The proof is given in the Appendix.

An estimate of standard error for $\hat{h}^2$ can be obtained as $\hat{\psi}/\sqrt{n} = \sqrt{\hat{\psi}^2/n}$, where

$$\frac{\hat{\psi}^2}{n} = \frac{2}{n} \left(\frac{m}{n\hat{\mu}_2} + 2\frac{\hat{\mu}_3}{\hat{\mu}_2^2}\hat{h}^2 - \hat{h}^4 \right). \quad (25)$$

is a plug-in estimate of the asymptotic variance ψ^2/n and $\hat{\mu}_3$ is an estimator of μ_3 (see Section 5.3). The asymptotic normality of $\hat{h}^2$ allows constructing an approximate two-sided 95% confidence interval for h^2 of the form $\hat{h}^2 \pm 1.96\hat{\psi}/\sqrt{n}$. In addition, the null hypothesis $H_0 : h^2 = 0$ may be tested against the alternative $H_A : h^2 > 0$ using the Wald statistic $\sqrt{n}\hat{h}^2/\hat{\psi}$ and declaring it significant at the 5% level if it exceeds the normal quantile 1.645.

5 Practical aspects in the context of GWAS

5.1 The GWASH estimator from summary statistics

In publicly available GWAS results, the original data $\mathbf{y}$ and $\mathbf{X}$ required to compute the correlation scores (16) are typically not available. Instead, it is possible to access the t-statistics (8) from the univariate regressions. Correlation scores are closely related to the t-statistics. In fact, the next result shows that the original data is not necessary, but it is possible to convert the squared t-statistics to squared correlation scores by a simple formula.

Proposition 3. *The square of the correlation scores (16) can be obtained from the squared t-statistics (8) via*

$$u_j^2 = \left(\frac{n-1}{n-2} \right) \frac{t_j^2}{1 + t_j^2/(n-2)}. \quad (26)$$

The squared correlation scores and the squared t-statistics are very close for large n , but not exactly. The transformation is needed because the residual variance (7) typically used in GWAS is a biased estimator of the true noise variance. The effect of the transformation is to "undo" the division by (7) and turn the t-statistic into a more appropriate score.

To compute the GWASH estimator (20), s^2 can be computed directly from the u-scores (26). The LD second spectral moment $\hat{\mu}_2$ cannot be computed from summary statistics. Note, however, that $\hat{\mu}_2$ is a property of the population from which the GWAS data was sampled. Following others, we make the assumption that the sampled population has similar properties to those in public datasets such as the 1000 genomes project (1KGP) [10]. Under this assumption, $\hat{\mu}_2$ can be estimated from any random sample assayed on the same set of predictors, even if the representative sample is of a different size. For example, if a representative auxiliary dataset of size $\tilde{n}$ is available on the same set of SNPs, then $\hat{\mu}_2$ can be estimated using the methods of Section 4.4 with $\tilde{n}$ instead of n . The same holds for $\hat{\mu}_3$ (see Section 5.3).

5.2 Efficient computation of the LD second moment estimator $\hat{\mu}_2$

From a computational point of view, we may take advantage of the fact that, in a randomly mating population, SNPs appreciably far away within the same chromosome, or on different chromosomes, should be segregating independently. For independent markers i, j , their squared correlation $\tilde{S}_{ij}^2$ has mean of about $1/(n-1)$, and so the terms $\tilde{S}_{ij}^2 - 1/(n-1)$ in (22) far from the diagonal are small and can be excluded from the calculation.

In general, suppose that only a set I_2 of index pairs (i, j) , $i \neq j$, are included in the calculation of $\hat{\mu}_2$. This results in the modified estimator

$$\hat{\mu}_{2,I_2} = 1 + \frac{1}{m} \sum_{(i,j) \in I_2} \left(\tilde{S}_{ij}^2 - \frac{1}{n-1} \right) = 1 + \frac{1}{m} \left[\sum_{(i,j) \in I_2} \tilde{S}_{ij}^2 - \frac{|I_2|}{n-1} \right], \quad (27)$$

where $|I_2|$ is the number of elements in the set I_2 . Note that the bias correction of $1/(n-1)$ is only applied to each of the terms actually included in the sum.

Specifically, for a single chromosome with m_k markers, $k = 1, \dots, K$, excluding all pairs more than $q > 0$ indexes away is equivalent to applying formula (22) to the restricted matrix $\tilde{\mathbf{S}}_q^{(k)}$ with entries

$$(\tilde{S}_q^{(k)})_{ij} = \begin{cases} 1 & i = j \\ \tilde{S}_{ij} & i \neq j, (i, j) \in I_2^{(k)} \\ 0 & i \neq j, (i, j) \notin I_2^{(k)} \end{cases}, \quad (28)$$

where $I_2^{(k)} = \{(i, j) : 1 < |i - j| \leq q\}$. It can be shown that $|I_2^{(k)}| = q(2m_k - q - 1)$, yielding the formula

$$\hat{\mu}_{2,q}^{(k)} = \frac{1}{m_k} \left[\text{tr} \left(\tilde{\mathbf{S}}_q^{(k)} \right)^2 - \frac{q(2m_k - q - 1)}{n-1} \right]. \quad (29)$$

In practice, the restricted matrix (28) can be stored as a sparse matrix and the trace above computed using the property that for any squared matrix $\mathbf{A}$, $\text{tr}(\mathbf{A}^2) = \sum_{i,j} A_{ij}^2$.

For any set of K chromosomes with $m_1 + \dots + m_K = m$, it can be shown using (27) that the overall estimate $\hat{\mu}_{2,q}$ is the weighted average of the per-chromosome estimates (29), weighted by the number of markers m_k in each chromosome. In what follows, we refer to the distance q as *correlation bandwidth*.

5.3 Estimation of the LD third spectral moment μ_3

To compute the variance (25), we need an estimator of μ_3 . From (15), an appropriate estimate of μ_3 is

$$\begin{aligned} \hat{\mu}_3 &= \frac{1}{m} \text{tr}(\tilde{\mathbf{S}}^3) - 3 \frac{m-1}{n-1} \hat{\mu}_2 - \frac{(m-1)(m-2)}{(n-1)^2} \\ &= \frac{1}{m} \left[\text{tr}(\tilde{\mathbf{S}}^3) - 3 \frac{m(m-1)}{n-1} \hat{\mu}_2 - \frac{m(m-1)(m-2)}{(n-1)^2} \right]. \end{aligned} \quad (30)$$

To understand this estimator, we realize that

$$\begin{aligned} \text{tr}(\tilde{\mathbf{S}}^3) &= \text{tr}(\tilde{\mathbf{S}}\tilde{\mathbf{S}}^2) = \sum_{i,j=1}^m \tilde{S}_{ij}(\tilde{S}^2)_{ij} = \sum_{i,j,k=1}^m \tilde{S}_{ij}\tilde{S}_{jk}\tilde{S}_{ik} \\ &= \sum_{i=j=k} + \sum_{i \neq j=k} + \sum_{i=j \neq k} + \sum_{i=k \neq j} + \sum_{i \neq j \neq k} \\ &= m + \sum_{i \neq j} \tilde{S}_{ij}^2 + \sum_{i \neq k} \tilde{S}_{ik}^2 + \sum_{j \neq k} \tilde{S}_{jk}^2 + \sum_{i \neq j \neq k} \tilde{S}_{ij}\tilde{S}_{jk}\tilde{S}_{ik}. \end{aligned}$$

Thus, the first subtracted term in (30) makes a bias correction of $\hat{\mu}_2/(n-1)$ for each of the $3m(m-1)$ second order terms in the sum above, while the second subtracted term makes a bias correction of $1/(n-1)^2$ for each of the $m(m-1)(m-2)$ third order terms.

