Title

On Negative Heritability and Negative Estimates of Heritability

Journal

Genetics, 215(2)

ISSN

0016-6731

Authors

Steinsaltz, David
Dahl, Andy
Wachter, Kenneth W

Publication Date

2020-06-01

DOI

10.1534/genetics.120.303161

Peer reviewed

On Negative Heritability and Negative Estimates of Heritability

David Steinsaltz,*[,1] Andy Dahl,[†,1] and Kenneth W. Wachter[§]

**Department of Statistics, University of Oxford, OX1 3TG, United Kingdom, [†]Department of Medicine, University of California, San Francisco, California, 94143, [‡]Section of Genetic Medicine, University of Chicago, Chicago, IL 60637, and [§]Department of Demography, University of California, Berkeley, California 94720

ORCID IDs: 0000-0001-6520-4766 (A.D.); 0000-0003-2802-2212 (K.W.W.)

ABSTRACT We consider the problem of interpreting negative maximum likelihood estimates of heritability that sometimes arise from popular statistical models of additive genetic variation. These may result from random noise acting on estimates of genuinely positive heritability, but we argue that they may also arise from misspecification of the standard additive mechanism that is supposed to justify the statistical procedure. Researchers should be open to the possibility that negative heritability estimates could reflect a real physical feature of the biological process from which the data were sampled.

KEYWORDS Heritability; GREML; Linear mixed model; Epistasis; Model misspecification

THE past decade has seen a proliferation of statistical methods for estimating heritability from large genome-wide genetic data sets. In particular, genomic-relatedness-based restricted maximum-likelihood (GREML; Visscher *et al.* 2006; Yang *et al.* 2010) has emerged as a standard tool in statistical genetics, along with related procedures such as Haseman–Elston (HE) and LD score regression (Bulik-Sullivan *et al.* 2015 *preprint*; Wu and Sankararaman 2018) that are constructed on the same statistical foundation. In many settings, these approaches can be invaluable for demonstrating the existence and approximate level of heritability by aggregating small genetic effects distributed across the genome. This is practically useful for complex traits given that mapping most causal genetic variants remains difficult (Manolio *et al.* 2009).

Despite its widespread and often indiscriminate application, GREML depends on very strong assumptions that are impossible to verify in detail, are not believed to be literally true, and are rarely subjected to any formal diagnostic or even qualitative critical consideration. In particular, the underlying statistical model assumes strictly additive causal genetic effects. Even if the additive model is accepted as a sensible foundation, nongenetic effects may reshape the appearance of genetic influences. In this paper we take an instrumental approach to GREML and related procedures: heritability measures arise from these calculations regardless of any connection to the additive model, and these need to be interpreted. In particular, we focus on a phenomenon that is typically dismissed as impossible or meaningless: negative heritability.

While the heritability parameter in the GREML model is mathematically compelled to be nonnegative, we explain how a broader view—not a new view, but one close to the root conception of heritability—implies that values of heritability meaningfully extend into the negative range, and hence that negative estimates of heritability can be taken seriously. It is only the extraneous (and not strictly credible) assumptions of GREML's motivating model that would exclude negative estimates. We buttress this intuition with a biologically plausible story linked to a mathematically coherent model, where negative heritability estimates arising from the standard GREML procedure are meaningful indicators of causal biological processes.

Operational Definitions of Heritability

As Albert Jacquard pointed out decades ago (Jacquard 1983), the narrow-sense heritability of a phenotype, commonly denoted h^2, has two distinct conventional meanings:

doi: https://doi.org/10.1534/genetics.120.303161
Manuscript received November 20, 2019; accepted for publication March 30, 2020; published Early Online April 14, 2020.
[1]Corresponding authors: Department of Statistics, University of Oxford, 24–29 St. Giles, Oxford OX1 3TG, United Kingdom. E-mail: steinsal@stats.ox.ac.uk; and The University of Chicago, Section of Genetic Medicine, 5841 S. Maryland Ave, Chicago, IL 60637. E-mail: andywdahl@gmail.com

1. The proportion of total variance attributable to additive genetic effects.
2. The slope of the linear regression of children's phenotypes on the mean parental phenotypes.

Both meanings appear in the earliest works to give a quantitative operational definition to heritability, in particular Lush (1940). For more on the history of the notion of heritability, see Bell (1977).

The nexus between these two meanings is an additive model, where genetic and nongenetic effects are independent and sum together to produce the phenotype. When we have general genetic relatedness (rather than parental relations with fixed 50% expected relatedness), heritability is analogous to a regression coefficient that relates phenotypic similarity to genotypic similarity.

We are particularly concerned here with the interpretation of negative estimates of heritability. The appearance of negative estimates for a parameter of crucial scientific interest that is *prima facie* positive is unusual, as has often been noted. Negative estimates of the heritability parameter are often dismissed as a mathematical abstraction, values in a range that arises purely formally and that may only be reported for formal purposes. For example, Johnston *et al.* (2010) obtain a point estimate of -0.109 for the heritability of horn length in Soay sheep, which is immediately dismissed with the statement that "it is impossible to have negative heritability." The inference is drawn that the true heritability must actually be a small positive number toward the upper end of the confidence interval.

One case where negative heritability estimates have been used in practice is for estimating the average heritability across a group of exchangeable phenotypic measurements, like gene expression. In this case, negative estimates are reported under the presumption that this yields a complete ensemble of estimates that are collectively unbiased. We illustrate one such analysis using RNA-sequencing data from the Genetic European Variation in Health and Disease (GEUVADIS) consortium (Lappalainen *et al.* 2013). One significant contribution of our work is our calculation of the bias imparted to the heritability estimate when negative estimates are suppressed, helping to elucidate the conditions under which this bias may be presumed negligible.

Our fundamental argument is that negative heritability estimates need to be taken more seriously. The confusion, we contend, comes from the overlap between statistical models that operationalize the two different interpretations of heritability described above. The argument for rejecting negative estimates appears compelling just so long as the focus is only on the additive random-effects model in Equation 1 that often motivates definition 1 of heritability. Variance is nonnegative, hence the ratio of two variances cannot be negative.

While "variance attributable to additive genetic effects" is a basic element of the genetic model in (1), it has no place in the statistical algorithms commonly used to fit the model to real data, including GREML. As our later discussion of (1) will make clear, the GREML estimate of heritability is defined to serve as an estimator of a ratio of two variances, where the numerator is a component of the denominator. The ratio is constrained to lie between 0 and 1, so the estimate seems intended to lie between 0 and 1. However, as we shall explain, the GREML estimate is realized under a more general multivariate normal model, where the natural constraint on h^2 is weaker: $h^2 \geq -1/(\max\{s_i^2\} - 1)$, where $(s_i)_{i=1}^n$ are the singular values of the genotype matrix. If the phenotypes were derived from summing independent additive genetic effects, then the true h^2 would have to be nonnegative, but that must be recognized as an additional assumption that would need to be scientifically warranted, as it is not compelled on strictly logical or mathematical grounds.

This discordance between common practice and formal probability theory manifests itself in two distinct roles in modern genetic analyses. First, alleles can exercise actual causal influences on traits, or can tag causal genetic influences through linkage, and such contributions cannot generate negative heritability. However, second, alleles also serve as markers of family and ancestry, markers of relatedness among individuals that may structure historical, behavioral, social, and environmental influences on traits. We argue that there is no reason to assume nonnegative heritability in this latter, more general class of models. As attention expands beyond basic additive genetic models to more complex characterizations of genome-wide genetic architecture, it is important to understand the behavior of h^2 beyond its intuitive definition grounded in classical quantitative genetics.

The meaning of negative heritability

Heritability is not a natural, measurable quantity. Heritability is defined only by its role in a model, and the model is inevitably misspecified. The normally distributed random genetic effects have no physical reality, and function instead primarily to justify a model of convenience. In general, the heritability of a trait will vary across populations, measurement devices, choice of scale, and countless environmental factors.

The term "negative heritability" appeared for the first time, so far as we are aware, in a paper by J. B. S. Haldane, written around 1960, but first published posthumously in 1996 (Haldane 1996). Haldane described how the maternal-effect trait of neonatal jaundice could be said to display negative heritability: Because the disease results from maternal antibodies against a fetal antigen, it will not arise in a fetus whose mother herself experienced neonatal jaundice (we thank Jonathan Marchini for pointing out this reference to us). Haldane then calculates a negative heritability from a model that is specialized to the peculiar inheritance structure of this condition.

