The accuracy of LD Score regression as an estimator of confounding and genetic correlations in genome-wide association studies

James J. Lee[1*]
Matt McGue[1]
William G. Iacono[1]
Carson C. Chow[2*]

[1]Department of Psychology
University of Minnesota Twin Cities
75 East River Parkway
Minneapolis, MN 55455, USA
(612) 625-4980

[2]Mathematical Biology Section
Laboratory of Biological Modeling, NIDDK
National Institutes of Health
10 Center Drive
Bethesda, MD 20892, USA
(301) 402-8250

[*]To whom correspondence should be addressed;
E-mail: leex2293@umn.edu, carsonc@niddk.nih.gov.

ARTICLE

RUNNING HEAD: LD Score regression

The authors declare no conflict of interest.

Abstract

In order to infer that a single-nucleotide polymorphism (SNP) either affects a phenotype or is linkage disequilibrium with a causal site, we must have some assurance that any SNP-phenotype correlation is not the result of confounding with some environmental variable that also affects the trait. In this work we study the properties of LD Score regression, a recently developed method for using summary statistics from genome-wide association studies (GWAS) to ensure that confounding does not inflate the number of false positives. We do not treat the effects of genetic variation as a random variable and thus are able to obtain results about the unbiasedness of this method. We demonstrate that LD Score regression can produce estimates of confounding at null SNPs that are nearly unbiased under fairly general conditions. This robustness holds in the case of the parent genotype affecting the offspring phenotype through some environmental mechanism, despite the resulting correlation over SNPs between LD Scores and the degree of confounding. LD Score regression is thus an even stronger technique for causal inference than foreseen by its developers. Additionally, we demonstrate that LD Score regression can produce reasonably robust estimates of the genetic correlation, even when its estimates of the genetic covariance and the two univariate heritabilities are substantially biased.

Key Words: causal inference; heritability; population stratification; quantitative genetics

1 Introduction

The goal of genome-wide association studies (GWAS) is to find regions in the genome where variation affects a phenotype. However, this must be accomplished from observed correlations, and inferring causation from correlation is a famously perilous endeavor (Freedman, 1999; Pearl, 2009). GWAS has been fortunate in that it offers a variety of methods to check whether confounding effects have produced spurious correlations between genetic and phenotypic variation. These methods have led to a strong consensus that confounding has a minimal impact on GWAS results (Goldstein, 2011; Visscher, Brown, McCarthy, & Yang, 2012; Lee, 2012; Lee, Vattikuti, & Chow, 2016).

One of the newer methods used to check the causal status of GWAS associations is known as LD Score regression (Bulik-Sullivan et al., 2015a), which can be applied to summary statistics assembled from the contributions of different research groups and thus does not require access to individual-level data. This technique relies on the simple linear regression of assayed single-nucleotide polymorphism (SNP) j's association chi-square statistic on

$$l_j = \sum_k \Gamma_{jk}^2, \tag{1}$$

the sum over all SNPs of each SNP's squared correlation with the focal SNP j. This latter quantity is called SNP j's "LD Score." Empirically, the regression curve relating chi-square statistics to LD Scores is always very close to an upwardly sloping straight line. This result is explicable because a SNP tagging more of its neighbors—and, thus, having a higher LD Score—is more likely to tag one or more causal sites affecting the phenotype. The lowest possible LD Score of a SNP is one, which is obtained when a SNP is in perfect linkage equilibrium (LE) with all other SNPs. A hypothetical SNP with an LD Score of zero fails to tag the causal effect of any SNP in the genome—including whatever effect the

SNP itself may have. Therefore, if the intercept of LD Score regression departs upward from unity (the theoretical expectation of the chi-square distribution with one degree of freedom), then intuitively the departure must be due to confounding, poor quality control, overlapping samples in the meta-analysis, or other artifacts. This simple and insightful method of estimating the distribution of truly null SNPs (or at least a certain subset of such SNPs) should in most cases lead to a much better global correction of the association statistics than the overly conservative genomic control (Devlin & Roeder, 1999).

The slope obtained from LD Score regression could in principle also provide an estimate of the trait's heritability—the fraction of the phenotypic variance ascribable to genetic differences in the population. The developers do not recommend this particular use of the method, and we will explain why LD Score regression is not a reliable estimator of heritability below.

Another use of LD Score regression is the estimation of genetic correlations (Bulik-Sullivan et al., 2015b). The dependent variable in this case is not the chi-square statistic from the GWAS of a single trait but rather the product of two Z statistics, each taken from a GWAS of a distinct trait. In principle, this use offers a means of determining whether a trait-trait correlation (as opposed to a SNP-trait correlation) is attributable to the presence of confounders affecting both traits. If the genetic correlation is statistically and quantitatively significant, then we can be sure that the total phenotypic correlation is not attributable solely to confounders that are entirely environmental in nature. Many interesting relationships have been confirmed or discovered by bivariate LD Score regression, including a high genetic correlation ($\sim$0.70) between years of education and age at first childbirth (Barban et al., 2016) and a moderate one ($\sim$0.35) between years of education and intracranial volume (Okbay et al., 2016).

In the classical era of quantitative genetics, genetic correlations were most commonly

estimated with twin data. Rather large samples of twinships are required for precise estimates with this design, and in some cases the estimates are not as robust against modeling assumptions as estimates of univariate heritabilities (Beauchamp, Cesarini, Johannesson, Lindqvist, & Apicella, 2011). For these reasons a welcome development in quantitative genetics has been the advent of GWAS, which can now reach sample sizes in the hundreds of thousands. The appearance of robustness offered by GWAS can be illusory, however, if estimates of genetic correlations are themselves subject to confounding. One can devise estimators of the genetic correlation that might be biased by environmental confounders that affect both phenotypes and happen to be correlated with genetic variation (Palla & Dudbridge, 2015; Okbay et al., 2016). An attractive feature of LD Score regression in this respect is that its control of confounding extends not just to the evidence of association at individual SNPs but also to its genome-wide estimates of genetic correlations. This is important because, again, it is precisely the issue of a phenotypic correlation's underlying causal nature that can call for an accurate estimate of the genetic correlation.