As before, an approximation can be obtained by only considering entries of $\tilde{\mathbf{S}}$ close to the diagonal. Let I_2 and I_3 be respectively the set of index pairs (i, j) , $i \neq j$, and index triplets (i, j, k) other than $i = j = k$ to be included in this calculation. Then we have the modified estimator

$$\hat{\mu}_{3,I_3} = 1 + \frac{1}{m} \left[\sum_{(i,j,k) \in I_3} \tilde{S}_{ij} \tilde{S}_{jk} \tilde{S}_{ik} - 3 \frac{|I_2|}{n-1} \hat{\mu}_2 - \frac{|I_3|}{(n-1)^2} \right], \quad (31)$$

where $|I_2|$ and $|I_3|$ are the number of elements in the sets I_2 and I_3 , respectively.

Specifically, for a single chromosome with m_k markers, let $\tilde{\mathbf{S}}_q$ be the restricted matrix (28) so that

$$\begin{aligned} I_2 &= \{(i, j) : 1 < |i - j| \leq q\}, \\ I_3 &= \{(i, j, k) : 1 < |i - j| \leq q, 1 < |i - k| \leq q, 1 < |j - k| \leq q\}. \end{aligned}$$

the number of elements $|I_2|$ and $|I_3|$ can be computed exactly as

$$\begin{aligned} |I_2| &= q(2m_k - q - 1) \\ |I_3| &= q(q - 1)(3m_k - 2q - 2). \end{aligned}$$

Replacing in (31) gives the formula

$$\hat{\mu}_{3,q}^{(k)} = \frac{1}{m_k} \left[\text{tr} \left(\tilde{\mathbf{S}}_q^{(k)} \right)^3 - 3 \frac{q(2m_k - q - 1)}{n-1} \hat{\mu}_2 - \frac{q(q-1)(3m_k - 2q - 2)}{(n-1)^2} \right]. \quad (32)$$

Note that the trace above can be computed using the property that for any squared matrix $\mathbf{A}$, $\text{tr}(\mathbf{A}^3) = \sum_{i,j} A_{ij}(\mathbf{A}^2)_{ij}$.

Again, for K chromosomes, the overall estimate $\hat{\mu}_{3,q}$ is the weighted average of the per-chromosome estimates (29), weighted by the number of markers m_k in each chromosome.

5.4 Relationship to LDSC regression

The LDSC regression method [2] is derived under different modeling assumptions to ours, the most important being the assumption that the β coefficients are random. In addition, the corresponding software is written for large GWAS applications and is not amenable to smaller scale simulations as we do here. To allow direct comparison both analytically and in simulations, here we consider a stylized version of LDSC regression that closely matches the GWASH estimator.

In its most basic version, written in our notation and based on our best understanding of [2], the LDSC method is essentially based on the approximation

$$\mathbb{E}[u_j^2 | \hat{\ell}_j] \approx h^2 \left(\frac{n}{m} \hat{\ell}_j - 1 \right) + 1, \quad j = 1, \dots, m \quad (33)$$

where $\hat{\ell}_j = \sum_{k=1}^m \tilde{r}_{jk}^2$ are the so-called LD-scores and $\tilde{r}_{jk}$ are the entries of $\tilde{\mathbf{S}} = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} / (n-1)$ (see Appendix A.4 for details). The LDSC method estimates h^2 by fitting a linear model based on (33) plus observation noise. Defining $\mathbf{u}^2 = (u_1^2, \dots, u_m^2)^T$ and $\boldsymbol{\ell} = ((n/m)\hat{\ell}_1 - 1, \dots, (n/m)\hat{\ell}_m - 1)^T$, the model (33) reads $\mathbb{E}(\mathbf{u}^2) = h^2 \boldsymbol{\ell} + \mathbf{1}$, leading to the least squares estimator

$$\hat{h}_{\text{LD}}^2 = |\boldsymbol{\ell}|^{-2} \boldsymbol{\ell}^T (\mathbf{u}^2 - \mathbf{1}), \quad (34)$$

fitted with a fixed intercept equal to 1.

The GWASH estimator is related to the LDSC regression estimator above in the following way. In linear regression, the fitted line always goes through the average of the point cloud. Therefore, the average $(\bar{\ell}, \bar{u^2}) = (\mathbf{1}^T \boldsymbol{\ell} / m, \mathbf{1}^T \mathbf{u}^2 / m)$ must satisfy the equation

$$\bar{u^2} = \hat{h}_{\text{LD}}^2 \bar{\ell} + 1. \quad (35)$$

But $\bar{u^2} = \sum_{j=1}^m u_j^2 / m = s^2$ by (21) and

$$\bar{\ell} = \frac{1}{m} \sum_{j=1}^m \left(\frac{n}{m} \hat{\ell}_j - 1 \right) = \frac{n}{m} \left[\frac{1}{m} \text{tr}(\tilde{\mathbf{S}}^2) \right] - 1 = \frac{n}{m} \left(\hat{\mu}_2 + \frac{m-1}{n-1} \right) - 1 = \frac{n}{m} \hat{\mu}_2 + O\left(\frac{1}{n}\right)$$

by (22), assuming that m/n converges to a constant. Replacing in (35) and comparing to (20), we obtain that

$$\hat{h}_{\text{LD}}^2 = \frac{\bar{u^2} - 1}{\bar{\ell}} = \frac{m(s^2 - 1)}{n\hat{\mu}_2} + O\left(\frac{1}{n}\right) = \hat{h}_{\text{GWASH}}^2 + O\left(\frac{1}{n}\right).$$

In other words, the LDSC estimator with fixed intercept equal to 1 and the GWASH estimator are asymptotically equivalent.

Another way of understanding the above derivation is that, if we consider a scatterplot of u_j^2 as a function of the LD scores $\hat{\ell}_j$, LDSC fits the least-squares straight line through the distribution, while GWASH targets the mean of the distribution directly. We will see in the simulations and data analysis that both give similar estimates but have different standard errors.

6 Finite sample performance

The purpose of the following simulations is to evaluate the statistical and computational performance of the GWASH estimator under various finite sample scenarios. To easily control the amount of correlation off the main diagonal, we use an autoregressive (AR) structure for the correlation matrix $\tilde{\mathbf{S}}$.

