Submitted to the Annals of Applied Statistics

A SIMPLE, CONSISTENT ESTIMATOR OF SNP HERITABILITY FROM GENOME-WIDE ASSOCIATION STUDIES: SUPPLEMENTARY MATERIAL

By Armin Schwartzman[*], Andrew J. Schork[†], Rong Zablocki[*] and Wesley K. Thompson[*,†]

[*] *University of California, San Diego, La Jolla, CA and*
[†] *Institute of Biological Psychiatry, Mental Health Center St. Hans, Mental Health Services Copenhagen, Roskilde, Denmark*

S1. Shorter derivations and proofs.

Derivation of the Dicker estimator (8). Written in our notation, Dicker (Dicker, 2014, Section 4.1) proposes the estimators

$$\hat\sigma^2 = \frac{m+n}{n(n-1)}\|\boldsymbol{y}\|^2 - \frac{1}{n(n-1)}\|\hat{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}\|^2,$$
$$\hat\tau^2 = -\frac{m}{n(n-1)}\|\boldsymbol{y}\|^2 + \frac{1}{n(n-1)}\|\hat{\boldsymbol{\Sigma}}^{-1/2}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}\|^2,$$

where n has been replaced by $n-1$ due to the centering of $\boldsymbol{y}$ and $\boldsymbol{X}$. The estimator (8) is obtained as the fraction $\hat h_I^2 = \hat\tau^2/(\hat\sigma^2+\hat\tau^2)$. □

Derivation of the Dicker estimator (10). Written in our notation, Dicker (Dicker, 2014, Section 4.2) proposes the estimators

$$\tilde\sigma^2 = \left(1+\frac{m\hat m_1^2}{n\hat m_2}\right)\frac{1}{n-1}\|\boldsymbol{y}\|^2 - \frac{\hat m_1}{n(n-1)\hat m_2}\|\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}\|^2,$$
$$\hat\tau^2 = -\frac{m\hat m_1^2}{n(n-1)\hat m_2}\|\boldsymbol{y}\|^2 + \frac{\hat m_1}{n(n-1)\hat m_2}\|\boldsymbol{X}^{\mathsf{T}}\boldsymbol{y}\|^2,$$

where n has been replaced by $n-1$ due to the centering of $\boldsymbol{y}$ and $\boldsymbol{X}$. The estimator (10) is obtained as the fraction $\hat h_{II}^2 = \tilde\tau^2/(\tilde\sigma^2+\tilde\tau^2)$. □

Proof of Proposition 1. From (25), using (19) and (20),

$$\sum_{k=1}^K \frac{\hat\mu_{2,k}}{\hat\mu_2}\hat h^2_{\mathrm{GWASH},k} = \sum_{k=1}^K \frac{m_k}{n\hat\mu_2}(s_k^2-1) = \frac{1}{n\hat\mu_2}\sum_{j\in J}(u_j^2-1) = \hat h^2_{\mathrm{GWASH}}.$$

imsart-aoas ver. 2014/10/16 file: Heritability-AOAS-Supplement.tex date: August 25, 2019

To show (28), using (21) and the second expression in (26),

$$\sum_{k=1}^{K} \frac{m_k}{m} \hat{\mu}_{2,k} = 1 + \frac{1}{m} \sum_{k=1}^{K} \sum_{i \neq j \in J_k} \left[\left(\tilde{S}_{ij}^{(k)} \right)^2 - \frac{1}{n-1} \right]$$
$$= 1 + \frac{1}{m} \sum_{i \neq j \in J} \left[\tilde{S}_{ij}^2 - \frac{1}{n-1} \right] + o\left(\frac{1}{n}\right) = \hat{\mu}_2 + o\left(\frac{1}{n}\right).$$

The second equality is due to independence between sets; in this case the squared cross-correlation terms between sets average approximately $1/(n-1)$ and so the overall average contribution of the cross-terms between sets to $\hat{\mu}_2$ is of smaller order. □

PROOF OF PROPOSITION 2. From (6) and (7),

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

so $\|\boldsymbol{y}\|^2 = \hat{\sigma}_j^2(t_j^2 + n - 2)$. Thus, using (5) and again (7), we can write

