A Unified Framework for Variance Component Estimation with Summary Statistics in Genome-wide Association Studies

Xiang Zhou[1,2*],

1 Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA

2 Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA

∗ **E-mail: xzhousph@umich.edu**

Abstract

Linear mixed models (LMMs) are among the most commonly used tools for genetic association studies. However, the standard method for estimating variance components in LMMs – the restricted maximum likelihood estimation method (REML) – suffers from several important drawbacks: REML is computationally slow, requires individual-level genotypes and phenotypes, and produces biased estimates in case control studies. To remedy these drawbacks, we present an alternative framework for variance component estimation, which we refer to as MQS. MQS is based on the method of moments (MoM) and the minimal norm quadratic unbiased estimation (MINQUE) criteria, and brings two seemingly unrelated methods – the renowned Haseman-Elston (HE) regression and the recent LD score regression (LDSC) – into the same unified framework. With this new framework, we provide an alternative but mathematically equivalent form of HE that allows for the use of summary statistics and is faster to compute. We also provide an exact estimation form of LDSC to yield unbiased and more accurate estimates with calibrated confidence intervals. A key feature of our method is that it can effectively use a small random subset of individuals for computation while still producing estimates that are almost as accurate as if the full data were used. As a result, our method produces unbiased and accurate estimates with calibrated standard errors, while it is computationally efficient for large data sets. Using simulations and applications to 33 phenotypes from 7 real data sets, we illustrate the benefits of our method for estimating and partitioning chip heritability. Our method is implemented in the GEMMA software package, freely available at `www.xzlab.org/software.html`.

Introduction

Linear mixed models (LMMs), sometimes referred to as variance component models, have been widely applied in many areas of genetics. For example, they have been used for linkage analysis and heritability estimations in family studies [1–5], for association analysis to control for individual relatedness and population stratification [6–15], for genomic selection and risk prediction by jointly modeling genome-wide SNPs [16–22], and for rare variant association tests by grouping individually weak effects to improve power [23]. More recently, with growing interest, LMMs have been applied to estimate the proportion of phenotypic variance explained by available SNPs [21, 22, 24–26] – a quantity often referred to as chip heritability – and to partition the chip heritability by different chromosome segments or by different functional genomic annotations [27–30]. These applications all require accurate estimation of variance components in LMMs. However, due to the increasingly large samples being collected today, estimating variance components in genetic data is becoming increasingly challenging.

Two standard statistical methods exist for variance component estimation. The first is the restricted maximum likelihood estimation (REML) method. REML is statistically efficient, but has three major drawbacks. First, despite recent computational innovations [7, 10–12, 31], REML is computationally expensive. It requires not only an iterative optimization algorithm that scales cubically with the number of individuals, but also a pre-computation step to construct a genetic relatedness matrix from genome-wide SNPs – a step that scales both quadratically with the sample size and linearly with the number of markers. Second, REML requires individual-level genotypes and phenotypes, both of which are not readily available in many large consortium studies or meta-analyses. Using individual-level data thus restricts the use of REML and limits the benefits of LMMs in large-scale studies. Finally, REML relies on the normality assumption of residual errors and is not robust to model misspecification. In particular,

in ascertained case control studies or studies with an extreme sample design, REML underestimates chip heritability [32, 33].

A long existing alternative to REML for variance component estimation is the minimal norm quadratic unbiased estimation (MINQUE) method, a method of moments (MoM) [34, 35]. Because the MINQUE estimates are relatively efficient and closely related to REML estimates [36, 37], MINQUE is widely used in many non-genetic settings and has also been applied to animal breeding programs [38]. However, MINQUE shares the drawbacks of REML: it requires computing a genetic relatedness matrix and is thus computationally slow, it does not allow the use of summary statistics, and it has unknown properties in case control studies.

To remedy these three drawbacks, we present a new, alternative framework for variance component estimation. We refer to our method as MQS (MinQue for Summary statistics) as it is based on MINQUE, but addresses its shortcomings with two approximations. Our first approximation allows us to use summary statistics and obtain unbiased estimates at the price of a small reduction of estimation accuracy. Our second approximation allows us to use only a small random subset of individuals to compute the genetic relatedness matrices. This sub-sampling strategy greatly improves computational efficiency, but does not significantly reduce the estimation accuracy in MQS. Most importantly, our framework unifies two seemingly unrelated methods – the renowned Haseman-Elston (HE) regression [32,39–44] and the recent LD score regression (LDSC) [30, 45, 46] – into the same umbrella. With simulations and applications to 33 phenotypes from 7 real data sets, we illustrate the benefits of our method.

Results

Method Overview

Our method applies to the following LMMs that can be used to partition chip heritability into k different non-overlapping categories

$$\mathbf{y} = \sum_{i=1}^{k} \mathbf{g}_i + \boldsymbol{\epsilon}, \quad \mathbf{g}_i \sim \text{MVN}(0, \sigma_i^2 \mathbf{K}_i), \quad \boldsymbol{\epsilon} \sim \text{MVN}(0, \sigma_{k+1}^2 \mathbf{M}), \tag{1}$$