Once we have accepted the GREML multivariate normal framework, which we will define precisely, we must admit the possibility that the joint distribution of phenotypes and genotypes in a given data set may be best described by an h^2 value that is negative. The question this raises is, can such a

negative heritability estimate be biologically sensible? The heritability parameter may be identified, in a precise way, with a correlation between genotypic similarity and phenotypic similarity. The model invites us to select an estimate of h^2 that best matches the genetic covariance between individuals to the similarity in their traits. Even if we *want* heritability to be interpreted as a partition of variance, in the sense of definition 1, this will not always be correct. All we have access to from the data are an estimate of something like heritability in the sense of definition 2. High heritability means that individuals with similar genotypes tend to have similar trait values. Zero heritability means that genotypes tell us nothing about similarities in trait values. Negative heritability, then, could be a perfectly sensible description of data where individuals with similar genotypes tend to have more discordant traits than random pairs. In the special case of twin studies, for example, negative heritability simply means that monozygotic twins are less phenotypically similar than dizygotic twins.

Saying that a given set of data might be best described by a negative heritability estimate goes only part of the way toward answering the biological plausibility of negative true heritability. We cannot assume that a small sample of data pairs that are known (for scientific reasons) to be positively correlated will indeed yield a positive empirical correlation. Negative heritability could arise in the same way, as a spurious effect of random fluctuations in data from a system with zero or small positive heritability. The essential question is, could there be a plausible stochastic mechanism that would produce genuine negative heritability, so that as the amount of data generated by the model goes to infinity, the estimate converges to a negative quantity?

GREML is an optimization procedure derived under a Gaussian model, with a heritability parameter that makes good mathematical sense in the negative range. It would be perfectly straightforward to generate data from this model, but it might be difficult to interpret such a procedure in biologically meaningful terms. We seek, then, a negative heritability mechanism that has a similar form to the random-genetic-effects model, but is misspecified in a way that suggests a plausible story. We propose one such mechanism, based on the phenomenon of "phenotypic repulsion." As with Haldane's model (which may be understood as a special case), this mechanism has implications that may be implausible or even obviously false in a given real data set. It involves interactions between individuals that are not primarily genetic. However, the point we want to suggest is that as an abstract physical mechanism that could be producing our data it is as mathematically plausible as the linear random-effects model that undergirds GREML. This is only one example of such a mechanism, and the conclusion we advocate is that negative heritability must be acknowledged as a genuine phenomenon for genotype–phenotype data, even if it may be reasonably excluded by the context of some studies. Speculation about the biological settings that could yield negative heritability can also prove an effective guide to understanding when negative heritability estimates may be reliably truncated or ignored.

Our position parallels the advice on "interpretation of negative components of variance" propounded in a very different context by J. A. Nelder in 1954 (Nelder 1954). Nelder considered the problem of ANOVA testing on split-plot experiments, where the error for main plots was found to be smaller than the error for subplots, producing a negative estimate for the residual subplot error. As we have done here, Nelder showed how the apparently negative "variance component" could arise either from sampling error on a positive variance component or from a misspecification of the model, where correlations between measurements have been neglected. "In any particular situation," Nelder concludes, "it is the statistician's responsibility to decide which model is more appropriate."

The GREML model as linear regression

For the remainder of this paper we follow Steinsaltz *et al.* (2018) in using the letter ψ to represent heritability, to avoid the confusing implication built into the nomenclature h^2 that this parameter formally cannot be negative. Our derivations draw on the analysis in that paper, which also discusses criticisms of GREML, such as those in Krishna Kumar *et al.* (2016).

Underlying GREML, as well as alternative approaches to heritability estimation such as LD score and HE regression, is a random-effects model. Our basic object is a data set consisting of an $n \times p$ matrix Z, taken to represent the genotypes of n individuals, measured at p different loci. There is a vector $\mathbf{y}$, representing a scalar trait observation for each of the n individuals. The raw genotypes are counts of alleles taking the values 0, 1, or 2, but the genotype matrix is centered to have mean zero in each column and normalized to have mean square over the whole matrix equal to 1. (Often columns are further standardized to variance 1, but we do not make this assumption.) The model posits the existence of a random vector $\mathbf{u} \in \mathbb{R}^p$ of genetic influences from the individual SNPs such that

$$\mathbf{y} = Z\mathbf{u} + \varepsilon. \tag{1}$$

The vectors $\mathbf{u}$ and ε are assumed to have independent Gaussian entries with zero means and equal variances. The variances are determined by two parameters, which are to be estimated: θ represents the precision (reciprocal variance) of the nongenetic noise and ψ represents the heritability, entering the model as the ratio of genetic variance to total variance. We also use the notation $\phi = \psi/(1-\psi)$ in some places, for concision.

The GREML model has been formulated as a random-effects model, but it is equivalent to a multivariate normal model corresponding to the covariance matrix

$$C^2 := \theta_0^{-1}(\phi_0 A + I_n), \tag{2}$$

where $A = ZZ^*/p$ is the genetic relatedness matrix (GRM), and θ_0 and Ψ_0 are the true values of the parameters. In this

section we describe how the model may also be understood as a linear regression model.

The initial GCTA paper (Yang *et al.* 2010) spelled out an analogy between GCTA and a different form of linear regression, regressing squared trait differences between pairs of individuals on corresponding off-diagonal elements of the GRM, with $n(n-1)/2$ points and correlated errors. This is essentially HE regression, which has recently become a popular heritability estimation method due to its speed and robustness to some degree of model misspecification (Chen 2014; Golan *et al.* 2014). Instead, we draw an approximate comparison between GREML and regression with n points and independent errors.

Let $Z/\sqrt{p} = U\,\mathrm{diag}(s_i)V^*$ be the singular-value decomposition of $Z/\sqrt{p}$, where X^* indicates the transpose of a matrix X. We rotate the observations to diagonalize the covariance matrix, obtaining

$$\mathbf{z} := U^*\mathbf{y}.$$

Because the columns of Z have zero means, one of the singular values is zero and the corresponding column of U is proportional to a vector with all elements equal to 1. Thus every other column of U sums to zero (because its columns are orthogonal), and hence each column defines a contrast between weighted groupings of individuals.

The elements of $\mathbf{z}$ are independent centered normal random variables, and z_i has variance $(1 + (\psi_0/(1-\psi_0))s_i^2)/\theta_0$. It follows that $z_i^2\theta_0(1-\psi_0)/(1+\psi_0(s_i^2-1))$ are independent chi-square random variables each on one degree of freedom and

$$\log z_i^2 = -\log(\theta_0) - \log(1-\psi_0) + \log\big(1+\psi_0(s_i^2-1)\big) + \epsilon_i^*,$$

where the ϵ_i^* are distributed as the logarithms of the independent chi-square variables, long-tailed to the left, with mean ≈ -1.302, SD ≈ 2.266, and skewness ≈ -1.643.

The mean of s_i^2 is 1, and when $\psi_0(s_i^2-1)$ are uniformly small we may approximate our equation by

$$\log z_i^2 \approx -\log(\theta_0) - [\psi_0 + \log(1-\psi_0)] + \psi_0 s_i^2 + \epsilon_i^*. \quad (3)$$

Here, Ψ_0 takes on the role of the true slope for a regression of $\log z_i^2$ on s_i^2. It can be estimated by least squares, bearing in mind that the skew of the ϵ_i^* affects SE of estimation.

Practitioners instead usually estimate ψ via (restricted) maximum likelihood. Obviously, the maximum likelihood estimate (MLE) is optimal when the underlying model assumptions hold. However, formally characterizing the behavior of the MLE is nontrivial, especially under nonindependent genotypes (*cf.* Jiang *et al.* 2016) or sparse, nonpolygenic architectures. For this reason, most theoretical mixed model analyses focus on regression-based approaches with simple analytic solutions, like HE regression or the eigenvalue regression in (3). In contrast, we derived an analytic approximation to the GREML estimate in Steinsaltz *et al.* (2018), which we used to demonstrate several important theoretical properties. First, the MLE has a small negative bias on the order of $1/n$, which is negligible in many settings. Second, if only k SNPs are causal, the MLE additionally suffers a nonrandom, nonasymptotically vanishing bias of order $1/k$. Finally, population structure tends to make GREML more efficient, at the cost of exposure to potential confounding. In this paper, we apply the same analytical framework to a different question: Are there plausible forms of model misspecification that yield truly negative MLE heritability?

Formally, Steinsaltz *et al.* (2018) shows how the MLEs can be expressed in terms of quantities $w_i(\psi) := (1-\psi)/(1+\psi(s_i^2-1))$ and $v_i(\psi) := w_i(\psi)z_i^2$. They satisfy

$$0 = \mathrm{Cov}\big(\mathbf{w}(\hat{\psi}), \mathbf{v}(\hat{\psi})\big)$$

$$\hat{\theta} = n \Big/ \sum_{i=1}^{n} v_i(\hat{\psi}) \quad (4)$$

Here, Cov is the empirical covariance of vector elements, an operation on vectors defined by $\mathrm{Cov}(\mathbf{x},\mathbf{y}) := n^{-1}\sum(x_i-\bar{x})(y_i-\bar{y})$, and Var is similarly defined. We also set $\tau_2(\psi) = \psi^{-2}\,\mathrm{Var}(w(\psi))$, and omit the dependence on ψ when helpful. When Ψ_0 is not too close to 1 and the variance of the squared singular values is small, the least-squares and MLEs are close to each other.