As appealing as the intuition behind LD Score regression may be, the mathematical justifications of this method given so far in the literature raise questions because of their assumption that the effects of genetic variants can be treated as a random variable. This assumption is a useful convenience for computations, but it is not biological. The effects of genetic polymorphisms should be invariant; it is genotypes and phenotypic residuals that vary between individuals (Lee & Chow, 2014; de los Campos, Sorensen, & Gianola, 2015). The assumption also precludes a quantitative treatment of the method's accuracy. Here we refrain from this assumption of random genetic effects and instead treat the effects as a vector of arbitrary fixed constants. Hence we are able to obtain precise expressions of the quantities estimated by LD Score regression, which can be compared with the quantities of actual interest to determine when they coincide. Here is a preview of our results:

1. If the effects of the standardized genotypes at SNP j and its correlated neighbors is not related to SNP j's LD Score, then the slope of LD Score regression provides an unbiased estimate of heritability. For both biological and evolutionary reasons, however, genetic effects are typically smaller near SNPs with higher LD Scores (Gazal et al., 2017). LD Score regression is therefore not a reliable way to estimate the heritability of a trait (or, by extension, the genetic covariance between two traits).

2. The intercept of LD Score regression reflects a useful measure of confounding in the GWAS even in an important case of a relationship between LD Scores and the correlations of SNPs with environmental confounders. This is perhaps the most novel and important conclusion of our analysis. The developers of LD Score regression warn that in the general case of such a relationship the intercept will not accurately estimate the contribution of confounding to the GWAS statistics (Bulik-Sullivan et al., 2015a). The most likely reason for such a relationship, however, is that the genotypes of the parents have an effect on the phenotype of the offspring that is not mediated by the offspring's own genotype. A prominent example of this phenomenon is parents with a genetic disposition to obtain more education creating an environment for their offspring that also promotes educational attainment (Sacerdote, 2007; Kong et al., 2018; Lee et al., in press). In this special but important case, the intercept of LD Score regression can still be used to correct the association statistics of null SNPs so that their average chi-square statistic is in line with the null hypothesis of no causality.

3. LD Score regression provides an accurate estimate of the genetic correlation between two traits, even if neither trait's heritability is well estimated.

2 Materials and methods

In the Supplementary Note, we present a mathematical analysis of LD Score regression's important properties. Importantly we derive our results without treating the average effects of gene substitution as random variables. To confirm our mathematical results, we conducted simulations using the Minnesota Center for Twin and Family Research (MCTFR) genetic data (Miller et al., 2012). The MCTFR cohort consists of 8,405 participants, clustered in families, each typically consisting of a father, mother, and two twin offspring. All cohort members were genotyped at 527,829 SNPs with the Illumina Human660W-Quad array. The other genotypes of the European-ancestry cohort members were subsequently imputed (1000 Genomes phase 1), producing calls at more than 8 million SNPs with a relatively high minor allele frequency (MAF). In the imputation step, data was obtained from only one member of each monozygotic (MZ) twinship, which led to a total sample size of roughly 6,700. There is a large degree of overlap between these imputed SNPs and those used in the calculation of LD Scores by Bulik-Sullivan et al. (2015a).

Our first set of simulations was intended to study the relationship between the LD Scores of causal SNPs and estimates of heritability. To minimize the computational burden of the simulations, we calculated our own MCTFR-specific LD Scores, using the Illumina genotyping data from the ∼4,000 parents. We limited the summation in Equation (1) to SNPs within a 1-cM window of SNP j, as recommended by the developers. We collected these LD Scores into one file and examined the quantiles of their distribution. We called the LD Scores below the 25th percentile (4.744) *very low*, those between the 25th percentile and the median (7.154) *low*, those between the median and the 75th percentile (10.47) *high*, and those above the 75th percentile *very high*. Any given simulation

condition used a sample of 5,000 causal SNPs from either just one of these categories or at random from all ∼500,000 genotyped SNPs, assigning them a normal distribution of average effects (Fisher, 1941; Lee & Chow, 2013) such that the total heritability equaled 0.50. We used PLINK 1.9 (Chang et al., 2015) to carry out a GWAS of the simulated genetic and phenotypic data. We then applied LD Score regression (downloaded September 2016 from https://github.com/bulik/ldsc) to the GWAS statistics to estimate the intercept and the heritability. A hundred replicates were conducted of each condition (*very low*, *low*, *random*, *high*, *very high*), each time keeping the same vector of average effects sampled for that condition but assigning the ∼4,000 subjects different non-genetic residuals. In this set of simulations only the MCTFR white parents were used as subjects.

We retained this simulation framework to study the accuracy with which bivariate LD Score regression estimates genetic correlations. Here we did not sample causal SNPs from just one quartile of LD Score, because in certain conditions this would preclude any nonzero genetic correlation. To simulate a genetic architecture tending to produce unbiased estimates of genetic covariance and heritability, we assigned all genotyped SNPs in MCTFR an average effect drawn from a normal distribution. To simulate a genetic architecture where lower-LD SNPs have larger effects, we multiplied the effects of all SNPs with the *low* annotation by two and then rescaled the vector of effects so that it satisfied the target heritability (0.8). Conversely, to simulate a genetic architecture where higher-LD SNPs have larger effects, we multiplied the effects of all SNPs with the *high* annotation by two and then rescaled. Draws from the bivariate normal distribution were used to induce the desired genetic correlation between the two traits. In scenarios where the two traits were related to LD in opposite ways, the correlation parameter of the bivariate normal distribution was fixed to be higher than the target genetic correlation so that the latter would end up being the correlation between the two vectors of average effects

after the multiplications of effects at disjoint SNPs. In all conditions we fixed the total heritability to an unrealistically high 0.8, because preliminary runs with a heritability of 0.5 sometimes led to the software returning an error rather than taking the square root of a negative heritability estimate. Four thousand subjects is a small sample by the standards of LD Score regression and can sometimes be inadequate for purposes of estimating a genetic correlation.

Our final set of simulations was intended to study whether the intercept continues to be an effective means of controlling the Type 1 error rate with respect to the null hypothesis that the SNP is neither causal nor in LD with a causal site, in a certain case of LD-dependent confounding. Here the MCTFR white offspring were used as the subjects in the simulated GWAS. To compensate for the resulting reduction in sample size, we both increased the number of replicates in each condition to 200 and used the precomputed whole-genome LD Scores available from https://data.broadinstitute.org/alkesgroup/LDSCORE. The latter step ensured a greater number of observations (SNPs) in the regression of chi-square statistics on LD Scores. We made all imputed SNPs on odd chromosomes causal and all imputed SNPs on even chromosomes non-causal; the SNPs on even chromosomes were thus rendered suitable for examining the Type 1 error, since they were all guaranteed to be null. In the conditions intended to simulate confounding that increases with LD Score, we took half the average breeding (additive genetic) value of each offspring's parents and added it to the part of the offspring's non-genetic residual that was independent of breeding value. (This latter part always had a variance of 0.5.) That is, using the same genetic architecture determining the "true polygenic scores" of the offspring, we calculated the true polygenic scores of the two parents in a family and treated the average as an environmental variable affecting the offspring phenotype with a path coefficient of 0.5. A path diagram representing this causal system is presented in Figure 1.