6.1 Estimation of heritability

To push the limits of the estimator, in this section we consider an AR covariance structure for the predictor matrix $\mathbf{X}$ where the AR parameter ranges from 0 to 0.8 and where the variances of the columns of $\mathbf{X}$ have a wide spread from 1 to m . We consider two different distributions for the entries of $\mathbf{X}$:

- The rows of $\mathbf{X}$ are i.i.d. multivariate normal with mean 0 and covariance matrix $\boldsymbol{\Sigma} = [\text{Diag}(\boldsymbol{\Sigma})]^{1/2} \tilde{\boldsymbol{\Sigma}}(\rho) [\text{Diag}(\boldsymbol{\Sigma})]^{1/2}$, where $\text{Diag}(\boldsymbol{\Sigma}) = \text{Diag}(1, 2, \dots, m)$ and $\tilde{\boldsymbol{\Sigma}}(\rho)$ is the $m \times m$ AR correlation matrix with entries $\tilde{\Sigma}_{ij} = \rho^{|i-j|}$.
- The rows of $\mathbf{X}$ are i.i.d. multivariate binomial, generated using a Gaussian multivariate copula [11, 12]. According to this method, a multivariate normal vector is generated with the same covariance matrix $\boldsymbol{\Sigma}$ as above and AR parameter denoted ρ^* . The multivariate normal vector is then transformed to binomial by a quantile transformation with the corresponding variance. Because of the copula, the correlation matrix that is obtained for the binomial variables is not exactly AR with the specified parameter. As a reminder, we use the notation ρ^* to indicate the AR parameter used as an input to the multivariate copula generating mechanism.

To simulate possible structures of the vector of coefficients β , we consider two different scenarios for this vector:

- β is a single realization of m i.i.d. $N(0, 1)$ variables.
- β is a mixture, containing 90% of 0's and 10% i.i.d. $N(0, 1)$ variables.

In all cases, the outcome y is generated according to model (3) with i.i.d. Gaussian errors. Given β and Σ , for any desired heritability $h^2 > 0$, the error variance is set to $\sigma^2 = \tau^2(1 - h^2)/h^2$ so that (5) gives heritability h^2 . For $h^2 = 0$, we set $\beta = 0$.

Figures 1 - 3 show the estimates of h^2 under the aforementioned combinations. The estimation methods shown are:

- The GWASH estimator (20) using the full sample correlation matrix ($q = m - 1$) to estimate μ_2 , as in (22).
- The GWASH estimator (20) using only q off-diagonals of the sample correlation matrix to estimate μ_2 , as in (29) (considered only when $\rho > 0$).
- The Dicker estimator for unestimable covariance (11).
- The simple LD regression estimator (34).

All h^2 estimates are hardly distinguishable and close to the true values (grey diagonal line) within simulation error. This is precisely the desired behavior, as it shows that the GWASH estimator can estimate heritability just as well as the Dicker and LDSC estimators using only summary statistics. Note too that the correlation bandwidth q has little influence on the results.

6.2 Estimation of spectral moments and standard error

Table 1: Estimates of μ_2 and μ_3 ($m = 1000$): values presented are the mean and empirical standard deviation over 100 repetitions. The symbol ρ^* represents the AR parameter of the Gaussian copula.

X	AR	μ_2	$\hat{\mu}_2$	$\hat{\mu}_{2,q}$	μ_3	$\hat{\mu}_3$	$\hat{\mu}_{3,q}$
		true	Full $\tilde{S}$	Partial $\tilde{S}$	true	Full $\tilde{S}$	Partial $\tilde{S}$
normal	$\rho = 0.8$	4.55	4.53 (0.03)	4.50 (0.02) _{$q=11$}	30.49	30.2 (0.48)	29.1 (0.32) _{$q=11$}
	$\rho = 0.4$	1.38	1.38 (0.006)	1.38 (0.003) _{$q=5$}	2.36	2.35 (0.04)	2.35 (0.01) _{$q=5$}
	$\rho = 0.2$	1.08	1.08 (0.004)	1.08 (0.001) _{$q=3$}	1.26	1.26 (0.02)	1.26 (0.004) _{$q=3$}
	$\rho = 0$	1	1.00 (0.004)		1	1.00 (0.01)	
binomial	$\rho^* = 0.8$	2.54	2.55 (0.02)	2.55 (0.02) _{$q=11$}	10.42	10.48 (0.24)	10.28 (0.18) _{$q=11$}
	$\rho^* = 0.4$	1.13	1.13 (0.005)	1.13 (0.003) _{$q=5$}	1.44	1.44 (0.02)	1.44 (0.009) _{$q=5$}
	$\rho^* = 0.2$	1.03	1.03 (0.004)	1.03 (0.001) _{$q=3$}	1.08	1.08 (0.01)	1.08 (0.003) _{$q=3$}
	$\rho^* = 0$	1	1.00 (0.004)		1	1.00 (0.01)	

To understand the effect of LD on the spectral moment estimators, estimates of μ_2 and μ_3 are shown in Table 1 under the different X structures considered above. Both $\hat{\mu}_2$ and $\hat{\mu}_3$ match their true values whether the full or partial sample correlation matrix is used in their

Figure 1: Estimates of h^2 for β normal and $\mathbf{X}$ normal ($m = 1000$, $n = 500$, 100 repetitions). The bars represent the empirical standard deviation of the estimates over 100 repetitions.

Figure 2: Estimates of h^2 for β mixture and $\mathbf{X}$ normal ($m = 1000$, $n = 500$, 100 repetitions).

Figure 3: Estimates of h^2 for β mixture and $\mathbf{X}$ binomial ($m = 1000$, $n = 500$, 100 repetitions).

Figure 4: Standard errors of h^2 estimates for β normal and $\mathbf{X}$ normal ($m = 1000$, $n = 500$, 100 repetitions).

Figure 5: Standard errors of h^2 estimates for β mixture and $\mathbf{X}$ normal ($m = 1000$, $n = 500$, 100 repetitions).

Figure 6: Standard errors of h^2 estimates for β mixture and $\mathbf{X}$ binomial ($m = 1000$, $n = 500$, 100 repetitions).

Trait	Ref.	Data source (website)
Height	[23]	http://portals.broadinstitute.org/collaboration/giant/images/0/01/GIANT_HEIGHT.Wood_et_al_2014_publicrelease_HapMapCeuFreq.txt.gz
BMI	[14]	http://portals.broadinstitute.org/collaboration/giant/images/1/15/SNP_gwas_mc_merge_nogc.tbl.uniq.gz
IQ	[20]	https://ctg.cncr.nl/documents/p1651/sumstats.txt.gz
Edu. Attain.	[17]	http://ssgac.org/documents/SSGAC.Rietveld2013.zip

Table 2: Reference information on four different GWAS studies.

estimation. Note that the empirical SE when using the partial $\tilde{\mathbf{S}}$ is slightly smaller than that when using the full $\tilde{\mathbf{S}}$.