Suppose, however, that the true variances of the z_i include a phenotypic contribution that varies inversely with the singular values of Z. In the phenotypic repulsion model to be specified shortly, to first order in $s_i^2 - 1$ the true slope is $(\psi_0 - \alpha - \psi_0^2)/(1-\psi_0)^2$ as a function of a repulsion parameter α. When α exceeds Ψ_0, the true slope turns negative and estimated slopes correctly include negative values. From this regression-based perspective, there is no reason *prima facie* to assume heritability must be nonnegative.

Bias from rejecting negative heritability estimates

The common practice of truncating the maximum likelihood calculation to nonnegative values introduces bias that is well-known and may be serious for samples of moderate size, both when estimates are truncated at zero and when negative estimates are simply suppressed.

The problem of estimating the probability of negative heritability estimates was studied 50 years ago by Gill and Jensen (1968). We add here a few comments about how the framework described in Steinsaltz *et al.* (2018) may contribute to understanding the magnitude of the negative heritability estimate problem that arises from sampling noise in settings where the true heritability is understood to be nonnegative, hence where truncation at zero (or rejection of negative estimates) is warranted and guarantees improved estimates in, say, mean squared error. We gain a rough idea of

the effect of rejecting negative estimates from a normal approximation $\hat{\psi} - \psi_0 \approx \sigma_0 X$, where $\sigma_0 = \sqrt{2}(1-\psi_0)/\sqrt{n\,\tau_2}$ and X has standard normal distribution [see Steinsaltz *et al.* (2018) for derivation]. Here, $\approx$ means that the difference between the left-hand and right-hand sides is bounded (in probability) by a constant times $(n\tau_2)^{-1}$, where the constant may depend on Ψ_0.

Truncating estimates where $\hat{\psi} < 0$ by setting them equal to 0 imposes the truncation bias

$$\begin{aligned}\mathbb{E}[\hat{\psi}] - \psi_0 &\approx -\psi_0 + \mathbb{E}\left[(\psi_0 + \sigma_0 X)1\left\{X > -\frac{\psi_0}{\sigma_0}\right\}\right] \\ &= -\psi_0 \Phi\left(-\frac{\psi_0}{\sigma_0}\right) + \sigma_0 \int_{-\psi_0/\sigma_0}^{\infty} x \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathrm{d}x \qquad (5) \\ &= -\psi_0 \Phi\left(-\frac{\psi_0}{\sigma_0}\right) + \frac{\sigma_0}{\sqrt{2\pi}} e^{-\psi_0^2/2\sigma_0^2},\end{aligned}$$

where Φ is the standard normal cumulative distribution function (c.d.f.) Note that by standard inequalities for Φ (Feller 1968) this is positive and bounded by $\frac{\sigma_0^3}{\sqrt{2\pi}\psi_0^2} e^{-\psi_0^2/2\sigma_0^2}$ when $\Psi_0 > 0$. This will be very small when $n\tau_2$ is even moderately large compared with $1/\psi_0^2$, which is to be expected except when Ψ_0 is zero, or nearly zero. When $\Psi_0 = 0$ we have a nonnegligible positive bias of approximately the same size as the SE σ_0, and will thus be highly relevant for any statistical tests of the null hypothesis $\Psi_0 = 0$.

Truncation at zero will at least be recognizable, whereas tacit rejection of negative estimates may leave no trace due to publication bias. If we have an ensemble of $\hat{\psi}$ estimates that have been selected to be nonnegative, the average has a conditioning bias that is identical to the expression in (5) divided by $\Phi(\psi_0/\sigma_0) := \mathbb{P}\{X > -\psi_0/\sigma_0\}$. In the special case $\Psi_0 = 0$, this doubles the bias relative to truncation.

The phenotypic repulsion model

The notion that new species force their way into phenotypic gaps in the existing ecological community was termed by Darwin as the "principle of divergence" and has been further developed by ecologists under the name "phenotypic repulsion" or "phylogenetic repulsion" (Webb *et al.* 2002). Species living in close proximity, which are often closely related phylogenetically, coexist by separating from each other phenotypically. A similar kind of competitive exclusion has been proposed (Sulloway 2011) on the individual level to explain observed patterns of developmental variation within human families. Social niche formation within families has also been proposed by Conley *et al.* (2013), without an explicit mathematical model, as the basis for an evaluation of gene–environment interaction based on misclassified twin types.

Phenotypic repulsion has been more commonly described on the level of species differences than within species. Cardillo (2012) has described negative correlation between phylogenetic distance and flowering period difference among fire-killed but not fire-resistant *Banksia* species in southwestern Australia. A study of Florida oak species found that many traits differed more, between closely related species, than would be expected by chance (Cavender-Bares *et al.* 2004). We have not found quantitative studies of phenotypic repulsion between individuals within a species, but it seems plausible that local competition for sunlight combined with range-limited seed dispersion would yield an effective phenotypic repulsion between related plants in a forest setting. In human populations anecdotal evidence suggests that monozygotic twins seek to differentiate themselves from their sibling, which may be a stronger force than genetic similarity for traits with a negligible causal genetic basis.

While our focus is on biologically meaningful phenotypic repulsion, as in the examples above, the repulsion may also result from pure experimental artifacts. For example, in mega-analyses across institutes or laboratories, similarity between analytical or experimental procedures may correlate negatively with similarity in genetic ancestry. This induces repulsion in the sense that genetic similarity predicts experimental dissimilarity. Nonetheless, in this situation the resulting repulsion is not connected to a biologically meaningful process and, rather, would disappear under proper experimental protocols and/or correcting for potential technical confounders like laboratory and batch.

The model we propose here is novel, so may be criticized for failing to provide an example of negative heritability in an established ecological model already in use. We would argue that this model does describe a phenomenon of interest in ecology that has not yet been formalized, and so either the behavior it describes should be taken seriously, or it should provoke a better model of the phenomenon.

We propose a model of phenotypic repulsion where individuals that are most closely related genetically strive to avoid each other phenotypically. This starts with a model like the GREML model described above, where individuals have phenotypes determined by normally distributed effect sizes acting on their individual genotypes. We introduce a penalty term to the probability, of the form

$$\exp\left\{-\alpha\theta_0\left(\sum_{1\le i<j\le n} A_{ij} y_i y_j + \frac{1}{2}\sum_{1\le i\le n}(A_{ii}-1)y_i^2\right)\right\},$$

where $A_{ij} = \frac{1}{p}\sum_{k=1}^{p} Z_{ik}Z_{jk}$ is the (i, j) entry of the GRM, and $\alpha \le 1$ is a parameter measuring the extent to which genetically similar individuals are pushed to have differing phenotypes. Of course, this setup could be generalized to higher-dimensional phenotypes, with $y_i y_j$ replaced by an arbitrary inner product. The penalty term is inspired by the statistical mechanics models that have been applied to geographically structured population dynamics, such as the Contact Process (Liggett 1999), used to model the spread of epidemics.

Combining this specification with (2), we see that the phenotypes will now be multivariate normal with mean zero and covariance matrix

$$\theta_0^{-1}\left[(\phi_0 A + I_n)^{-1} + \alpha(A - I_n)\right]^{-1}. \qquad (6)$$

It follows that the transformed phenotypes $\mathbf{z} = U^* \mathbf{y}$ are independent normal with mean zero and variance

$$\mathrm{Var}(z_i) = \theta_0^{-1} \frac{1 + \phi_0 s_i^2}{1 - \alpha + \alpha s_i^2 (1 - \phi_0) + \alpha \phi_0 s_i^4}.$$

Suppose the data have come from this phenotypic–repulsion model, and we analyze them using the misspecified random-effects model. While it is always possible to get $\hat{\psi} < 0$ because of random fluctuations, we would like to show that the heritability implied by this model is "really" negative, in the sense that the distribution of $\hat{\psi}$ converges to a strictly negative value as the number of subjects goes to infinity. This will follow from *Proposition 1* (below), when we take the function f in that result to be

$$f(t) = \frac{1 + \phi_0 t}{1 - \alpha + \alpha(1 - \phi_0)t + \alpha \phi_0 t^2}, \tag{7}$$

as long as $\phi_0 < \alpha$, since

$$f'(t) = \frac{\phi_0 - \alpha(1 + \phi_0 t)^2}{(1 - \alpha + \alpha(1 - \phi_0)t + \alpha \phi_0 t^2)^2},$$

which is less than 0 for all $t \geq 0$.