[Figure 1 about here.]

It is desirable to combine this special form of confounding with a more conventional form envisaged by GWAS investigators. An ideal way to simulate population stratification might be to use two cohorts sampled from opposite ends of Europe (or analyzed with different genotyping/imputation pipelines) and to give each cohort a different mean residual. We refrained from using principal components as a proxy for such structure because a subject's projection on a principal component is simply a linear combination of genotypes (Price et al., 2006). If the projection is then used as a basis for how to perturb the phenotype, it becomes very difficult to say how the simulated process is any different from a true causal effect of genotype on phenotype. Because we lack any way of discerning structure within the MCTFR whites independently of principal component analysis, we were forced to simulate bias through another means that happens to be highly convenient in MCTFR—the inclusion of strongly related individuals in the sample. In these simulation conditions, we augmented the sample with 694 individuals, each of whom is a dizygotic (DZ) twin of an original sample member. This raised the offspring sample size to 2,701. (At the outset we chose a twin at random from each DZ twinship to create a sample of unrelated individuals. In the conditions incorporating relatedness, we thus brought back the twin who was initially excluded.) Relatedness induces a spurious inflation of the GWAS chi-square statistics because the effective sample size is not as large as it seems. Note that when relatedness is combined with an effect of parent genotype on offspring phenotype (Figure 1), relatedness additionally becomes a kind of population stratification. Each family is its own population, represented by two members in the sample, and even null SNPs become associated with the phenotype because they are indicative of parentage and thus of a key environmental factor affecting the phenotype.

3 Results

3.1 The slope of univariate LD Score regression as an estimator of heritability

In the Supplementary Note, we show that the slope of LD Score regression provides an unbiased estimate of the heritability if a SNP's LD Score is unrelated to the per-SNP heritability of the SNP itself and its LD partners. The requirement of this null correlation for an unbiased estimate of heritability is stringent. Regressing chi-square statistics on LD Scores to estimate the heritability depends on a constant average per-SNP heritability regardless of LD. If average per-SNP heritability declines in higher-LD regions, say, then the estimated heritability must fall short of the true heritability. This sensitivity to LD is a feature shared with the heritability-estimation method GREML (Speed, Hemani, Johnson, & Balding, 2012; Lee & Chow, 2014; Yang et al., 2015; Chen, 2016).

A negative correlation between LD and heritability tagged per SNP may well be the rule (Gazal et al., 2017), for at least two reasons. First, if the region surrounding the focal SNP is under evolutionary constraint, then mutations occurring at nearby sites will typically be eliminated by selection and thereby fail to become present-day SNPs contributing to the focal SNP's LD Score. Second, the higher recombination rate in functionally important regions, such as those that are DNase I hypersensitive, leads to a more rapid attenuation of LD between the focal SNP and the neighboring polymorphisms that do manage to persist over evolutionary time. In this case of SNPs with higher LD Scores tagging less heritability, the slope of LD Score regression leads to an underestimation of the true heritability.

[Table 1 about here.]

We conducted a set of simulations to test these theoretical deductions. We chose

the causal SNPs either randomly or on the basis of their LD Scores and studied the impact of this choice on estimates of the heritability. The results are displayed in Table 1. A random selection of causal SNPs led to an average estimate of heritability (0.553) reasonably close to the true *in silico* heritability (0.50). The relationship between LD dependence and heritability estimate appears to be non-monotonic, and we will shortly discuss possible reasons for this. Nevertheless there is an overwhelmingly evident trend for the heritability estimates to be too low when the causal SNPs all have below-median LD Scores (and conversely too large when the causal SNPs all have above-median LD Scores), in accordance with our theory.

Our simulations testing the accuracy of bivariate LD Score regression as an estimator of genetic correlations produced as byproducts estimates of the two univariate heritabilities in each run, and Table 2 presents the results. Surprisingly, estimated heritabilities can vary substantially, beyond what is expected as a result of sampling error, even under the same values of the simulation parameters. For example, even though the average effects were drawn from the normal distribution and then rescaled in the same way, the estimated heritabilities of traits with a low-LD bias ranged from 0.32 to 0.62. It seems that even the small fluctuations in the relationship between LD and effect size induced by our scheme for generating the genetic architecture can have substantial effects on the heritability estimate returned by LD Score regression. (Recall that in a given condition we did not redraw the average effects of the SNPs for a new replicate. We only redrew the non-genetic residuals of the individuals.) Nevertheless we can see that the overall results bear out our theoretical arguments. The average of the estimates over all *unbiased heritabiilty* conditions is 0.768, reasonably close to the true *in silico* value of 0.8. The average of the estimates over all conditions intended to induce an upward bias is 0.882. The average of the estimates over all conditions intended to induce a downward bias is

0.477, suggesting an asymmetrically greater sensitivity to downward rather than upward bias.

[Table 2 about here.]

In summary, even though the simulations whose results are presented in Tables 1 and 2 assigned effects to SNPs in markedly different ways, they jointly affirmed that a dependence of per-SNP heritability on LD Score leads to inaccurate estimates of overall heritability.

3.2 The intercept of univariate LD Score regression as an estimator of confounding

A far more important use of LD Score regression is the estimation and correction of confounding (or any other bias that can inflate the association statistics, such as overestimation of the effective sample size as a result of highly related individuals in the sample). If the intercept of LD Score regression is truly equal to the average chi-square statistic of SNPs that neither affect the phenotype nor tag any causal sites, then dividing all of the GWAS chi-square statistics by the intercept should restore the average chi-square statistic of these null SNPs to the theoretically proper value of unity and bring the Type 1 error rate close to the targeted level. We now examine the extent to which this use of the method is valid.

We first suppose that the magnitude of a SNP's correlation with environmental factors affecting the phenotype is independent of its LD Score. Such independence implies that the conditional average increase in the chi-square statistic due to confounding at each possible LD Score does not in fact vary as a function of LD Score, and thus the entire regression line is elevated by a uniform amount. The intercept is expected to be very close to unity in the absence of confounding (Table 1, Figure 2), and therefore the amount by

which the regression line is moved upward can be determined from the departure of the intercept from unity. Furthermore, suppose that null SNPs do not differ from non-null SNPs in the average extent of confounding—which is extremely likely if LD Scores are indeed independent of confounding. After all, SNPs differing in LD Score also differ in their probability of being null (the probability increasing as the LD Score declines), and it is hard to see how the same spurious increase in the chi-square statistic can be maintained as the LD Score varies (and hence as the mixture of null and non-null proportions varies)—unless null SNPs do not in fact differ from non-null SNPs in the extent of their confounding with environmental factors. It follows that null SNPs have an average chi-square statistic equal to the intercept, and division of all chi-square statistics by the intercept will bring their average back to the required value of unity. This conclusion was also reached by Bulik-Sullivan et al. (2015a).