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

and the result follows. □

S2. 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 $(1 - \tilde{\Sigma}_{jk}^2)^2/n$ and $(1 - \tilde{\Sigma}_{lh}^2)^2/n$. The covariance $\mathrm{Cov}(\tilde{S}_{jk}, \tilde{S}_{lh})$, whose expression is given by **?**, 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

$$\text{(S1)} \qquad \begin{aligned} \mathrm{E}(\tilde{S}_{jk}^2) &= \tilde{\Sigma}_{jk}^2 + \frac{(1 - \tilde{\Sigma}_{jk}^2)^2}{n} + o\left(\frac{1}{n}\right), \\ \mathrm{Cov}(\tilde{S}_{jk}^2, \tilde{S}_{lh}^2) &= 2\left[\mathrm{Cov}(\tilde{S}_{jk}, \tilde{S}_{lh})\right]^2 + o\left(\frac{1}{n}\right) = O\left(\frac{1}{n^2}\right). \end{aligned}$$

Now, from (21) and (S1),

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

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

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

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

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

implying (22). □

LEMMA 2. *Let* $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{S})$ *and* $\mathbf{\Delta} = \mathrm{Diag}(\mathbf{\Sigma})$. *Under Assumption 1:*

$$(i) \qquad \frac{1}{m}\|\mathbf{\Delta}^{-1}\boldsymbol{D} - \boldsymbol{I}\|^2 = \frac{2}{n} + O_P\left(\frac{1}{n\sqrt{m}}\right),$$

$$(ii) \qquad \frac{1}{m}\|\mathbf{\Delta}^{-1}\boldsymbol{D}\|^2 - 1 = \frac{2}{n} + O_P\left(\frac{1}{n\sqrt{m}}\right).$$

PROOF. For the purpose of simplicity of the proof, assume that the columns of $\boldsymbol{X}$ have not been centered so that the rows of $\boldsymbol{X}$ are independent. This makes no difference asymptotically, only changing de number of degrees of freedom $n-1$ to n. Let $W_j = \|\boldsymbol{x}_j\|^2/[n\Sigma_{jj}]$ be the j-th diagonal entry of $\mathbf{\Delta}^{-1}\boldsymbol{D}$ 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 $\|\mathbf{\Delta}^{-1}\boldsymbol{D} - \boldsymbol{I}\|^2/m = \sum_{j=1}^m (W_j - 1)^2/m$ is equal to $\mathrm{Var}(W_j) = 2/n$. Its variance is

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

Since the rows of $\boldsymbol{X}$ 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 $\tilde{\mathbf{\Sigma}}$), 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 $\boldsymbol{X}$,

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

Hence,

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

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

(ii) Write

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

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

□

S3. Proof of Theorem 1. Consider a situation where, instead of $\boldsymbol{X}$, the observed covariate matrix is $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{\Delta}^{-1/2}$ with $\boldsymbol{\Delta} = \mathrm{Diag}(\boldsymbol{\Sigma})$. The idea of the proof is to approximate the GWASH estimator (19) by an estimator based on $\boldsymbol{Z}$. This is done in two steps by: (1) establishing the asymptotic properties of the estimator based on $\boldsymbol{Z}$; (2) establishing the asymptotic equivalence of the two estimators.

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

Step 1: Based on $\boldsymbol{Z}$, the full model (2) can be written as

$$\boldsymbol{y} = \boldsymbol{Z}\tilde{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}, \tag{S2}$$

where $\tilde{\boldsymbol{\beta}} = \boldsymbol{\Delta}^{1/2}\boldsymbol{\beta}$. Note that $\tau^2 = \boldsymbol{\beta}^{\mathsf{T}}\boldsymbol{\Sigma}\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}^{\mathsf{T}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\beta}}$, so the variance fraction (4) does not change. The matrix $\boldsymbol{Z}$ has rows with covariance $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Delta}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Delta}^{-1/2}$ and its sample covariance matrix is $\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{Z}/(n-1) = \boldsymbol{\Delta}^{-1/2}\boldsymbol{S}\boldsymbol{\Delta}^{-1/2}$. Following the form of Dicker's estimator (10) for unestimable covariance, the estimator of h^2 based on $\boldsymbol{Z}$ is