In other words, to the extent that we say that heritability is defined by the linear model, heritability can be negative if genotypes and phenotypes interact through the environment in a manner like the phenotypic repulsion model. We prove that this is the case—that the heritability to which the estimates converge with increasing population size is negative—in the following Proposition, which is proved in Appendix A.

Proposition 1: *Suppose we have a family of $n \times n$ GRMs A_n for $n \to \infty$, with eigenvalues $s_{n,i}^2$, with*

$$s_{\max}^2 := \limsup_{n \to \infty} \max_{1 \leq i \leq n} s_{n,i}^2 < \infty, \tag{8}$$

$$\sup_n n^{-1} \sum s_{n,i}^{-12} < \infty, \text{ and} \tag{9}$$

$$1 < C_2 + 1 := \liminf_{n \to \infty} n^{-1} \sum_{i=1}^{n} s_{n,i}^4. \tag{10}$$

We also write $C_3 := \limsup_{n \to \infty} n^{-1} \sum_{i=1}^{n} s_{n,i}^4$.

Let $U^{(n)}$ be the corresponding eigenvector matrix. For each n we have a multivariate normal random vector $\mathbf{y}^{(n)}$ with covariance matrix $U^{(n)} \mathrm{diag}(f(s_{n,i}^2)) U^{(n)*}$, where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a positive, strictly decreasing, continuously differentiable function. We assume that

$$C_1 := \inf_{0 \leq t \leq s_{\max}^2} (-f'(t)) > 0. \tag{11}$$

(We maintain the normalization assumption that $\sum_i s_{n,i}^2 = n$.)

Let $\hat{\psi}_n$ be the MLE for an observation $\mathbf{y}^{(n)}$, calculated from the random-effects model with GRM A_n. Then defining

$$\delta := \frac{1}{6} \left(\frac{C_3 f(0)}{C_1 C_2} + s_{\max}^2 - 1 \right)^{-1}, \tag{12}$$

$\hat{\psi}$ is bounded above in probability by the strictly negative quantity $-\delta$ as $n \to \infty$. That is, the probability of $\hat{\psi}_n > -\delta$ goes to 0 as $n \to \infty$.

Although we focus on GREML in this paper, two other prominent approaches to estimate heritability in unrelated samples are HE regression and LD score regression. In Appendix B, we show that HE regression also converges to a negative heritability estimate under the phenotypic repulsion model. While it is simpler to analyze HE, the proof is similar to the proof of *Proposition 1*: in both cases the key fact is that larger eigenvalues of the kinship matrix are actually associated with lower phenotypic variance under phenotypic repulsion (Equation A3), which is the essence of negative heritability. Moreover, based on established approximate equivalences between HE regression and LD score regression (Bulik-Sullivan 2015; Zhou 2017), LD score regression likely also converges to negative heritability estimates under phenotypic repulsion.

Broadly, these and other estimates of heritability may be understood as approximations for the same parameter as in the GREML model, and hence may be expected to target a negative value for large n, as the various estimates converge. The key point is that none of these procedures is directly estimating a variance component. Each of them is estimating a covariance, and it is easy to see how these covariances can be negative.

Finally, we note that the ordinary asymptotic SE for GREML are no longer accurate under the phenotypic repulsion model. In Appendix C, we derive the asymptotic behavior of the SE under repulsion using the sandwich estimator. However, there seems to be no simple interpretation of the relationship between the genotype distribution and the SE as there is for the well-specified model.

Transcriptome-wide gene expression heritability

We conclude with a brief example that illustrates the practical significance of negative heritability estimates. Although negative estimates of heritability for a single fixed trait are rarely published, it is common to include negative estimates when profiling heritability across a large number of roughly exchangeable traits (Yang *et al.* 2013; Wright *et al.* 2014; Bhatia *et al.* 2015 *preprint*; Finucane *et al.* 2015; Zhu *et al.* 2015; Gusev *et al.* 2016; Gymrek *et al.* 2016; Hernandez *et al.* 2019). Characterizing such -omic-wide heritability is common in functional genomics, where high-throughput measurements of some genomic feature are made at thousands of genomic positions. The most common measurement is (RNA) gene expression, but other prominent examples include methylation levels, chromatin accessibility, expression response to stimuli, or protein expression.

We analyzed an RNA-sequencing data set from the GEUVADIS consortium (Lappalainen *et al.* 2013). We aligned

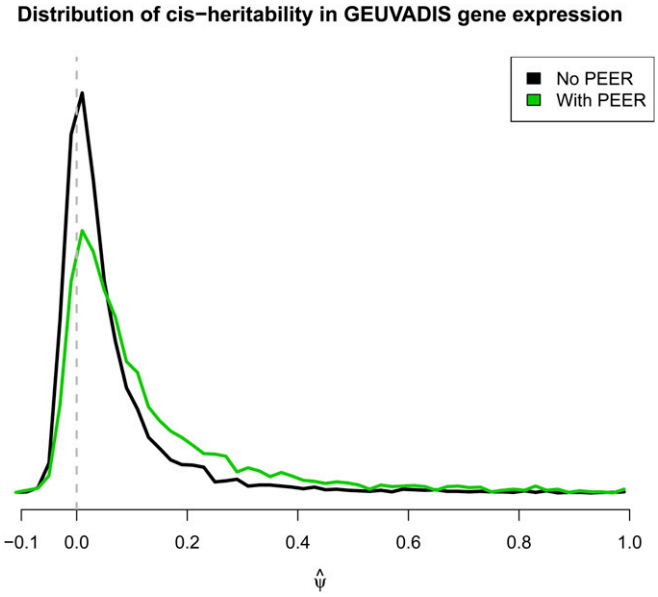

Figure 1 Transcriptome-wide density of gene expression *cis*-heritability estimates in the GEUVADIS data. For each gene, we estimate using GREML and a kinship based only on nearby SNPs.

the raw transcript reads from the European individuals to the reference hg19 transcriptome with the RSEM software package (Li and Dewey 2011). We removed perfectly correlated genes and genes with low expression mean or variance.

For each i in 375 people and j in 4154 genes, we define the phenotype $y_i^{(j)}$ as $\log(1 + n_{ij})$ where n_{ij} is the number of observed RNA reads for gene j measured in person i. We centered and scaled each gene to mean zero and variance one.

Separately for each gene $y^{(j)}$, we estimate its *cis*-heritability, that is, the heritability in expression levels explained by SNPs near to the gene. We do so by fitting our standard model (1) with a genotype matrix $Z^{(j)}$ whose columns correspond to SNPs located up to 1 megabase upstream or downstream of gene j's transcription start site. Restricting to SNPs near a gene is a common way to enrich for causal SNPs. We discard rare SNPs, which we define as SNPs with minor allele frequencies below 2.5%. Finally, we remove genes with fewer than 1000 corresponding SNPs, which excludes 35 genes.

The column dimensions (p) of the *cis* genotype matrices range from 1000 to 20,523 across genes, with a mean of 3027 and median of 2754. We fit each $\hat{\psi}$ with the maximum likelihood routine from Hernandez *et al.* (2019), yielding 4119 values reflecting systematic variation across genes in their *cis*-heritability, within the limits imposed by sampling error.

The distribution of the resulting transcriptome-wide *cis*-heritability estimates is shown in Figure 1 in the form of a smoothed histogram. Clearly, many of the estimates are negative. The mode is close to zero. Removing negative heritability estimates increases the transcriptome-wide average heritability from 6.2 to 9.0%, and truncating at zero increases it from 6.2 to 6.6%.

We repeated the analysis after adjusting for unobserved confounding estimated by 10 probabilistic estimation of expression residuals factors (Stegle *et al.* 2010). This practice, or variants based on gene expression principal components (Alter *et al.* 2000) or surrogate variables (Leek and Storey 2007), is standard practice in functional genomics (Stegle *et al.* 2012). The common aim of these approaches is to approximate latent confounding variation, like experimental batch effects, which can often be partially captured by dimensionality reduction. The confounder estimates are treated as known covariates and residualized from the phenotype and genotype data.

Correcting for 10 probabilistic estimation of expression residuals factors increases many of the $\hat{\psi}$ values and reduces the incidence of negative $\hat{\psi}$ as shown in the green curve in Figure 1. However, it is clear that many negative estimates remain. Negative estimates are bound to be part of the picture whenever Ψ_0 is small and estimated with low precision, both conditions that will likely hold in most functional genomic analyses for at least the near future.