We will now show that division by the intercept can still be viable means of correcting confounding in some situations where LD Scores and SNP-environment correlations are related.

Suppose that the spurious increase in the chi-square statistic depends linearly on LD Score. This is extremely likely if the regression of the total chi-square statistic on LD Score is linear, since it would be quite a coincidence if the superposition of terms with markedly nonlinear relationships with LD Score produced a closely linear relationship. SNPs with an LD Score of one can tag at most one causal SNP, SNPs with an LD Score of two can tag more than one, and so on. The linear extrapolation to an LD Score of zero represented by the intercept is thus very close to the confounding-induced inflation of the chi-square statistics at SNPs that are null by virtue of tagging very few SNPs. If the trait is so highly polygenic that virtually all SNPs with even moderate LD Scores tag at least one causal site, then the intercept estimated under these conditions is sufficient to rescale

the chi-square statistics of the few null SNPs so as to bring their average in line with the theoretical value under the hypothesis of no causality or LD with a causal site.

If the trait is not sufficiently polygenic, then there will be many SNPs with moderate or large LD Scores that happen to be null. Then the intercept reflects the average confounding-induced inflation at only a certain subset of null SNPs, and one might worry that the chi-square statistics of null SNPs outside of this subset are not properly corrected.

When considering realistic reasons for a dependence of confounding on LD Score, however, we find that the intercept continues to be robust. The most likely case of dependence is a direct (non-genetic) effect of parent on offspring phenotype, such as when highly educated parents can help their offspring (even if adopted) become highly educated in turn (Sacerdote, 2007). Figure 1 depicts this situation. The causal effect of the offspring's own genotype on phenotype will be accurately estimated in within-family studies (Laird & Lange, 2006; Lee & Chow, 2013), but the within-family estimate will fall short of the estimate obtained from GWAS of unrelated individuals because the latter also reflects the confounding influence of the parent genotype. Although we have depicted the parent years of education as the mediator of the parent genotype's causal influence (above and beyond its influence through the offspring genotype), the mediator does not necessarily have to be the same phenotype studied in the GWAS of the offspring. When the offspring phenotype is years of education, parent characteristics acting as mediators might also include intelligence, income, and other determinants of social status (Clark, 2014). Our results below are applicable whenever there is a high genetic correlation between the offspring phenotype and the mediating parent characteristic (Marioni et al., 2014; Hill et al., 2016).

In this case the spurious increase in the chi-square statistic is equal to the square of the true genetic coefficient up to a constant factor, plus whatever part of the spurious increase

does not depend on the true coefficient (Lee, 2012). Critically this affine dependence means that null SNPs, regardless of their LD Scores, will have an average chi-square statistic equal to the intercept so long as the confounding that is independent of the true genetic coefficient does not depend on LD Score. Division of all chi-square statistics by the intercept again leads to the subset of statistics corresponding to null SNPs having the required average of unity. Note that the inability to factor out the contribution of confounding to the chi-square statistics of non-null SNPs in these cases simply leaves us with more or less statistical power to detect such SNPs without affecting the Type 1 error rate.

We provide more justification of this argument in the Supplementary Note. Here, we use simulations based on our MCTFR genetic data to provide further support. Briefly, we conducted GWAS of a simulated phenotype potentially affected by the genotypes of the parents (Figure 1) and applied LD Score regression to the resulting summary statistics. Figure 2 displays the results. One consequence of augmenting the sample with DZ twins of the original sample members is that the estimated heritability increased even in the condition with no confounding. This may be the result of SNPs with higher LD Scores having even more LD partners as a result of the "bottleneck" imposed by recent common ancestry, an effect analogous to the increase in heritability estimates obtained with GREML in samples with relatedness (Vattikuti, Guo, & Chow, 2012; Zaitlen et al., 2013). This should not affect our conclusions. What is important is that in both the *unrelated* and *DZ twin* conditions, the heritability estimate increased upon allowing the parent genotype to have an environmentally mediated effect on the offspring phenotype ($P < 0.002$). This form of confounding thus contributes more to the chi-square statistics of the SNPs with the largest LD Scores.

[Figure 2 about here.]

In the *unrelated* conditions, the intercept remained close to unity even when there was confounding by parent genotype. This is consistent with our argument that the intercept is unaffected by this type of confounding despite dependence on LD Score. In the *DZ twin* conditions, the intercept increased as a result of the relatedness between sample members. It further increased upon the addition of confounding by parent genotype. This is consistent with relatedness becoming a type of population stratification when combined with an effect of parent genotype on offspring phenotype; each family is its own population, and even null SNPs become associated with the phenotype because they are indicative of parentage and thus of a key environmental variable affecting the phenotype. Crucially, however, in both the *no confounding* and *confounding by parent genotype* conditions at this level of the *DZ twin* factor, the intercept increased by almost exactly the right amount to offset the inflation of the chi-square statistic at null SNPs. In the rightmost panel of Figure 2, we can see that the chi-square statistics of SNPs on even chromosomes (simulated to be non-causal) became very close to unity upon division by the intercept, irrespective of relatedness and confounding by parent genotype.

We now turn to an important caveat. Our argument concerning the robustness of the intercept depends on the linearity of the LD Score regression. It is certainly possible to create gross violations of linearity in simulations (Bulik-Sullivan et al., 2015a, Supplementary Figure 7). For example, if we depopulate high-LD regions of causal SNPs, then the regression curve can be non-monotonic, rising at first and then declining as the LD Score increases. In this case the slope of LD Score regression can be negative and the intercept greater than unity even in the absence of confounding. For this reason it is a salutary practice to inspect the binned scatterplot for any evidence of substantial nonlinearity, although unfortunately such a plot may not be informative if the sample size is small.

A mild degree of nonlinearity might have some effect on the intercept if the SNPs

with largest LD Scores deviate from the linear trend extrapolated from the SNPs with the smallest LD Scores. This kind of nonlinearity might have been responsible in our first set of simulations for the small but significant deviations of the intercept from unity (Table 1), which roughly tracked how the estimate of heritability deviated from the true value. For this reason it is fortunate that in practice LD Score regression is a weighted regression where the SNPs with the smallest LD Scores receive the largest weights. The purpose of this weighting is to address heteroskedasticity and non-independence; if the regression curve is perfectly linear, then the effect of this weighting is to improve the standard errors. If the curve is nonlinear, then an additional effect is to bring the entire regression line closer to the linear extrapolation from the SNPs with the smallest LD Scores and the intercept thereby closer to the average chi-square statistic of truly null SNPs.