$$\tilde{h}^2 = \frac{m\tilde{m}_1^2}{n\tilde{m}_2}\left(\frac{\|\boldsymbol{Z}^{\mathsf{T}}\boldsymbol{y}\|^2}{m\tilde{m}_1\|\boldsymbol{y}\|^2} - 1\right) = \frac{m\tilde{m}_1^2}{n\tilde{m}_2}\left(\frac{\|\boldsymbol{Z}^{\mathsf{T}}\tilde{\boldsymbol{y}}\|^2}{m\tilde{m}_1(n-1)} - 1\right), \tag{S3}$$

where
(S4)
$$\tilde{m}_1 = \frac{1}{m}\mathrm{tr}(\boldsymbol{\Delta}^{-1/2}\boldsymbol{S}\boldsymbol{\Delta}^{-1/2}), \quad \tilde{m}_2 = \frac{1}{m}\mathrm{tr}[(\boldsymbol{\Delta}^{-1/2}\boldsymbol{S}\boldsymbol{\Delta}^{-1/2})^2] - \frac{m}{n-1}\tilde{m}_1^2.$$

Applying Proposition 2 of Dicker (2014) to the estimator (S3) gives that this estimator is asymptotically Gaussian with mean h^2 and variance

$$\frac{\tilde{\psi}^2}{n} = \frac{2}{n}\left(\frac{m\mu_1^2}{n\mu_2} + 2\frac{\mu_1\mu_3}{\mu_2^2}h^2 - h^4\right), \tag{S5}$$

where $\mu_1 = 1$, μ_2 and μ_3, given by (17), are the spectral moments of the covariance of $\boldsymbol{Z}$.

Step 2: To show that the GWASH estimator (19) is asymptotically equivalent to the estimator (S3) based on $\boldsymbol{Z}$, we need to approximate: (i) $\|\tilde{\boldsymbol{X}}^{\mathsf{T}}\tilde{\boldsymbol{y}}\|^2/m$ by $\|\boldsymbol{Z}^{\mathsf{T}}\tilde{\boldsymbol{y}}\|^2/m$, and (ii) $\hat{\mu}_1$ by $\tilde{m}_1$ and $\hat{\mu}_2$ by $\tilde{m}_2$.

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

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

so

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

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 3,

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

Thus

$$\frac{1}{m}\|\tilde{\boldsymbol{X}}^{\mathsf{T}}\tilde{\boldsymbol{y}}\|^2 = \frac{1}{m}\|\boldsymbol{Z}^{\mathsf{T}}\tilde{\boldsymbol{y}}\|^2\left(1 + O_P\left(\frac{1}{\sqrt{n}}\right)\right). \tag{S6}$$

(ii) As stated in the proof of Proposition 2 of Dicker (2014), applied to the estimator (S3) based on $\boldsymbol{Z}$, we have that
(S7)
$$\tilde{m}_1 = \mu_1 + O_P\left(\frac{1}{\sqrt{mn}}\right) = \hat{\mu}_1 + O_P\left(\frac{1}{\sqrt{mn}}\right), \qquad \tilde{m}_2 = \mu_2 + O_P\left(\frac{1}{\sqrt{mn}}\right),$$

because $\hat{\mu} = \mu_1 = 1$. Furthermore, (22) together with (S7) imply

$$\tilde{m}_2 = \hat{\mu}_2 + O_P\left(\frac{1}{n}\right) \tag{S8}$$

because m/n converges to a constant.

Putting (S6), (S7) and (S8) together in (S3) and comparing to (18), we obtain that

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

Hence $\hat{h}^2_{\text{GWASH}}$ has the same asymptotic distribution as $\tilde{h}^2$. Note that the variance (S5) is the same as (23).