On the question of whether some negative estimates may be meaningful reflections of nongenetic phenotypic structure, it is best to keep an open mind.

Data availability

GEUVADIS data were obtained from the GEUVADIS consortium. We fit GREML heritability estimates using the LMM implementation in the singher R package (Hernandez *et al.* 2019), available at https://github.com/andywdahl/SingHer.

Discussion

Negative heritability estimates are common results of standard statistical procedures. Linear random-effects models of genetic causality make negative heritability impossible, inviting us to treat the negative parameter estimates as spurious results produced by statistical noise that should be truncated back to zero, the closest meaningful value. However, these generative linear models are not physically validated: it is not in any sense literally true that phenotypes are produced by additive contributions of alleles and independent noise. We have shown here that other biologically plausible stochastic models would indeed generate data in the negative range of heritability parameters. These are misspecified from the point of view of the linear random-effects models, but they are correctly specified from the point of view of the Gaussian likelihood that is used to estimate the heritability parameter. Our phenotypic repulsion example demonstrates that truly negative heritability can convey a biological fact when individuals tend to differentiate themselves from their relatives. Meaningfully negative heritability should not always be ruled out.

There has long been some dispute about whether these "spurious" negative estimates ought to be included for the sake of unbiasedness, so that the whole ensemble of estimates from multiple studies might be properly centered.

We use an approximation for the GREML heritability estimate that we previously derived (Steinsaltz *et al.* 2018) to formally support this argument as well as to quantify the bias from truncation.

More importantly, we also suggest that the problem should be considered with more nuance: The very possibility of negative heritability depends strongly on the nature of the trait, of the population, and of the sampling procedure. True, asymptotically persistent negative heritability requires strong nonlinear contributions, increasingly strong as the negative parameter approaches the true negative lower bound. This suggests that it may be reasonable to replace truncation at zero by an appropriate shrinkage of negative estimates toward zero, perhaps based on context. This would affect not only negative point estimates, but also confidence intervals centered at small positive values. In a Bayesian framework this would correspond, of course, to assigning heritability a prior distribution with small, nonzero weight on negative values. Statistical models of convenience, such as the variance-component model that underlies GREML and many other heritability estimation approaches, should never drive substantive scientific conclusions, such as declaring that negative heritability is impossible.

Acknowledgments

We thank David Siegel for help processing the GEUVADIS data. D.S. is supported by grant ES/N011856/1 from the UK Economic and Social Research Council and by grant BB/S001824/1 from the Biotechnology and Biological Sciences Research Council. A.D. is supported by grant 1U01HG009080-01 from the U.S. National Institutes of Health. K.W.W. is supported by grant 5P30AG012839 from the U.S. National Institute on Aging.

Literature Cited

Alter, O., P. O. Brown, and D. Botstein, 2000 Singular value decomposition for genome-wide expression data processing and modeling. Proc. Natl. Acad. Sci. USA 97: 10101–10106. https://doi.org/10.1073/pnas.97.18.10101

Bell, A. E., 1977 Heritability in retrospect. J. Hered. 68: 297–300. https://doi.org/10.1093/oxfordjournals.jhered.a108840

Bhatia, G., A. Gusev, P.-R. Loh, B. J. Vilhjálmsson, S. Ripke *et al.*, 2015 Haplotypes of common SNPs can explain missing heritability of complex diseases. bioRxiv (Preprint posted July 12, 2015).https://doi.org/doi: 10.1101/022418

Boucheron, S., G. Lugosi, and P. Massart, 2013 *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199535255.001.0001

Bulik-Sullivan, B., 2015 Relationship between LD score and Haseman-Elston regression. bioRxiv (Preprint posted April 20, 2015). https://doi.org/doi: 10.1101/018283

Bulik-Sullivan, B. K., P.-R. Loh, H. K. Finucane, S. Ripke, J. Yang *et al.*, 2015 LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47: 291–295. https://doi.org/10.1038/ng.3211

Cardillo, M., 2012 The phylogenetic signal of species co-occurrence in high-diversity shrublands: different patterns for fire-killed and fire-resistant species. BMC Ecol. 12: 21. https://doi.org/10.1186/1472-6785-12-21

Cavender-Bares, J., D. D. Ackerly, D. A. Baum, and F. A. Bazzaz, 2004 Phylogenetic overdispersion in Floridian oak communities. Am. Nat. 163: 823–843. https://doi.org/10.1086/386375

Chen, G.-B., 2014 Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman–Elston regression. Front. Genet. 5: 107. https://doi.org/10.3389/fgene.2014.00107

Conley, D., E. Rauscher, C. Dawes, P. K. E. Magnusson, and M. L. Siegal, 2013 Heritability and the equal environments assumption: evidence from multiple samples of misclassified twins. Behav. Genet. 43: 415–426. https://doi.org/10.1007/s10519-013-9602-1

Feller, W., 1968 *An Introduction to Probability and Its Applications*. Vol. 1 Ed. 3. John Wiley & Sons, New York.

Finucane, H. K., B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef *et al.*, 2015 Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47: 1228–1235. https://doi.org/10.1038/ng.3404

Gill, J.L., and E. Jensen, 1968 Probability of obtaining negative estimates of heritability. Biometrics 24: 517–526. https://doi.org/10.2307/2528315

Golan, D., E. S. Lander, and S. Rosset, 2014 Measuring missing heritability: inferring the contribution of common variants. Proc. Natl. Acad. Sci. USA 111: E5272–E5281. https://doi.org/10.1073/pnas.1419064111

Gusev, A., A. Ko, H. Shi, G. Bhatia, W. Chung *et al.*, 2016 Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. 48: 245–252. https://doi.org/10.1038/ng.3506

Gymrek, M., T. Willems, A. Guilmatre, H. Zeng, B. Markus *et al.*, 2016 Abundant contribution of short tandem repeats to gene expression variation in humans. Nat. Genet. 48: 22–29. https://doi.org/10.1038/ng.3461

Haldane, J. B. S., 1996 The negative heritability of neonatal jaundice. Ann. Hum. Genet. 60: 3–5. https://doi.org/10.1111/j.1469-1809.1996.tb01165.x

Hernandez, R. D., L. H. Uricchio, K. Hartman, C. Ye, A. Dahl *et al.*, 2019 Ultrarare variants drive substantial cis heritability of human gene expression. Nat. Genet. 51: 1349–1355. https://doi.org/10.1038/s41588-019-0487-7

Jacquard, A., 1983 Heritability: one word, three concepts. Biometrics 39: 465–477. https://doi.org/10.2307/2531017

Jiang, J., C. Li, D. Paul, C. Yang, and H. Zhao *et al.*, 2016 On high-dimensional misspecified mixed model analysis in genome-wide association study. Ann. Stat. 44: 2127–2160. https://doi.org/10.1214/15-AOS1421

Johnston, S. E., D. Beraldi, A. F. McRae, J. M. Pemberton, and J. Slate, 2010 Horn type and horn length genes map to the same chromosomal region in Soay sheep. Heredity 104: 196–205. https://doi.org/10.1038/hdy.2009.109

Krishna Kumar, S., M. W. Feldman, D. H. Rehkopf, and S. Tuljapurkar, 2016 Limitations of GCTA as a solution to the missing heritability problem. Proc. Natl. Acad. Sci. USA 113: E61–E70 (erratum: Proc. Natl. Acad. Sci. USA 113: E813). https://doi.org/10.1073/pnas.1520109113

Lappalainen, T., M. Sammeth, M. R. Friedländer, P. A. 't Hoen, J. Monlong, 2013 Transcriptome and genome sequencing uncovers functional variation in humans. Nature 501: 506–511. https://doi.org/10.1038/nature12531

Leek, J. T., and J. D. Storey, 2007 Capturing heterogeneity in gene expression studies by Surrogate Variable Analysis. PLoS Genet. 3: e161. https://doi.org/10.1371/journal.pgen.0030161

Li, B., and C. N. Dewey, 2011 RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323. https://doi.org/10.1186/1471-2105-12-323

Liggett, T., 1999 *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes*. Springer Verlag, New York. https://doi.org/10.1007/978-3-662-03990-8

Lush, J. L., 1940 Intra-sire correlations or regressions of offspring on dam as a method of estimating heritability of characteristics. Proceedings of the American Society of Animal Nutrition 1940: 293–301.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff *et al.*, 2009 Finding the missing heritability of complex diseases. Nature 461: 747–753. https://doi.org/10.1038/nature08494

Nelder, J., 1954 The interpretation of negative components of variance. Biometrika 41: 544–548. https://doi.org/10.1093/biomet/41.3-4.544

Pollard, D., 1990 *Empirical Processes: Theory and Applications, Volume 2 of CBMS-NSF Regional Conference Series in Probability and Statistics*. Institute of Mathematical, Hayward, CA.