Our conclusion regarding the robustness of LD Score regression as a safeguard against confounding is a novel result of our analysis. Bulik-Sullivan et al. (2015a) went to some lengths to show that LD Scores are uncorrelated with F_{ST} (a measure of population differentiation in allele frequencies) at various geographical scales within Europe. This is very convincing evidence in support of the assumption that confounding is uncorrelated with LD Scores—at least when the confounding takes the form of population stratification usually contemplated by GWAS researchers, the sampling of the individuals in the study from geographically distinct subpopulations differing in both allele frequencies and exposure to environmental factors. But we have found that even if confounding is correlated with LD Scores, the intercept of LD Score regression can still be used to ensure that null SNPs have an average chi-square statistic of close to unity in some important cases, including extreme polygenicity and an environmentally mediated effect of the parent genotype.

With all of these considerations in mind, we turn to the recent work of de Vlaming,

Johannesson, Magnusson, Ikram, and Visscher (2017a). These authors found that a very large degree of population stratification in their simulations leads to an intercept falling short of the magnitude required to restore the Type 1 error rate and also an overestimate of the heritability. Whatever the problem may be, evidence for it can be seen in their binned scatterplot of chi-square statistics and LD Scores (an average over many replicates), which shows a nonlinearity in the leftmost simulated data points that we have never observed in real empirical data. It is also worth noting that the problems in these simulations only arise when population stratification is quite extreme, leading to an intercept greater than 1.5 with rather small sample sizes. In this regime, the small multivariate fourth cumulant approximation may no longer be valid, although we think this is unlikely to be the explanation of the simulation results. In any event this example shows the importance of inspecting the binned scatterplot if this has stabilized and being cautious when the intercept is large enough to indicate substantial undiagnosed problems.

3.3 Bivariate LD Score regression as an estimator of genetic correlations

We now consider LD Score regression as an estimator of the genetic correlation between the two traits. Previous studies have found the output of bivariate LD Score regression to be consistently close to what is produced by wholly different methods (Bulik-Sullivan et al., 2015b; Okbay et al., 2016; Shi, Mancuso, Spendlove, & Pasaniuc, 2017), and our goal now is to explain this robustness.

In bivariate LD Score regression, the slope is naively expected to be proportional to the genetic covariance. In the absence of confounding and sample overlap, the intercept is zero since the expected product of two independent and null-distributed Z statistics is zero. Any upward departure of the intercept from zero in this case is indicative of

confounders affecting both traits, just as an upward departure from unity is analogously indicative of confounders affecting the focal trait in the univariate case. Herein lies the power of bivariate LD Score regression as a method; its estimate of the genetic correlation relies on the respective slopes of three regressions on LD Scores, the dependent variables being the product of Z statistics, the chi-square statistics of trait 1, and the chi-square statistics of trait 2. As a result any influence of confounders affecting one or both traits is minimized. Because of this key property, it is important to demonstrate that the estimate of the genetic correlation is reasonably accurate.

The Supplementary Note contains our mathematical treatment of this problem, including our demonstration that the estimate returned by bivariate LD Score regression is unaffected by a direct effect of parent genotype on offspring phenotype. (The intuitive explanation of this result is that this type of confounding increases the magnitude of each GWAS coefficient by an amount that depends on its true value, leaving unaltered the relationship between the entire vector of coefficients and another vector corresponding to a differently defined phenotype.) Here we present the results of our simulations (Figure 3). Recall that the heritability estimates produced by this set of simulations vary substantially even for the same values of the governing parameters (Table 2). Another way to put this is that two traits with the same true heritabilities, selection pressures, extent of pleiotropy, and so forth might have different realized heritabilities as estimated by LD Score regression in the limit of infinite sample size, due to evolutionary contingency. Such variability may affect estimates of genetic correlations as well. We can see in Figure 3, however, that the genetic correlation appears to be more robust. Although two of the three average estimates in the *unbiased heritability* conditions are significantly different from the *in silico* true value, this average is never far from the truth in a practical sense.

[Figure 3 about here.]

For the most part, the simulation results affirm that bivariate LD Score regression is a robust estimator of the genetic correlation, even when estimates of the heritabilities are tremendously biased (Table 2). The largest discrepancies from the true genetic correlation, approaching 0.10 in magnitude, occurred when the correlation was fixed to the fairly low value of 0.2. Our analysis in the Supplemental Note shows that the bias in genetic correlation is expected to be larger for smaller genetic correlation. Although our simulations show that the estimate returned by bivariate LD Score regression may be too large when the true genetic correlation is small and the estimated heritabilities of one or both traits are biased upward by per-SNP heritability increasing with LD Score, such a bias is not likely to be the rule. Of the 20 roughly independent traits analyzed by Gazal et al. (2017), not a single one showed a tendency for per-SNP heritability to increase with LD Score. The opposite trend is to be expected for the evolutionary and biological reasons given in our earlier discussion. At all values of the genetic correlation, then, the estimate returned by bivariate LD Score regression is likely to be reasonably accurate and even to miss low if there is any bias at all.

4 Discussion

The regression of GWAS association statistics on LD Scores partitions the statistics into a part that covaries with LD Scores (the slope) and a part that does not (the intercept). Polygenic causal signal contributes to the first part by necessity, whereas confounding and other biases spuriously inflating the statistics need not—and typically do not—make any such contribution. This insight lies at the heart of LD Score regression.

The reason that the slope of LD Score regression cannot be used to estimate the heritability of a trait (or the genetic covariance between two traits) is that per-SNP heritability (genetic covariance) will itself vary as a function of LD Score, such that naive

estimates based on LD Score regression will typically fall short of the target quantities.

It had been presumed that in order for division by the intercept to restore the average chi-square statistic of null SNPs to the theoretically prescribed value of unity, LD Scores must be uncorrelated over SNPs with the extent of confounding with environmental influences on the phenotype. In the framework of Bulik-Sullivan et al. (2015a), this is equivalent to the absence of a correlation between LD Scores and the F_{ST} characterizing the two subpopulations. There may be such a correlation, however, in certain cases such as when the phenotype of the parent affects the phenotype of the offspring through some environmental mechanism. Remarkably we found that LD Score regression remains a robust means of correcting the association statistics, for in such a case the deviation of the intercept from unity reflects the degree of confounding at just those SNPs that are neither causal themselves nor in LD with any causal sites—that is, at precisely those SNPs where otherwise an excess of false positives might occur. We have focused on this case of confounding by parent genotype because of evidence for its occurrence reported in recent work (Kong et al., 2018; Lee et al., in press) and also because we have not been able to think of any other means by which LD Scores might become correlated with the extent of confounding with environmental influences on the phenotype. If any such means is discovered in the future, it will be important to consider whether the extent of confounding at SNPs that are null by virtue of tagging few other SNPs of any kind can be generalized to a SNP with a larger LD Score that is nevertheless null because none of the many SNPs tagged by it happen to be causal. If this generalization is warranted, then the intercept of LD Score regression will continue to be a robust means of bringing the Type 1 error rate closer to the desired level.