Finally, Figures 4 - 6 compare estimates of standard error (SE) according to the following methods:

- Empirical SE from 100 repetitions of the GWASH estimator (20) using the full sample correlation matrix ($q = m - 1$) to estimate μ_2 .
- Empirical SE from 100 repetitions of the GWASH estimator (20) using only q off-diagonals of the sample correlation matrix to estimate μ_2 (considered only when $\rho > 0$).
- The theoretical asymptotic SE of the GWASH estimator (square root of (25)), using the full sample correlation matrix ($q = m - 1$) to estimate μ_2 and μ_3 .
- The theoretical asymptotic SE of the GWASH estimator (square root of (25)), using only q off-diagonals of the sample correlation matrix to estimate μ_2 and μ_3 .
- Empirical SE from 100 repetitions of the simple LDSC regression estimator (34).
- Theoretical SE of the simple LD regression estimator (34) obtained from the linear model fit.

In all plots, the asymptotic SE formula for the GWASH estimator approximates the empirical SE closely. For LD regression, however, overestimates or underestimates the corresponding empirical SE, explaining why the estimation of SE in [2] requires computational methods such as jackknife and bootstrapping to estimate the SE more accurately.

7 Real data examples

To demonstrate the performance of our estimator on real GWAS data, we estimated h^2 from publicly available GWAS summary statistics for four phenotypes, height [23], body mass index (BMI) [14], IQ [20] and educational attainment [17] (Edu. Attain.). The results of the GWASH estimator were compared to the LDSC estimator [2] with intercept = 1, as follows.

7.1 Pre-processing of GWAS summary statistics

Summary statistics from the four GWAS studies listed in Table 2 were downloaded from the authors cited public repositories. For each study we kept the SNP name, effect allele (A1), non-effect allele (A2), per SNP sample size (n), association p-value (p) and corresponding test statistic (t). Where per SNP sample sizes were not available (Edu. Attain.), we used the sample

Trait	n	m (total)	m (used)	GWASH	LDSC (int=1)
Height	246,770	2,550,859	872,188	0.40 (0.003)	0.42 (0.015)
BMI	233,018	2,554,638	872,188	0.05 (0.002)	0.05 (0.006)
IQ	75,270	12,104,295	872,188	0.21 (0.006)	0.20 (0.008)
Edu. Attain.	328,917	9,444,231	872,188	0.10 (0.002)	0.10 (0.003)

Table 3: Heritability estimates for the traits in Table 2. Estimated SEs are given in parentheses.

size reported in the paper for each SNP. Where test statistics were not reported (BMI, Edu. Attain, Height), we converted two-tailed p-values to z-scores via the inverse of the normal CDF, maintaining the sign from the regression coefficients.

To enable comparisons between our approach and LDSC, we kept only the subset of SNPs that was present in each of the four GWAS, had an LDSC authors precomputed LD score, and had genotype data available in the 1KGP data. We also excluded SNPs with a minor allele frequency less than 0.1% in any of the five 1KGP European subpopulations as these may be less reliably genotyped or more population variable, limiting their representativeness. This left a common set of 872,188 SNPs for each GWAS study that could be analyzed with each approach.

7.2 GWASH estimates and comparison to LDSC

To perform LDSC regression for each of the four GWAS studies, we downloaded all necessary software and reference data from the authors repository (<https://github.com/bulik/ldsc>). Both GWASH and LDSC require information about the LD among SNPs that is not typically made available alongside summary statistics. The LDSC authors address this by providing pre-computed LD scores estimated from a subset of representative individuals genotypes available as part of the 1KGP data, along with their software (<https://github.com/bulik/ldsc>). We used these pre-computed values and their recommended protocols as faithfully as possible, following their tutorial provided (<https://github.com/bulik/ldsc/wiki/Heritability-and-Genetic-Correlation>). We performed the estimation with the intercept constrained to 1, as our method has theoretical similarity to the constrained intercept case.

For our estimator, calculation of $\hat{\mu}_2$ and $\hat{\mu}_3$ require LD information not provided with summary statistics. To compute representative values, we used a sample of the same 1KGP data (see Section 7.3 below). Table 3 shows the performance of our methods and LDSC regression on the same GWAS data. The estimated values are very similar, but the SEs for the GWASH estimator are smaller. Needless to say, all heritability estimates are highly significantly greater than zero.

7.3 Estimation of μ_2 and μ_3 from the 1KGP data

A sample of 503 individuals of European ancestry were used to compute representative values of $\hat{\mu}_2$ and $\hat{\mu}_3$. Genome-wide genotypes are available for these individuals through the 1KGP data, phase 36 (<http://www.internationalgenome.org/>).

The computation of $\hat{\mu}_2$ and $\hat{\mu}_3$ require manipulations of the LD matrix. We computed this matrix using the statistical genetic software package **plink2** [4], which provides fast routines for manipulating large genotype data sets. We can take advantage of the fact that SNPs far away from each other, or on different chromosomes are expected to have correlation (LD) approaching 0. For our reference set of genotypes we computed a banded correlation matrix with correlation

Figure 7: Sensitivity of GWASH estimates to the calculation of $\hat{\mu}_2$ as a function of correlation bandwidth q . The LDSC (int=1) estimator is added for reference.

bandwidth $q = 1000$ (using the notation of Section 4.4). This was done using the `plink2` commands `--r` and `--ld-window 1000`, which only returns correlations up to 1000 rows off the diagonal, for each chromosome in parallel. Matrix calculations for $\hat{\mu}_2$ and $\hat{\mu}_3$ were performed in an R routine using sparse matrix operations in the package `matrix`. The values so obtained $\hat{\mu}_2 = 16.93$ and $\hat{\mu}_3 = 617.35$ were used in the calculation of the GWASH estimate and its std. error in Table 3.

To evaluate the choice of correlation bandwidth q , the GWASH estimate was recomputed for a range of values of q up to 5000 used in the calculation of $\hat{\mu}_2$. Figure 7 shows that the GWASH estimate is fairly insensitive to the correlation bandwidth q , the chosen value $q = 1000$ being a reasonable compromise between accuracy and computation.

8 Discussion

8.1 Estimation of SE

A nice property of the GWASH estimator, inherited from the Dicker estimator and not available with other currently used estimators, is that the precision (24) of the estimator is known theoretically based on the number of SNPs m , the sample size n , the second and third spectral moments μ_2 and μ_3 of the LD matrix, and the true heritability h^2 . The first two quantities are known from the study, while the second two can be estimated from a public resource (e.g. 1KGP). The true heritability is unknown. In this paper we have used a plug-in estimate of h^2 from the study itself.

To assess the sensitivity of the SE to the value of h^2 , Figure 8 shows the SE (square root of (25)) as a function of the sample size n and the true heritability h^2 using the values $m = 872,188$, $\mu_2 = 16.93$ and $\mu_3 = 617.35$ from the data analysis. The plot shows that the SE is almost insensitive to the value of h^2 , increasing only slightly as h^2 increases for any fixed n . As a consequence, as an alternative to (25), a slightly conservative but more stable estimate

Figure 8: The SE of the GWASH estimator as a function of n and h^2 , for $m = 872,188$, $\mu_2 = 16.93$ and $\mu_3 = 617.35$. The red curve indicates the pairs (n, h^2) for which $h^2 = 1.645$ SE; values of n to the right of the curve allow detection of a non-zero heritability at the 5% level.

of the SE can be obtained by plugging in the worst-case value $h^2 = 1$ instead of the estimated value $\hat{h}^2$.