Stegle, O., L. Parts, R. Durbin, and J. Winn, 2010 A Bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eQTL studies. PLoS Comput. Biol. 6: e1000770. https://doi.org/10.1371/journal.pcbi.1000770

Stegle, O., L. Parts, M. Piipari, J. Winn, and R. Durbin, 2012 Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. 7: 500–507. https://doi.org/10.1038/nprot.2011.457

Steinsaltz, D., A. Dahl, and K. W. Wachter, 2018 Statistical properties of simple random-effects models for genetic heritability. Electron. J. Stat. 12: 321–358. https://doi.org/10.1214/17-EJS1386

Sulloway, F. J., 2011 Why siblings are like darwin's finches: birth order, sibling competition, and adaptive divergence within the family, pp. 87–119 in *The Evolution of Personality and Individual Differences*, edited by D. M. Buss and P. H. Hawley. Oxford University Press, New York.

Visscher, P. M., S. E. Medland, M. A. R. Ferreira, K. I. Morley, G. Zhu *et al.*, 2006 Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genet. 2: e41. https://doi.org/10.1371/journal.pgen.0020041

Webb, C. O., D. D. Ackerly, M. A. McPeek, and M. J. Donoghue, 2002 Phylogenies and community ecology. Annu. Rev. Ecol. Syst. 33: 475–505. https://doi.org/10.1146/annurev.ecolsys.33.010802.150448

White, H., 1982 Maximum likelihood estimation of misspecified models. Econometrica 50: 1–25.

Wright, F. A., P. F. Sullivan, A. I. Brooks, F. Zou, W. Sun *et al.*, 2014 Heritability and genomics of gene expression in peripheral blood. Nat. Genet. 46: 430–437. https://doi.org/10.1038/ng.2951

Wu, Y., and S. Sankararaman, 2018 A scalable estimator of SNP heritability for biobank-scale data. Bioinformatics 34: i187–i194. https://doi.org/10.1093/bioinformatics/bty253

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42: 565–569. https://doi.org/10.1038/ng.608

Yang, J., T. Lee, J. Kim, M. C. Cho, B. G. Han *et al.*, 2013 Ubiquitous polygenicity of human complex traits: genome-wide analysis of 49 traits in Koreans. PLoS Genet. 9: e1003355. https://doi.org/10.1371/journal.pgen.1003355

Yuan, K.-H., 1997 A theorem on uniform convergence of stochastic functions with applications. J. Multivariate Anal. 62: 100–109. https://doi.org/10.1006/jmva.1997.1674

Zhou, X., 2017 A unified framework for variance component estimation with summary statistics in genome-wide association studies. Ann. Appl. Stat. 11: 2027–2051. https://doi.org/10.1214/17-AOAS1052

Zhu, Z., A. Bakshi, A. A. Vinkhuyzen, G. Hemani, S. H. Lee *et al.*, 2015 Dominance genetic variation contributes little to the missing heritability for human complex traits. Am. J. Hum. Genet. 96: 377–385. https://doi.org/10.1016/j.ajhg.2015.01.001

Communicating editor: S. Browning

Appendix A: Proof of Proposition 1

We wish to show that $\limsup_{n\to\infty}\hat{\psi}_n \le -\delta$. This will follow if every increasing sequence n_i has a subsequence n_{i_j} such that $\limsup_{j\to\infty}\hat{\psi}_{n_{i_j}} \le -\delta$. Define the empirical measure $\sigma_n := \sum_{i=1}^{n}\delta_{s_{n,i}}$, where δ_x represents unit point mass at x. Since the space of probability measures on $[0, \sup s_{n,i}]$ is compact, given an increasing sequence n_i we may find a subsequence n_{i_j} such that $\hat{\sigma}_{n_{i_j}}$ converges weakly to a measure σ on $[0, s_{\max}]$. Thus, it will suffice to prove the proposition under the assumption that $\sigma_n \xrightarrow[n\to\infty]{} {}_w\sigma$.

We follow the general principle, enunciated by White (1982), that the MLE for the misspecified model will converge to the closest fit in the Kullback–Leibler sense. In other words, the parameter estimate converges in probability to the location of the maximum expected value of the log-likelihood function. The arguments of White (1982) do not apply directly here, because we are not sampling identically and independently (i.i.d.) random variables; however, by Equation (22) of Steinsaltz *et al.* (2018), the score function may be written

$$\frac{1}{2\bar{v}(\psi)}\cdot G_n(\psi;\mathbf{x}) := \frac{1}{2n\bar{v}(\psi)}\sum_{i=1}^{n} a_{n,i}(\psi)X_i, \tag{A1}$$

for $-1/(s_{\max}^2-1) < \psi \le 1$, where (X_i) are i.i.d. χ_1^2 random variables and

$$\begin{aligned} a_{n,i}(\psi) &:= a\left(s_{n,i}^2,\psi\right) \\ &:= \frac{f\left(s_{n,i}^2\right)}{(1-\psi)\left(1-\psi+\psi s_{n,i}^2\right)}\left(\frac{s_{n,i}^2}{1-\psi+\psi s_{n,i}^2} - \frac{1}{n}\sum_{j=1}^{n}\frac{s_{n,j}^2}{1-\psi+\psi s_{n,j}^2}\right) \\ &= \frac{f\left(s_{n,i}^2\right)}{1-\psi+\psi s_{n,i}^2}\left(\frac{s_{n,i}^2-1}{1-\psi+\psi s_{n,i}^2} - n^{-1}\sum_{j=1}^{n}\frac{s_{n,j}^2-1}{1-\psi+\psi s_{n,j}^2}\right). \end{aligned}$$

Since $\left(\max\{{s_{n',i}}^2 : 1\le i\le n',\ n'\ge n\}-1\right)^{-1} > \delta$ for n sufficiently large, we may assume without loss of generality that this holds for all n, perhaps after truncating an initial portion of the sequence. It follows that $a_{n,i}(\psi)$ is defined for any $\psi\in[-\delta,1]$, and by (9) that

$$n^{-1}\sum_{i=1}^{n}\left|a_{n,i}(\psi)\right| \quad\text{and}\quad n^{-1}\sum_{i=1}^{n}\left|a_{n,i}^{'}(\psi)\right|$$

are both uniformly bounded over all n, and $\psi\in[-\delta,1]$. By a variant of the central result of Yuan (1997), $G_n(\psi;\mathbf{x})$ converges uniformly in ψ to the function that is the limit of the expected values

$$G(\psi) = \lim_{n\to\infty} G_n(\psi;1) = \mathrm{Cov}_\sigma\left(\frac{f(S^2)}{1-\psi+\psi S^2}, \frac{S^2-1}{1-\psi+\psi S^2}\right).$$

The covariance is understood here to be with respect to S having distribution σ. [This result does not satisfy exactly the conditions of Yuan (1997), so we provide a proof of the result, stated as *Lemma 1*.]

We need to show that

$$G(\psi) < 0 \ \text{ for } \ \psi \ge -\delta. \tag{A2}$$

From this it will follow that $\inf_{\psi\in[-\delta,1]} G(\psi) < 0$, hence

$$\begin{aligned} &\lim_{n\to\infty}\sup\mathbb{P}\{\hat{\psi}_n \ge -\delta\} \le \\ &\lim_{n\to\infty}\sup\mathbb{P}\left\{\sup_{\psi\in[-\delta,1]}|G_n(\psi)-G(\psi)| \ge \inf_{\psi\in[-\delta,1]}|G(\psi)|\right\} = 0. \end{aligned}$$

It remains to verify (A2). We note that the definition of δ makes

$$\frac{C_1C_2}{\delta} \geq 3f(0)C_3\big(1+\psi\big(S^2-1\big)\big)^{-3}$$

for any $\psi \in [-\delta, 0]$ and $S \in [0, s_{\max}]$. Since $f(t) + C_1 t$ is a decreasing function of t, for $t \in [0, s_{\max}^2]$, we have by Harris's inequality (Boucheron *et al.* 2013, Theorem 2.15)

$$\begin{aligned} G(0) &= \mathrm{Cov}_\sigma\big(f\big(S^2\big) + C_1S^2, S^2\big) - C_1\mathrm{Var}_\sigma\big(S^2\big) \\ &\leq -C_1C_2 \\ &< 0. \end{aligned} \tag{A3}$$

We also have

$$\begin{aligned} (1-\psi)G'(\psi) = &-\mathrm{Cov}_\sigma\Big(\frac{f\big(S^2\big)}{1-\psi+\psi S^2}, \frac{S^2\big(S^2-1\big)}{\big(1-\psi+\psi S^2\big)^2}\Big) \\ &-\mathrm{Cov}_\sigma\Big(\frac{\big(S^2-1\big)f\big(S^2\big)}{\big(1-\psi+\psi S^2\big)^2}, \frac{S^2}{1-\psi+\psi S^2}\Big) \\ &+(1-\psi)^{-1}\mathrm{Cov}_\sigma\Big(\frac{f\big(S^2\big)}{1-\psi+\psi S^2}, \frac{S^2}{1-\psi+\psi S^2}\Big). \end{aligned}$$

For $\psi \in [-\delta, 0]$ and $0 \leq S \leq s_{\max}$ we have

$$\big(1-\psi+\psi S^2\big)^{-3} \leq \big(1-\delta\big(s_{\max}^2-1\big)\big)^{-3}.$$

Since f is decreasing, we have for $-\delta \leq \psi \leq 0$ the bound

$$\begin{aligned} |G'(\psi)| &\leq 3f(0)C_3\big(1-\delta\big(s_{\max}^2-1\big)\big)^{-3} \\ &\leq \frac{C_1C_2}{\delta}. \end{aligned}$$

This proves that $G(\psi) < 0$ for $-\delta \leq \psi \leq 0$.