These conclusions depend importantly on the linearity of the relationship between LD Scores and the GWAS chi-square statistics (product of Z statistics). This is essen-

tially because without linearity there is no guarantee that the intercept of a particular least-squares regression equals the conditional expected value of the dependent variable characterizing observations with a zero value of the independent variable. In real-data applications of LD Score regression to date, the chi-square vs. LD Score scatterplots have always borne out approximate linearity, and they should continue to be inspected in future applications. When users follow the developers' recommendations for weighting of the SNPs in the regression, those SNPs with smaller LD Scores will receive larger weights, which in the case of nonlinearity brings the intercept closer to the conditional expected chi-square statistic of null SNPs.

Despite the inability of LD Score regression to estimate the heritability (genetic covariance) without bias, the method is able to estimate the genetic correlation with reasonable accuracy. Our mathematical analysis and simulation results suggest that the estimate should be treated with caution if it is statistically significant but nevertheless small and particularly if the heritability estimate of either trait seems biased upward. The latter possibility can be checked by judging whether either estimate of the SNP-based heritability seems much larger relative to the total heritability than is typically the case (about 50 percent) (Shi, Kichaev, & Pasaniuc, 2016) or whether stratified LD regression with the annotations introduced by Gazal et al. (2017) (discussed further below) indicate that per-SNP heritability increases with LD. Our mathematical argument about the robustness of the estimated genetic correlation in the Supplementary Note will be valid if the genetic correlation depends primarily on direct overlap of the causal sites affecting the two traits—and negligibly on SNPs in LD with more potential causal sites thereby being more likely to tag one site affecting trait 1 and a distinct site affecting trait 2, with the signs of the alleles coupled with the reference allele at the tagging SNP showing a consistency across the genome. Such a genome-wide pattern seems quite implausible; for example, if it is to

create a misleading nonzero estimate of the genetic correlation when the true value is in fact zero, it amounts to causal sites that affect the two traits occurring in the same genes and regulatory elements, with the appropriate coupling of alleles, but never coinciding. Furthermore, one might argue that this biologically implausible scenario does not necessarily invalidate bivariate LD Score regression as an estimator of the genetic correlation when this quantity is defined properly. We have adopted the definition $r := \alpha_1' \alpha_2 / \sqrt{h_1^2 \, h_2^2}$ (see the Supplementary Note) because this seems most consistent with the definition of heritability given in the original LD Score regression paper (Bulik-Sullivan et al., 2015a), but other authors have included contributions from LD and consistent coupling of allele signs to the definition of the genetic correlation (Lynch & Walsh, 1998).

A use of LD Score regression that we did not study in this work is the functional partition of heritability between different parts of the genome (Finucane et al., 2015). Simulation studies conducted by the authors suggest that this use is also quite robust, and this is probably the result of a similar cancellation of biases from numerator and denominator. A more recent work has introduced functional annotations describing many properties of SNPs related to per-SNP heritability, including MAF, local recombination rate, and extent of LD with neighbors (Gazal et al., 2017). Because these factors related to per-SNP heritability are thus effectively controlled, we might expect that the heritability estimate produced by stratified LD Score regression with these new annotations will be closer to the true heritability. Supplementary Table 8b of Gazal et al. (2017) does bear out a weak tendency for heritability estimates to increase in this manner. This same table, however, reveals a much stronger influence on heritability estimates; when stratified LD Score regression is applied to the summary statistics of a single large study rather than a meta-analysis of multiple studies, its heritability estimate becomes markedly higher and even approaches the estimate returned by the GREML method. Imperfect genetic

correlations between studies thus seem to affect this output of GWAS as well (de Vlaming et al., 2017b). Applying stratified LD Score regression with the LD-related annotations to a large sample of a homogeneous population, analyzed with a uniform pipeline, appears to be promising strategy if the goal is to estimate heritability accurately.

In a field already marked by remarkable progress toward the goal of elucidating the causal relationship between its variables of interest without undue hindrance by confounding, LD Score regression adds a powerful new tool that allows whatever confounding there may be in a GWAS to be estimated and removed. In addition, it is a robust estimator of the genetic correlation, which is valuable in its own right because of its relevance to the causal nature of the phenotypic correlation (Duffy & Martin, 1994).

Acknowledgements

This research was supported in part by the Intramural Research Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

References

Barban, N., Jansen, R., de Vlaming, R., Vaez, A., Mandemakers, J., Tropf, F. C., … Mills, M. C. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature Genetics*, *48*(12), 1462–1472. doi:10.1038/ng.3698

Beauchamp, J. P., Cesarini, D., Johannesson, M., Lindqvist, E., & Apicella, C. (2011). On the sources of the height-intelligence correlation: New insights from a bivariate ACE model with assortative mating. *Behavior Genetics*, *41*(2), 242–252. doi:10.1007/s10519-010-9376-7

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., … Neale, B. M. (2015b). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. doi:10.1038/ng.3406

Bulik-Sullivan, B., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, … Neale, B. M. (2015a). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. doi:10.1038/ng.3211

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7. doi:10.1186/s13742-015-0047-8

Chen, G.-B. (2016). On the reconciliation of missing heritability for genome-wide association studies. *European Journal of Human Genetics*, *24*(12), 1810–1816. doi:10.1038/ejhg.2016.89

Clark, G. (2014). *The son also rises: Surnames and the history of social inequality*. Princeton, NJ: Princeton University Press.

de los Campos, G., Sorensen, D., & Gianola, D. (2015). Genomic heritability: What is it? *PLOS Genetics*, *11*(5), e1005048. doi:10.1371/journal.pgen.1005048

de Vlaming, R., Johannesson, M., Magnusson, P. K. E., Ikram, M. A., & Visscher, P. M. (2017a). Equivalence of LD-score regression and individual-level-data methods. *bioRxiv*. doi:10.1101/211821

de Vlaming, R., Okbay, A., Rietveld, C. A., Johannesson, M., Magnusson, P. K. E., Uitterlinden, A. G., … Koellinger, P. D. (2017b). Meta-GWAS Accuracy and Power (MetaGAP) calculator shows that hiding heritability is partially due to imperfect genetic correlations across studies. *PLOS Genetics*, *13*(1), e1006495. doi:10.1371/journal.pgen.1006495

Devlin, B. & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, *55*(4), 997–1004. doi:10.1111/j.0006-341X.1999.00997.x

Duffy, D. L. & Martin, N. G. (1994). Inferring the direction of causation in cross-sectional twin data: Theoretical and empirical considerations. *Genetic Epidemiology*, *11*(6), 483–502. doi:10.1002/gepi.1370110606

Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., … Price, A. L. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, *47*(11), 1228–1235. doi:10.1038/ng.3404

Fisher, R. A. (1941). Average excess and average effect of a gene substitution. *Annals of Eugenics*, *11*, 53–63.