8.2 Sample size and power calculations for prospective GWAS

Relation (24) can be used in a prospective study to determine the number of subjects required to estimate heritability according to a desired accuracy. Given any fixed set of m SNPs, the values of μ_2 and μ_3 may be estimated from a public resource (e.g. 1KGP) and then the SE can be designed as a function of n and the targeted h^2 . For the values of m , μ_2 and μ_3 in the data analysis, Figure 8 shows that the SE can be quite large for small n , but it drops as n increases.

From a design point of view, the sample size n can be chosen to achieve a desired SE. For example, for a heritability of $h^2 = 0.5$, an SE of 0.05 is achieved with $n = 7234$ (red circle in Figure 8). The SE can also help design studies with the goal of detecting a heritability that is significantly greater than zero. As mentioned at the end of Section 4.5, a one-sided Wald test will be significant at the 5% level if the estimate of h^2 is greater than 1.645 SEs. In Figure 8, this corresponds to choosing n to the right of the red curve. For example, to detect a heritability of $h^2 = 0.8$, the minimal sample size is $n = 673$; to detect a heritability of $h^2 = 0.2$, the minimal sample size is $n = 2699$ (red triangles).

A Proofs

A.1 Proofs of shorter propositions

Proof of Proposition 1. Dicker [6, Section 4.1] proposes the estimators

$$\begin{aligned}\hat{\sigma}^2 &= \frac{m+n}{n(n-1)} \|\mathbf{y}\|^2 - \frac{1}{n(n-1)} \|\hat{\Sigma}^{-1/2} \mathbf{X}^T \mathbf{y}\|^2, \\ \hat{\tau}^2 &= -\frac{m}{n(n-1)} \|\mathbf{y}\|^2 + \frac{1}{n(n-1)} \|\hat{\Sigma}^{-1/2} \mathbf{X}^T \mathbf{y}\|^2,\end{aligned}$$

where n has been replaced by $n-1$ due to the centering of $\mathbf{y}$ and $\mathbf{X}$. The estimator of heritability is the fraction

$$\hat{h}_I^2 = \frac{\hat{\tau}^2}{\hat{\sigma}^2 + \hat{\tau}^2} = \frac{\|\hat{\Sigma}^{-1/2} \mathbf{X}^T \mathbf{y}\|^2 - m \|\mathbf{y}\|^2}{n \|\mathbf{y}\|^2} = \frac{m}{n} \left(\frac{\|\hat{\Sigma}^{-1/2} \mathbf{X}^T \mathbf{y}\|^2}{m \|\mathbf{y}\|^2} - 1 \right).$$

□

Proof of Proposition 2. Dicker [6, Section 4.2] proposes the estimators

$$\begin{aligned} \hat{\sigma}^2 &= \left(1 + \frac{m \hat{m}_1^2}{n \hat{m}_2} \right) \frac{1}{n-1} \|\mathbf{y}\|^2 - \frac{\hat{m}_1}{n(n-1) \hat{m}_2} \|\mathbf{X}^T \mathbf{y}\|^2, \\ \hat{\tau}^2 &= -\frac{m \hat{m}_1^2}{n(n-1) \hat{m}_2} \|\mathbf{y}\|^2 + \frac{\hat{m}_1}{n(n-1) \hat{m}_2} \|\mathbf{X}^T \mathbf{y}\|^2, \end{aligned}$$

where n has been replaced by $n-1$ due to the centering of $\mathbf{y}$ and $\mathbf{X}$. The estimator of heritability is the fraction

$$\hat{h}_{II}^2 = \frac{\hat{\tau}^2}{\hat{\sigma}^2 + \hat{\tau}^2} = \frac{\hat{m}_1 \|\mathbf{X}^T \mathbf{y}\|^2 - m \hat{m}_1^2 \|\mathbf{y}\|^2}{n \hat{m}_2 \|\mathbf{y}\|^2} = \frac{m \hat{m}_1^2}{n \hat{m}_2} \left(\frac{\|\mathbf{X}^T \mathbf{y}\|^2}{m \hat{m}_1 \|\mathbf{y}\|^2} - 1 \right).$$

□

Proof of Proposition 3. From (7) and (8),

$$(n-2) \hat{\sigma}_j^2 = \|\mathbf{y}\|^2 - \|\mathbf{x}_j\|^2 \hat{\beta}_j^2 = \|\mathbf{y}\|^2 - \hat{\sigma}_j^2 t_j^2,$$

so $\|\mathbf{y}\|^2 = \hat{\sigma}_j^2 (t_j^2 + n - 2)$. Thus, using (6) and again (8), we can write

$$u_j^2 = (n-1) \left(\frac{\mathbf{x}_j^T \mathbf{y}}{\|\mathbf{x}_j\| \|\mathbf{y}\|} \right)^2 = (n-1) \frac{\|\mathbf{x}_j\|^2 \hat{\beta}_j^2}{\|\mathbf{y}\|^2} = (n-1) \frac{\|\mathbf{x}_j\|^2 \hat{\beta}_j^2}{\hat{\sigma}_j^2 (t_j^2 + n - 2)} = (n-1) \frac{t_j^2}{t_j^2 + n - 2}$$

and the result follows. □

A.2 Supporting lemmas for the proof of Theorem 1

Proof of Lemma 1. Any two sample correlations $\tilde{S}_{jk}$ and $\tilde{S}_{lh}$ are asymptotically bivariate normal with means $\tilde{\Sigma}_{jk}$ and $\tilde{\Sigma}_{lh}$, and variances $(1 - \tilde{\Sigma}_{jk}^2)/n$ and $(1 - \tilde{\Sigma}_{lh}^2)/n$. The covariance $\text{Cov}(\tilde{S}_{jk}, \tilde{S}_{lh})$, whose expression is given by [7], is a polynomial of order 4 in $\tilde{\Sigma}_{jk}$ and $\tilde{\Sigma}_{lh}$ and proportional to $1/n$. From the asymptotic normality, we may write

$$\begin{aligned} \mathbb{E}(\tilde{S}_{jk}^2) &= \tilde{\Sigma}_{jk}^2 + \frac{(1 - \tilde{\Sigma}_{jk}^2)^2}{n} + o\left(\frac{1}{n}\right), \\ \text{Cov}(\tilde{S}_{jk}^2, \tilde{S}_{lh}^2) &= 2 \left[\text{Cov}(\tilde{S}_{jk}, \tilde{S}_{lh}) \right]^2 + o\left(\frac{1}{n}\right) = O\left(\frac{1}{n^2}\right). \end{aligned} \tag{36}$$