For $\psi \in [0, 1]$ $f(t)/(1-\psi+\psi t)$ is a decreasing function of t, and $t/(1-\psi+\psi t)$ is increasing, so (again by Harris's Inequality) $G(\psi) < 0$, which completes the proof.

We now prove the key uniform convergence result for G_n. The range of s and of ψ in this result may be rescaled arbitrarily, so for convenience of notation we will denote these by [0,1].

Lemma 1.
Let $a : [0,1]^2 \to \mathbb{R}$ be a continuous function such that for all $s \in (0,1]$

$$K_s := \sup_{\psi\in[0,1]}\Big|a(s,\psi)\Big| \text{ and } L_s := \sup_{\psi\neq\psi'\in[0,1]}\frac{\Big|a(s,\psi')-a(s,\psi)\Big|}{\Big|\psi-\psi'\Big|}$$

are both finite. Let $\sigma_n = n^{-1}\sum_{i=1}^n \delta_{s_{n,i}}$ be atomic probability measures on $(0,1]$ concentrated at n points $0 < s_{n,1} \leq \cdots \leq s_{n,n} \leq 1$. We suppose that the measures σ_n converge weakly to a probability measure $\sigma = \sigma_\infty$ on $(0, s_*)$, and that there is a $\delta \in (0,1]$ such that

$$K_*^2 := \sup_n \int \Big(1 \vee K_S\Big)^2 \mathrm{d}\sigma_n(S) < \infty \text{ and} \tag{A4}$$

$$L_*^2 := \sup_n \int \Big(1 \vee L_S\Big)^2 \mathrm{d}\sigma_n(S) < \infty. \tag{A5}$$

Let $\{X_{n,i} : 1 \le i \le n,\ n \in \mathbb{N}\}$ *be independent random variables with* $\mathbb{E}[X_{n,i}] = 1$ *and* $V := \sup \mathrm{Var}(X_{n,i}) < \infty$. *Define for* $\psi \in (0,1]$

$$G_n(\psi) := n^{-1} \sum_{i=1}^{n} X_{n,i} a\big(s_{n,i}, \psi\big).$$

Then G_n *converges uniformly in probability to the function* $G : \to \mathbb{R}$ *defined by*

$$G(\psi) := \int a(s,\psi) \mathrm{d}\sigma(s).$$

That is,

$$\sup\{|G(\psi) - G_n(\psi)| : \psi \in [0,1]\} \xrightarrow{n \to \infty}_{P} 0.$$

The condition (A4) may be weakened by replacing $(1 + K_S)^2$ by $(1 + K_S)^{1+\delta}$, for δ positive, and equivalently for L_S, as long as we have correspondingly stronger moment bounds on $X_{n,i}$. We have stated it in this form for simplicity.

Proof. The sublinearity of the Lipschitz constant implies that the Lipschitz constant of G_n is a random variable bounded by

$$L_{(n)} := n^{-1} \sum_{i=1}^{n} X_{n,i} L_{s_{n,i}}.$$

We have

$$\mathbb{E}\Big[L_{(n)}\Big] = \int L_S \mathrm{d}\sigma_n(S) \le L_*$$

by the Cauchy–Schwarz Inequality. Also by the Cauchy–Schwarz Inequality, we have

$$\mathrm{Var}\Big(L_{(n)}\Big) = n^{-2} \sum_{i=1}^{n} L_{s_{n,i}}^2 \mathrm{Var}\big(X_{n,i}\big)$$

$$\le \frac{V L_*^2}{n}.$$

Thus $\mathbb{P}\{\mathrm{Lip}(G_n) \le 2L_*\} \overset{n \to \infty}{\to} 1$.

We have

$$\sup_{\psi \in [0,1]} \Big| G(\psi) - G_n(\psi) \Big| \le \sup_{\psi \in [0,1]} \left| n^{-1} \sum_{i=1}^{n} \big(X_{n,i} - 1\big) a\big(s_{n,i}, \psi\big) \right|$$

$$+ \sup_{\psi \in [0,1]} \left| \int a(s,\psi) \mathrm{d}\sigma_n(s) - \int a(s,\psi) \mathrm{d}\sigma(s) \right|. \tag{A6}$$

Fix any positive integer k. Because of the bounds on the Lipschitz constants of G and G_n, the second term is bounded by

$$\frac{3L_*}{k} + \max_{1 \le j \le k} \left| \int a(s, j/k) \mathrm{d}\sigma_n(s) - \int a(s, j/k) \mathrm{d}\sigma_n(s) \right|. \tag{A7}$$

Because of the assumed weak convergence $\sigma_n \to \sigma$, this converges to $3L_*/k$ as $n \to \infty$ for each fixed k. Since k is arbitrary, the second term in fact converges to 0 as $n \to \infty$.

To deal with the first term we use the standard method of *chaining* (*cf*. Pollard 1990, chapter 3): we define finite skeletons of [0,1], subsets $D_0 \subset D_1 \subset \cdots$ with $|D_j| = 2^j$, defined by

$$D_j := \left\{ \frac{2\ell+1}{2^{j+1}} : \ell = 0, \cdots, 2^j - 1 \right\}.$$

We then proceed by approximating any point $\psi \in [0,1]$ by a sequence of nearest neighbors $\psi_j \in D_j$, so that $|\psi_j - \psi_{j-1}| = 2^{-j-1}$. Since for any continuous function f

$$f(\psi) = f(0) + \sum_{j=1}^{\infty} \Big(f(\psi_j) - f(\psi_{j-1}) \Big),$$

we have the basic chaining inequality

$$\sup_{\psi \in [0,1]} \left| n^{-1} \sum_{i=1}^{n} (X_{n,i} - 1) a(s_{n,i}, \psi) \right| \le n^{-1/2} \left(R_0 + \sum_{j=1}^{\infty} R_j \right), \tag{A8}$$

where

$$R_0 := n^{-1/2} \left| \sum_{i=1}^{n} (X_{n,i} - 1) a(s_{n,i}, 1/2) \right|,$$

$$R_j := \max_{\psi_j \in D_j} n^{-1/2} \left| \sum_{i=1}^{n} (X_{n,i} - 1) a(s_{n,i}, \psi_j) - \right.$$

$$\left. \sum_{i=1}^{n} (X_{n,i} - 1) a(s_{n,i}, \psi_{j-1}) \right|.$$

We have

$$\mathbb{E}[R_0^2] \le n^{-1} V \sum_{i=1}^{n} a\left(s_{n,i}^2, 0\right)^2 \le K_*^2 V.$$

Now note that

$$\mathbb{E}\left[\left| n^{-1/2} \sum_{i=1}^{n} (X_{n,i} - 1) \left(a(s_{n,i}, \psi_j) - a(s_{n,i}, \psi_{j-1}) \right) \right|^2 \right]$$

$$\le 2^{-2j-2} \mathbb{E}\left[n^{-1} \left| \sum_{i=1}^{n} (X_{n,i} - 1) L_{s_{n,i}} \right|^2 \right]$$

$$\le 2^{-2j-2} V L_*^2.$$

For any collection of random variables $\xi_1, \ldots, \xi_m$ we know that

$$\mathbb{E}\left[\max \xi_k^2 \right] \le m \max \mathbb{E}\left[\xi_k^2 \right].$$

so for $j \ge 1$

$$\mathbb{E}[R_j^2] \le 2^j \cdot 2^{-2j-2} V L_*^2 = 2^{-j-2} V L_*^2$$

By Minkowski's Inequality, we have

$$\mathbb{E}\left[(R_0 + \cdots + R_k)^2 \right] \le \left(\sum_{j=0}^{k} \mathbb{E}[R_j^2]^{1/2} \right)^2 \le (K_* + 2L_*)^2 V.$$

So finally, by (A8), we have

$$\mathbb{E}\left[\sup_{\psi\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}(X_{n,i}-1)a(s_{n,i},\psi)\right|^2\right] \leq \frac{(K_*+2L_*)^2V}{n}. \tag{A9}$$

Applying Markov's inequality, and combining this with (A7), completes the proof.