S4. Computationally efficient estimation of the third spectral moment $\hat{\boldsymbol{\mu}}_3$. As in Section 4.4, an approximation to $\hat{\mu}_3$ can be obtained by only considering entries of $\tilde{\boldsymbol{S}}$ close to the diagonal. Let $\mathcal{I}_2$ and $\mathcal{I}_3$ 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

$$\hat{\mu}_{3,\mathcal{I}_3} = 1 + \frac{1}{m}\left[\sum_{(i,j,k)\in\mathcal{I}_3} \tilde{S}_{ij}\tilde{S}_{jk}\tilde{S}_{ik} - 3\frac{|\mathcal{I}_2|}{n-1}\hat{\mu}_2 - \frac{|\mathcal{I}_3|}{(n-1)^2}\right], \tag{S9}$$

where $|\mathcal{I}_2|$ and $|\mathcal{I}_3|$ are the number of elements in the sets $\mathcal{I}_2$ and $\mathcal{I}_3$, respectively.

Specifically, for a single chromosome with m_k markers, let $\tilde{\boldsymbol{S}}_q$ be the restricted matrix (31) so that

$$\mathcal{I}_2 = \{(i,j) : 1 < |i-j| \leq q\},$$
$$\mathcal{I}_3 = \{(i,j,k) : 1 < |i-j| \leq q, 1 < |i-k| \leq q, 1 < |j-k| \leq q\}.$$

the number of elements $|\mathcal{I}_2|$ and $|\mathcal{I}_3|$ can be computed exactly as

$$|\mathcal{I}_2| = q(2m_k - q - 1)$$
$$|\mathcal{I}_3| = q(q-1)(3m_k - 2q - 2).$$

Replacing in (S9) gives the formula

$$\hat{\mu}^{(k)}_{3,q} = \frac{1}{m_k}\left[\operatorname{tr}\left(\tilde{\boldsymbol{S}}^{(k)}_q\right)^3 - 3\frac{q(2m_k - q - 1)}{n-1}\hat{\mu}_2 - \frac{q(q-1)(3m_k - 2q - 2)}{(n-1)^2}\right]. \tag{S10}$$

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

Again, for K chromosomes, the overall estimate $\hat{\mu}_{3,q}$ is the weighted average of the per-chromosome estimates (32), weighted by the number of markers m_k in each chromosome.

S5. Derivations related to LDSC regression.

DERIVATION OF THE LDSC REGRESSION EQUATION (35). In our notation, the chi-squared statistics defined in Section 1.2 of the Supplementary Note to **?** can be written as $\chi_j^2 = n(\tilde{\boldsymbol{x}}_j^{\mathsf{T}}\tilde{\boldsymbol{y}}/n)^2$, because both the predictors and the response in **?** are assumed standardized. Using (15), we can write $\chi_j^2 = u_j^2(n-1)/n \approx u_j^2$ for large n. Then Eq. (1.3) in **?** becomes

$$\text{(S11)} \qquad \mathrm{E}[u_j^2 \mid \ell_j] \approx h^2 \frac{n}{m}\ell_j + 1, \qquad j = 1, \ldots, m,$$

where $\ell_j = \sum_{k=1}^m r_{jk}^2$ and $r_{jk} = \mathrm{E}[\tilde{\boldsymbol{x}}_j^{\mathsf{T}}\tilde{\boldsymbol{x}}_k/n]$. However, the Online Methods section in **?** explains that the empirical LD-scores are biased and cannot be used directly in (S11). 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 (S11), we have that

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

yielding (35) for large n. □

DERIVATION OF THE LDSC REGRESSION EQUIVALENCE (38). From (37), note that $\overline{u^2} = \sum_{j=1}^m u_j^2/m = s^2$ by (20) and

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

by (21), assuming that m/n converges to a constant. Replacing in (37) and comparing to (19), we obtain that

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

yielding (38). □

References.

Dicker, L. H. (2014). Variance estimation in high-dimensional linear models. *Biometrika* **101** 269-284.

Division of Biostatistics
University of California, San Diego
La Jolla, CA
E-mail: armins@ucsd.edu
wes.stat@gmail.com

Department of Psychiatry
University of California, San Diego
La Jolla, CA
E-mail: rzablocki@ucsd.edu

Institute of Biological Psychiatry
Mental Health Center St. Hans
Mental Health Services Copenhagen
Roskilde, Denmark
E-mail: Andrew.Schork@regionh.dk