Now, from (22) and (36),

$$\begin{aligned} \mathbb{E}(\hat{\mu}_2) &= 1 + \frac{1}{m} \sum_{j \neq k} \left[\mathbb{E}(\tilde{S}_{jk}^2) - \frac{1}{n-1} \right] = 1 + \frac{1}{m} \sum_{j \neq k} \left[\tilde{\Sigma}_{jk}^2 - \frac{2\tilde{\Sigma}_{jk}^2 - \tilde{\Sigma}_{jk}^4}{n} + o\left(\frac{1}{n}\right) \right] \\ &= \mu_2 - \frac{1}{m} \sum_{j \neq k} \left[\frac{2\tilde{\Sigma}_{jk}^2 - \tilde{\Sigma}_{jk}^4}{n} + o\left(\frac{1}{n}\right) \right] = \mu_2 + O\left(\frac{1}{n}\right), \end{aligned}$$

since the spectral moments of $\tilde{\Sigma}$ up to order 4 are assumed bounded by Assumption 1. Furthermore,

$$\text{Var}(\hat{\mu}_2) = \frac{1}{m^2} \sum_{j \neq k} \sum_{l \neq h} \text{Cov}(\tilde{S}_{jk}^2, \tilde{S}_{lh}^2) = O\left(\frac{1}{n^2}\right),$$

again because the spectral moments of $\tilde{\Sigma}$ up to order 4 are assumed bounded. Thus,

$$\text{E}(\hat{\mu}_2 - \mu_2)^2 = \text{Var}(\hat{\mu}_2) + [\text{E}(\hat{\mu}_2) - \mu_2]^2 = O\left(\frac{1}{n^2}\right),$$

implying (23). $\square$

Lemma 2. Let $\mathbf{D} = \text{Diag}(\mathbf{S})$ and $\mathbf{\Delta} = \text{Diag}(\mathbf{\Sigma})$. Suppose Assumption 1 holds. Then

$$\begin{aligned} (i) \quad & \frac{1}{m} \|\mathbf{\Delta}^{-1} \mathbf{D} - \mathbf{I}\|^2 = \frac{2}{n} + O_P\left(\frac{1}{n\sqrt{m}}\right), \\ (ii) \quad & \frac{1}{m} \|\mathbf{\Delta}^{-1} \mathbf{D}\|^2 - 1 = \frac{2}{n} + O_P\left(\frac{1}{n\sqrt{m}}\right). \end{aligned}$$

Proof. For the purpose of simplicity of the proof, assume that the columns of $\mathbf{X}$ have not been centered so that the rows of $\mathbf{X}$ are independent. This makes no difference asymptotically, only changing the number of degrees of freedom $n - 1$ to n . Let $W_j = \|\mathbf{x}_j\|^2 / [n\Sigma_{jj}]$ be the j -th diagonal entry of $\mathbf{\Delta}^{-1} \mathbf{D}$ where Σ_{jj} is the j -th diagonal entry of $\mathbf{\Sigma}$ (or $\mathbf{\Delta}$). Then $W_j \sim \chi_n^2/n$ has mean 1 and variance $2/n$.

(i) The expectation of $\|\mathbf{\Delta}^{-1} \mathbf{D} - \mathbf{I}\|^2 / m = \sum_{j=1}^m (W_j - 1)^2 / m$ is equal to $\text{Var}(W_j) = 2/n$. Its variance is

$$\begin{aligned} \text{Var}\left\{\frac{1}{m} \|\mathbf{\Delta}^{-1} \mathbf{D} - \mathbf{I}\|^2\right\} &= \frac{1}{m^2} \sum_{j,k} \text{Cov}[(W_j - 1)^2, (W_k - 1)^2] \\ &= \frac{1}{m^2} \sum_{j,k} \text{Cov}\left[\left(\frac{1}{n} \sum_{i=1}^n \frac{x_{ij}^2}{\Sigma_{jj}} - 1\right)^2, \left(\frac{1}{n} \sum_{\ell=1}^n \frac{x_{\ell k}^2}{\Sigma_{kk}} - 1\right)^2\right] \\ &= \frac{1}{m^2 n^4} \sum_{j,k} \left\{ \sum_{i,g,\ell,h} \text{Cov}\left[\left(\frac{x_{ij}^2}{\Sigma_{jj}} - 1\right)\left(\frac{x_{gj}^2}{\Sigma_{jj}} - 1\right), \left(\frac{x_{\ell k}^2}{\Sigma_{kk}} - 1\right)\left(\frac{x_{hk}^2}{\Sigma_{kk}} - 1\right)\right] \right\}. \end{aligned}$$

Since the rows of $\mathbf{X}$ are independent, most terms in the inner sum vanish except for the cases $i = g = \ell = h$, where the inner covariance is $8\tilde{\Sigma}_{jk}^2(3\tilde{\Sigma}_{jk}^2 + 4)$ (and $\tilde{\Sigma}_{jk}$ is the (j, k) entry of $\tilde{\Sigma}$), and the terms $i = \ell \neq g = h$ and $i = h \neq g = \ell$, where the inner covariance is $4\tilde{\Sigma}_{jk}^4$. Here we have used the fact that within any row i of $\mathbf{X}$,

$$\begin{aligned} \text{Cov}(\Sigma_{jj}^{-1} x_{ij}^2, \Sigma_{kk}^{-1} x_{ik}^2) &= 2\tilde{\Sigma}_{jk}^2, \\ \text{Cov}(\Sigma_{jj}^{-2} x_{ij}^4, \Sigma_{kk}^{-1} x_{ik}^2) &= 12\tilde{\Sigma}_{jk}^2, \\ \text{Cov}(\Sigma_{jj}^{-2} x_{ij}^4, \Sigma_{kk}^{-2} x_{ik}^4) &= 24\tilde{\Sigma}_{jk}^2(\tilde{\Sigma}_{jk}^2 + 3). \end{aligned}$$

Hence,

$$\text{Var}\left\{\frac{1}{m} \|\mathbf{\Delta}^{-1} \mathbf{D} - \mathbf{I}\|^2\right\} = \frac{1}{m^2 n^4} \sum_{j,k} \left[2n(n-1) \cdot 4\tilde{\Sigma}_{jk}^4 + n \cdot 8\tilde{\Sigma}_{jk}^2(3\tilde{\Sigma}_{jk}^2 + 4) \right] = \frac{8}{mn^2} \mu_4 + O\left(\frac{1}{mn^3}\right),$$

where $\sum_{jk} \tilde{\Sigma}_{jk}^4/m \leq \text{tr}(\Sigma^4)/m = m_4$ is bounded by Assumption 1. This yields (i).
(ii) Write

$$\frac{1}{m} \|\Delta^{-1} \mathbf{D} - \mathbf{I}\|^2 = \frac{1}{m} \|\Delta^{-1} \mathbf{D}\|^2 - \frac{2}{m} \text{tr}(\Delta^{-1} \mathbf{D}) + \frac{1}{m} \|\mathbf{I}\|^2 = \frac{1}{m} \|\Delta^{-1} \mathbf{D}\|^2 - \frac{2}{m} \sum_{j=1}^m W_j + 1.$$

Thus $\|\Delta^{-1} \mathbf{D}\|^2/m$ has expectation $1 + 2/n$. Similar to the proof of (ii), it can be shown that the variance is of the same order. $\square$

A.3 Proof of Theorem 1

Consider a situation where, instead of $\mathbf{X}$, the observed covariate matrix is $\mathbf{Z} = \mathbf{X} \Delta^{-1/2}$ with $\Delta = \text{Diag}(\Sigma)$. The idea of the proof is to approximate the GWASH estimator (20) by an estimator based on $\mathbf{Z}$. This is done in two steps by: (1) establishing the asymptotic properties of the estimator based on $\mathbf{Z}$; (2) establishing the asymptotic equivalence of the two estimators.