Freedman, D. (1999). From association to causation: Some remarks on the history of statistics. *Statistical Science*, *14*(3), 243–258.

Gazal, S., Finucane, H. K., Furlotte, N. A., Loh, P.-R., Palamara, P. F., Liu, X., … Price, A. L. (2017). Linkage disequilibrium-dependent architecture of human complex traits shows action of negative selection. *Nature Genetics*, *49*(10), 1421–1427. doi:10.1038/ng.3954

Goldstein, D. B. (2011). The importance of synthetic associations will only be resolved empirically. *PLOS Biology*, *9*(1), e1001008.

Hill, W. D., Hagenaars, S. P., Marioni, R. E., Harris, S. E., Liewald, D. C. M., Davies, G., … Deary, I. J. (2016). Molecular genetic contributions to social deprivation and household income in UK Biobank. *Current Biology*, *26*(22), 3083–3089. doi:10.1016/j.cub.2016.09.035

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjálmsson, B. J., Young, A. I., Thorgeirsson, T. E., … Stefansson, K. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424–428. doi:10.1126/science.aan6877

Laird, N. M. & Lange, C. (2006). Family-based designs in the age of large-scale gene-association studies. *Nature Reviews Genetics*, *7*(5), 385–394. doi:10.1038/nrg1839

Lee, J. J. (2012). Correlation and causation in the study of personality (with discussion). *European Journal of Personality*, *26*(4), 372–412. doi:10.1002/per.1863

Lee, J. J. & Chow, C. C. (2013). The causal meaning of Fisher's average effect. *Genetics Research*, *95*(2–3), 89–109. doi:10.1017/S0016672313000074

Lee, J. J. & Chow, C. C. (2014). Conditions for the validity of SNP-based heritability estimation. *Human Genetics*, *133*(8), 1011–1022. doi:10.1007/s00439-014-1441-5

Lee, J. J., Vattikuti, S., & Chow, C. C. (2016). Uncovering the genetic architectures of quantitative traits. *Computational and Structural Biotechnology Journal*, *14*, 28–34. doi:10.1016/j.csbj.2015.10.002

Lee, J. J., Wedow, R., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., … Cesarini, D. (in press). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nature Genetics*.

Lynch, M. & Walsh, B. (1998). *Genetics and the analysis of quantitative traits*. Sunderland, MA: Sinauer.

Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., … Deary, I. J. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence*, *44*, 26–32. doi:10.1016/j.intell.2014.02.006

Miller, M. B., Basu, S., Cunningham, J., Eskin, E., Malone, S. M., Oetting, W. S., … McGue, M. (2012). The Minnesota Center for Twin and Family Research genome-wide association study. *Twin Research and Human Genetics*, *15*(06), 767–774. doi:10.1017/thg.2012.62

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., … Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539–542. doi:10.1038/nature17671

Palla, L. & Dudbridge, F. (2015). A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *American Journal of Human Genetics*, *97*(2), 250–259. doi:10.1016/j.ajhg.2015.06.005

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). New York: Cambridge University Press.

Price, A. L., Patterson, N., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*(8), 904–909. doi:10.1038/ng1847

Sacerdote, B. (2007). How large are the effects from changes in family environment? A study of Korean American adoptees. *Quarterly Journal of Economics*, *122*(1), 119–157. doi:10.1162/qjec.122.1.119

Shi, H., Kichaev, G., & Pasaniuc, B. (2016). Contrasting the genetic architecture of 30 complex traits from summary association data. *American Journal of Human Genetics*, *99*(1), 139–153. doi:10.1016/j.ajhg.2016.05.013

Shi, H., Mancuso, N., Spendlove, S., & Pasaniuc, B. (2017). Local genetic correlation gives insights into the shared genetic architecture of complex traits. *American Journal of Human Genetics*, *101*(5), 737–751. doi:10.1016/j.ajhg.2017.09.022

Speed, D., Hemani, G., Johnson, M. R., & Balding, D. J. (2012). Improved heritability estimation from genome-wide SNPs. *American Journal of Human Genetics*, *91*(6), 1011–1021. doi:10.1016/j.ajhg.2012.10.010

Vattikuti, S., Guo, J., & Chow, C. C. (2012). Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLOS Genetics*, *8*(3), e1002637. doi:10.1371/journal.pgen.1002637

Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. *American Journal of Human Genetics*, *90*(1), 7–24. doi:10.1016/j.ajhg.2011.11.029

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A. E., Lee, S. H., … Visscher, P. M. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics*, *47*(10), 1114–1120. doi:10.1038/ng.3390

Zaitlen, N. A., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., & Price, A. L. (2013). Using extended genealogy to estimate components of heritability for 23

quantitative and dichotomous traits. *PLOS Genetics*, *9*(5), e1003520. doi:10.1371/journal.pgen.1003520

List of Tables

Table 1: Properties of univariate LD Score regression in simulations based on the MCTFR genetic data

LD Scores of causal SNPs	Intercept	Heritability
Very low	1.017 (1.015, 1.019)	0.227 (0.200, 0.255)
Low	1.017 (1.015, 1.019)	0.134 (0.106, 0.162)
Random	0.995 (0.994, 0.996)	0.553 (0.531, 0.575)
High	0.981 (0.979, 0.982)	0.869 (0.846, 0.893)
Very high	0.989 (0.987, 0.990)	0.893 (0.867, 0.920)

Causal SNPs were selected to have the kind of LD Score described in the first column. LD Score regression was then used to estimate the intercept and heritability of the simulated trait in the MCTFR parents. For each of the 5 conditions, 100 replicates were conducted. The parentheses enclose the 95% confidence intervals. The true heritability was fixed to 0.50 in all conditions.