Appendix B: HE Regression Under Phenotypic Repulsion

According to Equation (8) of Wu and Sankararaman (2018) the HE regression estimate of genetic variance may be defined by

$$\hat{\sigma}_g^2 = \frac{y^*(A-I_n)y/n}{\mathrm{tr}(A^2)/n-1}.$$

where tr $(\cdot)$ is the trace of a matrix. As the heritability is the ratio of $\hat{\sigma}_g^2$ to a positive estimate of phenotypic variance, the estimated heritability will be negative whenever $\hat{\sigma}_g^2$ is negative.

Under our model, $y = UTx$ is the vector of phenotypes, with U an orthogonal matrix, x a vector of i.i.d. standard normal random variables, and T the diagonal matrix with $\sqrt{t_i}$ on the diagonal, with

$$t_i = \frac{1+\phi_0 s_i^2}{1-\alpha+\alpha s_i^2(1-\phi_0)+\alpha\phi_0 s_i^4}.$$

These t_i are the same as $f(s_i^2)$, where f is given in Equation 7.

The denominator in the HE estimator of $\hat{\sigma}_g^2$ is $n^{-1}\sum_{i=1}^n(s_i^4-1)$, which converges to the constant we have called C_3-1. The numerator is

$$n^{-1}x^*T(S^2-I_n)Tx = n^{-1}\sum_{i=1}^{n}t_i(s_i^2-1)x_i^2,$$

where S is the diagonal matrix with s_i on the diagonal. By the same argument as in the proof of *Proposition 1*, where we apply Harris's inequality to show that $\mathrm{Cov}(t_i, s_i^2) < 0$ (Equation A3), we see that $\lim\ \sup_{n\to\infty} n^{-1}\sum_{i=1}^n t_i(s_i^2-1) < -C_1C_2$. The Weak Law of Large Numbers implies that the numerator remains bounded above (in probability) as $n\to\infty$ by $-C_1C_2$. Hence the HE regression estimate targets a negative number smaller than $-C_1C_2/C_3$ for large n.

Appendix C: SE Under Repulsion

Asymptotically correct SE for the standard genotypic random-effects model may be calculated from the Fisher information. We carry out the calculations here for the parametrization in terms of θ and $\phi = \psi/(1-\psi)$, to simplify the notation. In the region of ψ negative or close to 0 the variance and covariance of ψ hardly differs from that of φ, and in any case the transformation is straightforward.

Starting from the transformed phenotypes $\mathbf{z} = U^*\mathbf{y}$, and using the definitions of v_i and w_i provided above Equation 4, we have the log likelihood

$$\ell(\theta,\phi) = \frac{n}{2}\log\theta - \frac{1}{2}\sum\log(1+\phi s_i^2) - \frac{\theta}{2}\sum v_i(\phi).$$

The first two derivatives may be written as

$$D\ell(\theta,\phi) = \frac{n}{2}\begin{pmatrix} \frac{1}{\theta}-\langle v\rangle \\ -\langle s^2 w\rangle + \theta\langle s^2 vw\rangle \end{pmatrix}$$

and

$$D^2\ell(\theta,\phi) = -\frac{n}{2}\begin{pmatrix} \frac{1}{\theta^2} & \langle s^2 vw\rangle \\ \langle s^2 vw\rangle & -\langle s^2 w^2\rangle + 2\theta\langle s^4 vw^2\rangle \end{pmatrix},$$

where $\langle a\rangle$ is used to denote the mean of a sequence (a_i). Thus $\langle s^4 vw^2\rangle = n^{-1}\sum_{i=1}^n s_i^4 v_i w_i^2$.

We immediately have $\hat{\theta} = 1/\langle v \rangle$. When samples are drawn from the true model with parameters (θ_0, ϕ_0) we have that $\theta_0 v_i(\phi_0)$ are independent chi-squared random variables with 1 degree of freedom. Thus, the expected Fisher information is

$$\mathcal{I}(\theta_0, \phi_0) = \frac{n}{2} \begin{pmatrix} 1/\theta_0^2 & (1 - \langle w \rangle)/(\phi_0 \theta_0) \\ (1 - \langle w \rangle)/(\phi_0 \theta_0) & \left\langle (1-w)^2 \right\rangle / \phi_0^2 \end{pmatrix}.$$

The covariance matrix for $(\hat{\theta}, \hat{\phi})$ is the inverse

$$\mathcal{I}(\theta_0, \phi_0)^{-1} = \frac{2}{n \mathrm{Var}(w)} \begin{pmatrix} \theta_0^2 \left\langle (w-1)^2 \right\rangle & \phi_0 \theta_0 (\langle w \rangle - 1) \\ \phi_0 \theta_0 (\langle w \rangle - 1) & \phi_0^2 \end{pmatrix}.$$

In particular, the asymptotic variance of $\hat{\phi}$ is $2\phi_0^2/(n\ \mathrm{Var}(w))$. It follows immediately that the asymptotic variance of $\hat{\psi}$ is $2\psi_0^2(1-\psi_0)^2/(n\ \mathrm{Var}(w)) = 2(1-\psi_0)^2/(n\tau_2)$.

For the (misspecified) phenotypic repulsion model three changes are needed: First, the expected value of $v_i(\phi)$ is no longer $(1 + \phi_0 s_i^2)/((1 + \phi s_i^2)\theta_0)$, but

$$\beta_i(\phi) := \theta_0^{-1} \frac{1 + \phi_0 s_i^2}{1 + \phi s_i^2} \frac{1}{1 - \alpha + \alpha s_i^2 (1 - \phi_0) + \alpha \phi_0 s_i^4}.$$

Second, the Fisher information is evaluated not at the parameters (θ_0, ϕ_0),which no longer define the distribution from which the data are sampled, but rather at the best-fit parameters (θ_*, ϕ_*). We still have $\theta_* = 1/\langle v(\phi_*) \rangle$, but there is no simple representation for ϕ_*, which will solve the equation $\mathrm{Cov}(w(\phi_*), \beta(\phi_*)) = 0$. The expected Fisher information is

$$\mathcal{I}(\theta_*, \phi_*) = \frac{n}{2} \begin{pmatrix} \langle \beta \rangle^2 & \dfrac{\langle \beta \rangle - \langle \beta w \rangle}{\phi_*} \\ \dfrac{\langle \beta \rangle - \langle \beta w \rangle}{\phi_*} & \dfrac{2\left\langle \beta (1-w)^2 \right\rangle}{\langle \beta \rangle \phi_*^2} - \dfrac{\left\langle (1-w)^2 \right\rangle}{\phi_*^2} \end{pmatrix}.$$

Here and below β and w are evaluated at ϕ_*.) The inverse is

$$\mathcal{I}\left(\theta_*, \phi_*\right)^{-1} = \frac{2}{n} \Big(2\langle \beta \rangle \cdot \left\langle \beta (1-w)^2 \right\rangle - \langle \beta \rangle^2 \left\langle (1-w)^2 \right\rangle - (\langle \beta \rangle - \langle \beta w \rangle)^2 \Big)^{-1}$$

$$\times \begin{pmatrix} \dfrac{2\left\langle \beta (1-w)^2 \right\rangle}{\langle \beta \rangle} - \left\langle (1-w)^2 \right\rangle & \phi_* (\langle \beta w \rangle - \langle \beta \rangle) \\ \phi_* (\langle \beta w \rangle - \langle \beta \rangle) & \phi_*^2 \langle \beta \rangle^2 \end{pmatrix}.$$

The third change is that the asymptotic variance for a misspecified model is not given by $\mathcal{I}^{-1}$, but by the sandwich estimator White (1982) $\mathcal{I}^{-1} V \mathcal{I}^{-1}$, where V is the covariance matrix of $D\ell$, which is

$$V = \begin{pmatrix} 2\langle \beta^2 \rangle & \dfrac{\langle \beta^2 (1-w) \rangle}{\phi_* \langle \beta \rangle} \\ \dfrac{\langle \beta^2 (1-w) \rangle}{\phi_* \langle \beta \rangle} & \dfrac{\left\langle \beta^2 (1-w)^2 \right\rangle}{\phi_*^2 \langle \beta \rangle^2} \end{pmatrix}.$$