Let $\|A\|^2 = \text{tr}(A^T A)$ denote the squared Frobenius norm of the matrix A and recall the Cauchy-Schwartz inequality $\text{tr}(A^T B) \leq \|A\| \|B\|$.

Step 1: Based on $\mathbf{Z}$, the full model (3) can be written as

$$\mathbf{y} = \mathbf{Z} \tilde{\beta} + \varepsilon, \quad (37)$$

where $\tilde{\beta} = \Delta^{-1/2} \beta$. Note that $\tau^2 = \beta^T \Sigma \beta = \tilde{\beta}^T \tilde{\Sigma} \tilde{\beta}$, so the heritability (5) does not change. The matrix $\mathbf{Z}$ has rows with covariance $\tilde{\Sigma} = \Delta^{-1/2} \Sigma \Delta^{-1/2}$ and its sample covariance matrix is $\mathbf{Z}^T \mathbf{Z} / (n-1) = \Delta^{-1/2} \mathbf{S} \Delta^{-1/2}$. Following the form of Dicker's estimator (11) for unestimable covariance, the heritability estimator based on $\mathbf{Z}$ is

$$\tilde{h}^2 = \frac{m \tilde{m}_1^2}{n \tilde{m}_2} \left(\frac{\|\mathbf{Z}^T \mathbf{y}\|^2}{m \tilde{m}_1 \|\mathbf{y}\|^2} - 1 \right) = \frac{m \tilde{m}_1^2}{n \tilde{m}_2} \left(\frac{\|\mathbf{Z}^T \tilde{\mathbf{y}}\|^2}{m \tilde{m}_1 (n-1)} - 1 \right), \quad (38)$$

where

$$\tilde{m}_1 = \frac{1}{m} \text{tr}(\Delta^{-1/2} \mathbf{S} \Delta^{-1/2}), \quad \tilde{m}_2 = \frac{1}{m} \text{tr}[(\Delta^{-1/2} \mathbf{S} \Delta^{-1/2})^2] - \frac{m}{n-1} \tilde{m}_1^2. \quad (39)$$

Applying Proposition 2 of [6] to the estimator (38) gives that this estimator is asymptotically Gaussian with mean h^2 and variance

$$\frac{\tilde{\psi}^2}{n} = \frac{2}{n} \left(\frac{m \mu_1^2}{n \mu_2} + 2 \frac{\mu_1 \mu_3}{\mu_2^2} h^2 - h^4 \right), \quad (40)$$

where $\mu_1 = 1$, μ_2 and μ_3 , given by (18), are the spectral moments of the covariance of $\mathbf{Z}$.

Step 2: To show that the GWASH estimator (20) is asymptotically equivalent to the estimator (38) based on $\mathbf{Z}$, we need to approximate: (i) $\|\tilde{\mathbf{X}}^T \tilde{\mathbf{y}}\|^2/m$ by $\|\mathbf{Z}^T \tilde{\mathbf{y}}\|^2/m$, and (ii) $\hat{\mu}_1$ by $\tilde{m}_1$ and $\hat{\mu}_2$ by $\tilde{m}_2$.

(i) Let $\mathbf{D} = \text{Diag}(\mathbf{S})$ so that $\tilde{\mathbf{X}}^T = \mathbf{D}^{-1/2} \mathbf{X}^T = \mathbf{D}^{-1/2} \Delta^{1/2} \mathbf{Z}^T$. Define $\mathbf{v} = \mathbf{Z}^T \tilde{\mathbf{y}}$. We may write

$$\|\tilde{\mathbf{X}}^T \tilde{\mathbf{y}}\|^2 = \|\mathbf{D}^{-1/2} \Delta^{1/2} \mathbf{Z}^T \tilde{\mathbf{y}}\|^2 = \mathbf{v}^T \mathbf{D}^{-1} \Delta \mathbf{v},$$

so

$$\begin{aligned} \left| \|\tilde{\mathbf{X}}^T \tilde{\mathbf{y}}\|^2 - \|\mathbf{Z}^T \tilde{\mathbf{y}}\|^2 \right| &= |\mathbf{v}^T (\mathbf{D}^{-1} \mathbf{\Delta} - \mathbf{I}) \mathbf{v}| = |\text{tr} [\mathbf{v} \mathbf{v}^T (\mathbf{D}^{-1} \mathbf{\Delta} - \mathbf{I})]| \\ &\leq \|\mathbf{v} \mathbf{v}^T\| \|\mathbf{D}^{-1} \mathbf{\Delta} - \mathbf{I}\| = \|\mathbf{v}\|^2 \|\mathbf{D}^{-1} \mathbf{\Delta} - \mathbf{I}\|, \end{aligned}$$

where the bars around matrices denote the Frobenius norm and the inequality is due to the Cauchy-Schwarz inequality for the Frobenius norm. By Lemma 2,

$$\|\mathbf{D}^{-1} \mathbf{\Delta} - \mathbf{I}\| = \left[\frac{1}{m} \|\mathbf{\Delta}^{-1} \mathbf{D}\|^2 \right]^{-1/2} \left[\frac{1}{m} \|\mathbf{\Delta}^{-1} \mathbf{D} - \mathbf{I}\|^2 \right]^{1/2} = O_P \left(\frac{1}{\sqrt{n}} \right).$$

Thus

$$\frac{1}{m} \|\tilde{\mathbf{X}}^T \tilde{\mathbf{y}}\|^2 = \frac{1}{m} \|\mathbf{Z}^T \tilde{\mathbf{y}}\|^2 \left(1 + O_P \left(\frac{1}{\sqrt{n}} \right) \right). \quad (41)$$

(ii) As stated in the proof of Proposition 2 of [6], applied to the estimator (38) based on $\mathbf{Z}$, we have that

$$\tilde{m}_1 = \mu_1 + O_P \left(\frac{1}{\sqrt{mn}} \right) = \hat{\mu}_1 + O_P \left(\frac{1}{\sqrt{mn}} \right), \quad \tilde{m}_2 = \mu_2 + O_P \left(\frac{1}{\sqrt{mn}} \right), \quad (42)$$

because $\hat{\mu} = \mu_1 = 1$. Furthermore, (23) together with (42) imply

$$\tilde{m}_2 = \hat{\mu}_2 + O_P \left(\frac{1}{n} \right) \quad (43)$$

because m/n converges to a constant.

Putting (41), (42) and (43) together in (38) and comparing to (19), we obtain that

$$\tilde{h}^2 = \hat{h}_{\text{GWASH}}^2 \left(1 + O_P \left(\frac{1}{\sqrt{n}} \right) \right).$$

Hence $\hat{h}_{\text{GWASH}}^2$ has the same asymptotic distribution as $\tilde{h}^2$. Note that the variance (40) is the same as (24).