Table 2: The estimated heritabilities in simulations based on the MCTFR genetic data to test the accuracy of estimated genetic correlations

Condition	Trait 1 heritability	Trait 2 heritability
True genetic correlation 0.2		
Unbiased heritability	0.631 (0.613, 0.648)	0.660 (0.643, 0.677)
Heritability biased low	0.379 (0.361, 0.398)	0.380 (0.364, 0.396)
Heritability biased high	0.832 (0.812, 0.852)	0.886 (0.868, 0.904)
One trait low, other high	0.552 (0.533, 0.571)	0.867 (0.848, 0.886)
True genetic correlation 0.5		
Unbiased heritability	0.854 (0.838, 0.870)	0.838 (0.821, 0.856)
Heritability biased low	0.588 (0.571, 0.606)	0.618 (0.601, 0.635)
Heritability biased high	0.873 (0.856, 0.890)	1.105 (1.086, 1.123)
One trait low, other high	0.467 (0.449, 0.485)	0.861 (0.841, 0.880)
True genetic correlation 0.8		
Unbiased heritability	0.822 (0.804, 0.840)	0.800 (0.782, 0.817)
Heritability biased low	0.454 (0.436, 0.473)	0.539 (0.519, 0.559)
Heritability biased high	0.799 (0.779, 0.819)	0.886 (0.869, 0.903)
One trait low, other high	0.320 (0.308, 0.342)	0.830 (0.813, 0.847)

LD Score regression was used to estimate the heritabilities of the two simulated traits in the MCTFR parents. For each of the conditions, 100 replicates were conducted. The parentheses enclose the 95% confidence intervals. The true heritability was fixed to 0.8 in all conditions.

List of Figures

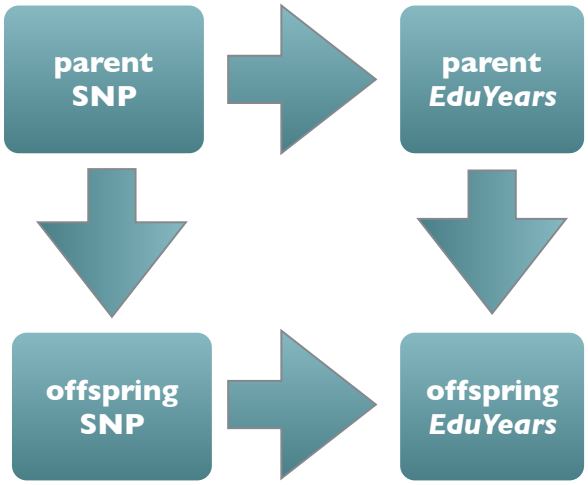

Figure 1: A causal graph (path diagram) displaying a mechanism where the extent of confounding increases with LD Score. Evidence for this mechanism has been uncovered in recent studies (Kong et al., 2018; Lee et al., in press). See the Supplementary Note for a more detailed explication.

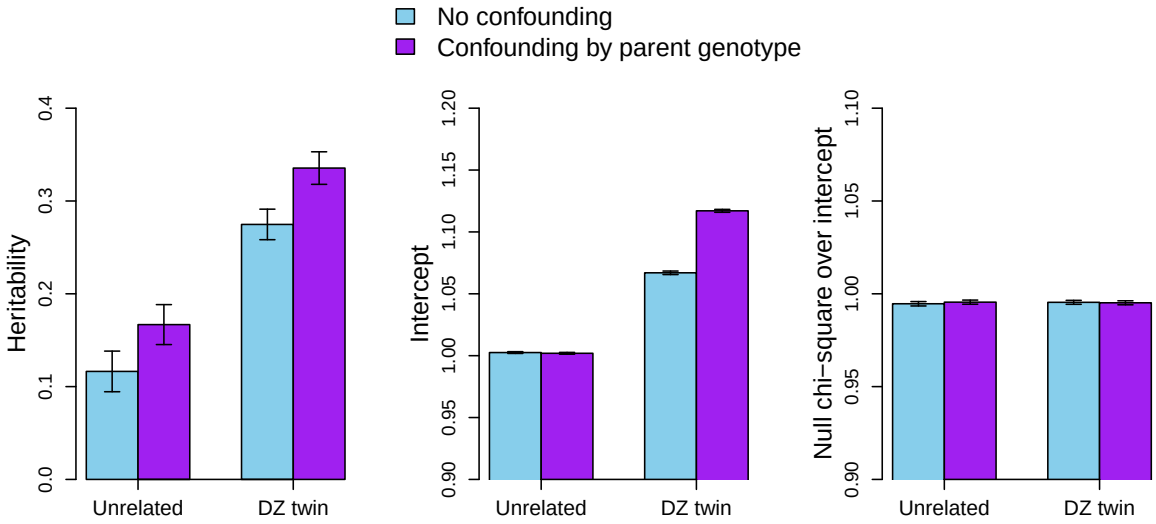

Figure 2: The performance of the LD Score regression intercept in simulations based on the MCTFR genetic data. Each bar and its error correspond to the average of the quantity over 200 replicates and 95% confidence interval. *Unrelated* is a sample of 2,007 nominally unrelated offspring; *DZ twin* augments that sample with 694 individuals, each of whom is a dizygotic twin of an original sample member. In the *confounding by parent genotype* condition, half the average of the two parental breeding (genetic) values was added to the offspring's phenotype. The left panel displays the estimate of heritability based on the regression slope, the center panel displays the intercept, and the right panel displays the mean chi-square statistics of SNPs on even chromosomes (simulated to be non-causal) divided by the intercept.

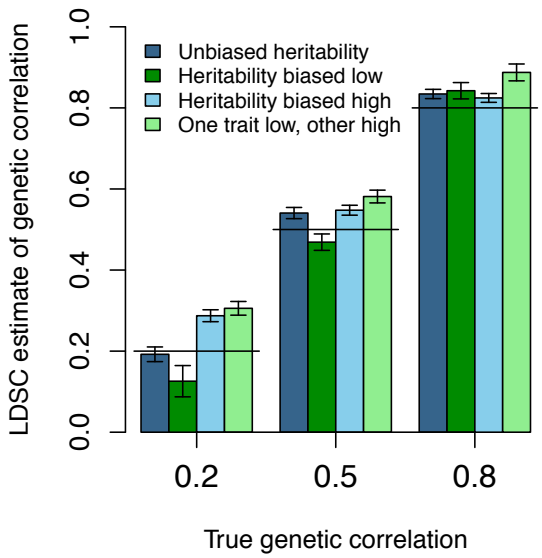

Figure 3: Bivariate LD Score regression was used to estimate the genetic correlation between the simulated traits in the MCTFR parents. For each condition, 100 replicates were conducted. The error bars enclose the 95% confidence intervals. The horizontal line through a given group of bars is placed at the height of the true genetic correlation.