A.4 Derivation of the LDSC regression equation (33)

In our notation, the chi-squared statistics defined in Section 1.2 of the Supplementary Note to [2] can be written as $\chi_j^2 = n(\tilde{\mathbf{x}}_j^T \tilde{\mathbf{y}}/n)^2$, because both the predictors and the response in [2] are assumed standardized. Using (16), we can write $\chi_j^2 = u_j^2(n-1)/n \approx u_j^2$ for large n . Then Eq. (1.3) in [2] becomes

$$\mathbb{E}[u_j^2 \mid \ell_j] \approx h^2 \frac{n}{m} \ell_j + 1, \quad j = 1, \dots, m, \quad (44)$$

where $\ell_j = \sum_{k=1}^m r_{jk}^2$ and $r_{jk} = \mathbb{E}[\tilde{\mathbf{x}}_j^T \tilde{\mathbf{x}}_k/n]$. However, the Online Methods section in [2] explains that the empirical LD-scores are biased and cannot be used directly in (44). Instead, they use the adjusted LD-scores

$$\ell_{j,\text{adj}} = \sum_{k=1}^m r_{jk,\text{adj}}^2 = \sum_{k=1}^m \left(\tilde{r}_{jk}^2 - \frac{1 - \tilde{r}_{jk}^2}{n-1} \right) = \hat{\ell}_j - \frac{m - \hat{\ell}_j}{n-1},$$

where $\hat{\ell}_j = \sum_{k=1}^m \tilde{r}_{jk}^2$ and we have used $n - 1$ in the adjustment instead of $n - 2$ to reflect the fact that the relevant LD-regression fit in our case uses a fixed intercept. Using the adjusted LD-scores in (44), we have that

$$E[u_j^2 \mid \hat{\ell}_j] \approx h^2 \frac{n}{m} \left(\hat{\ell}_j - \frac{m - \hat{\ell}_j}{n - 1} \right) + 1 = h^2 \left[\frac{n}{m} \hat{\ell}_j \left(1 + \frac{1}{n - 1} \right) - \frac{n}{n - 1} \right] + 1,$$

yielding (33) for large n .

Acknowledgments

This work was partially supported by NIH grant 1R01GM104400 and The Lundbeck Foundation Initiative for Integrative Psychiatric Research.

References

- [1] Peter J. Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *Ann. Statist.*, 36:199–227, 2008.
- [2] B. K. Bulik-Sullivan, P. R. Loh, H. K. Finucane, S. Ripke, J. Yang, C. Schizophrenia Working Group of the Psychiatric Genomics, N. Patterson, M. J. Daly, A. L. Price, and B. M. Neale. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*, 47:291–295, 2015.
- [3] T. Tony Cai, Cun-Hui Zhang, and Harrison Zhou. Optimal rate of convergence for covariance matrix estimation. *Ann. Statist.*, 38:2118–2144, 2010.
- [4] C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4:7, 2015.
- [5] C. A. de Leeuw, J. M. Mooij, T. Heskes, and D. Posthuma. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology*, 11:e1004219, 2015.
- [6] Lee H. Dicker. Variance estimation in high-dimensional linear models. *Biometrika*, 101(2): 269–284, 2014.
- [7] R. C. Elston. On the correlation between correlations. *Biometrika*, 62(1):133–140, 1975.
- [8] D. S. Falconer and T. F. C. Mackay. *Introduction to quantitative genetics*, 4th ed. Longman, 1996.
- [9] R. A. Fisher. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52:399–433, 1918.
- [10] C. Genomes Project, A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, and G. R. Abecasis. A global reference for human genetic variation. *Nature*, 526:68–74, 2015.
- [11] Marius Hofert, Ivan Kojadinovic, Martin Maechler, and Jun Yan. *copula*: Multivariate dependence with copulas. *r* package version 0.999-9, 2014. URL <http://CRAN.R-project.org/package=copula>.

- [12] Ivan Kojadinovic and Jun Yan et al. Modeling multivariate distributions with continuous margins using the copula r package. *Journal of Statistical Software*, 34(9):1–20, 2010.
- [13] Y. Li, C. Willer, S. Sanna, and G. Abecasis. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Genotype imputation*, 10:387–406, 2009.
- [14] A. E. et al. Locke. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518:197–206, 2015.
- [15] M. Lynch and B. Walsh. *Genetics and analysis of quantitative traits. Vol. 1*. Sinauer Sunderland, MA, 1998.
- [16] J. MacArthur, E. Bowler, M. Cerezo, L. Gil, P. Hall, E. Hastings, H. Junkins, A. McMahon, A. Milano, J. Morales, Z. M. Pendlington, D. Welter, T. Burdett, L. Hindorff, P. Flicek, F. Cunningham, and H. Parkinson. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res*, 45, 2017.
- [17] A. et al. Okbay. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, 533:539–542, 2016.
- [18] B. Pasaniuc and A. L. Price. Dissecting the genetics of complex traits using summary association statistics. *Nat Rev Genet*, 18:117–127, 2017.
- [19] A. J. Schork, W. K. Thompson, P. Pham, A. Torkamani, J. C. Roddey, P. F. Sullivan, J. R. Kelsoe, M. C. O’Donovan, H. Furberg, C. Tobacco, Genetics, C. Bipolar Disorder Psychiatric Genomics, C. Schizophrenia Psychiatric Genomics, N. J. Schork, O. A. Andreassen, and A. M. Dale. All SNPs are not created equal: genome-wide association studies reveal a consistent pattern of enrichment among functionally annotated SNPs. *PLoS Genetics*, 9:e1003449, 2013.
- [20] S. Snieder, S. Stringer, K. Watanabe, P. R. Jansen, J. R. I. Coleman, E. Krapohl, E. Taske-sen, A. R. Hammerschlag, A. Okbay, D. Zabaneh, N. Amin, G. Breen, D. Cesarini, C. F. Chabris, W. G. Iacono, M. A. Ikram, M. Johannesson, P. Koellinger, J. J. Lee, P. K. E. Magnusson, M. McGue, M. B. Miller, W. E. R. Ollier, A. Payton, N. Pendleton, R. Plomin, C. A. Rietveld, H. Tiemeier, C. M. van Duijn, and D. Posthuma. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics*, 49:1107–1112, 2017.
- [21] S. L. Spain and J. C. Barrett. Strategies for fine-mapping complex traits. *Hum Mol Genet*, 24:R111–119, 2015.
- [22] D. Speed, N. Cai, U. Consortium, M. R. Johnson, S. Nejentsev, and D. J. Balding. Reevaluation of SNP heritability in complex human traits. *Nature Genetics*, 49:986–992, 2017.
- [23] A. R. et al. Wood. Defining the role of common variation in the genomic and biological architecture of adult human height. *Nature Genetics*, 46:1173–1186, 2014.
- [24] J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42:565–569, 2010.
- [25] X. Zhou, P. Carbonetto, and M. Stephens. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genetics*, 9:e1003264, 2013.