where $\mathbf{y}$ is an n-vector of phenotypes for n individuals; $\mathbf{g}_i$ is an n-vector of random effects representing the combined genetic effects of SNPs in ith category; $\mathbf{K}_i = \mathbf{X}_i \mathbf{X}_i^T / p_i$ is an n by n genetic relatedness matrix computed from the n by p_i genotype matrix for p_i SNPs in ith category; $\boldsymbol{\sigma}^2 = (\sigma_1^2, \cdots, \sigma_k^2)$ are the corresponding variance components; $\boldsymbol{\epsilon}$ is a n-vector of residual errors; σ_{k+1}^2 is the residual error variance; $\mathbf{M} = \mathbf{I} - \mathbf{1}_n \mathbf{1}_n^T / n$ is a projection matrix; and MVN denotes a multivariate normal distribution. Both $\mathbf{y}$ and every column of $\mathbf{X}$ have been centered to have mean zero, allowing us to ignore the intercept and use $\mathbf{M}$ instead of the usual identity matrix $\mathbf{I}$ to constrain the errors to have mean zero. We denote the phenotype variance $s_y^2 = \mathbf{y}^T \mathbf{y} / (n-1)$. We also define the scaled version of the variance components as $\mathbf{h}^2 = (h_1^2, \cdots, h_k^2) = (\sigma_1^2 / s_y^2, \cdots, \sigma_k^2 / s_y^2) = \boldsymbol{\sigma}^2 / s_y^2$ and $h_{k+1}^2 = \sigma_{k+1}^2 / s_y^2$.

We have described our method for variance component estimation in detail in the Methods section. Briefly, our framework, which we refer to as MQS, is based on MINQUE and provides unbiased estimates with calibrated standard errors. MQS has a simple, closed-form solution for estimating the k variance components: $\hat{\boldsymbol{\sigma}}^2 = \mathbf{S}^{-1}\mathbf{q}$, with a k-vector $\mathbf{q}$ and a k by k matrix $\mathbf{S}$ (equation 11). Intuitively, the ith element of $\mathbf{q}$ measures the proportion of variance in phenotypes explained (PVE) by SNPs in ith category when SNPs are independent, while ijth element of $\mathbf{S}$ accounts for the linkage disequilibrium (LD) between SNPs in ith and jth categories. MQS requires the marginal z-scores for computing $\mathbf{q}$, the usual genetic relatedness matrices for computing $\mathbf{S}$ (or the genetic relatedness matrices from a reference panel for estimating $\mathbf{S}$; see below), and a set of pre-specified SNP weights that are used for both $\mathbf{q}$ and $\mathbf{S}$. This set of pre-specified SNP weights is used in MQS to approximate the optimal MINQUE estimating

equations. Different choices of weights represent different ways of approximation and can lead to unbiased estimates with different accuracy.

MQS is a unified framework because different SNP weighting options lead to different estimation methods. We consider two particular weighting options here. The first option is equal SNP weights (equation 9). We refer to the variation of MQS under this weighting option MQS-HEW. MQS-HEW is mathematically equivalent to the renowned Haseman-Elston (HE) cross-product regression [32, 39–44]. However, our particular MQS formulation allows us to both make use of summary statistics and compute the standard errors faster than before. The equivalence between HE and MQS-HEW guarantees the MQS-HEW estimates to be unbiased for case control studies [32,33]. The second option assigns SNP weights as a function of both the LD scores and *a priori* set of variance components (equation 10). We refer to the variation of MQS under this weighting option MQS-LDW. For $k = 1$, MQS-LDW is effectively an exact form of LDSC [45]. However, in contrast to LDSC, MQS-LDW uses $\mathbf{S}$ instead of LD scores to measure SNP correlations. Because LD scores are inevitably under-estimated, using LD scores to measure SNP correlations is expected to yield biased estimates. In particular, when $k = 1$, LDSC will over-estimate the variance components. When $k > 1$, the variance components from LDSC will show different degrees of bias. By providing an exact form of LDSC and a new computing form of standard errors (CI1; see below), MQS-LDW yields unbiased and more accurate estimates with calibrated confidence intervals.

One particular feature of MQS is that we can use a random subset of individuals to obtain an estimate of $\mathbf{S} - \hat{\mathbf{S}}$ – without significantly reducing the accuracy of the variance component estimates. Estimating $\mathbf{S}$ with a smaller sub-sample instead of computing $\mathbf{S}$ with the full data not only improves computational efficiency, but also allows us to apply MQS to data from many consortium studies: thus, we can pair $\mathbf{q}$ computed from the available marginal z-scores in the consortium study with $\hat{\mathbf{S}}$ estimated from a random sub-sample of the study. When such a random sub-sample of the study is not available, we can also use a reference panel, such as the 1000 genome project [47], to estimate $\mathbf{S}$, so long as individuals in the reference panel can be viewed as a sub-sample of the study (e.g. of the same ethnic origin). Our statistical arguments for using the sub-sampling strategy, as detailed in the Methods section, is based on examining the variances of $\mathbf{q}$ and $\hat{\mathbf{S}}$. Briefly, the variance of $\hat{\boldsymbol{\sigma}}^2$, $V(\hat{\boldsymbol{\sigma}}^2)$, is contributed by both $V(\mathbf{q})$ and $V(\hat{\mathbf{S}})$. However, $V(\mathbf{q})$ dominates the contribution and can be $\sqrt{2p'm/n}$ times larger than $V(\hat{\mathbf{S}})$, where p' is the effective number of independent SNPs and m is the number of sub-samples. Thus, we can use a much smaller number of individuals m to estimate $\mathbf{S}$ without increasing $V(\hat{\boldsymbol{\sigma}}^2)$ significantly. In addition, our computation of standard errors explicitly accounts for the uncertainty introduced by using a much smaller set of individuals to estimate $\mathbf{S}$ instead of computing it.

Finally, MQS provides three different options to compute the standard errors. The first option, CI, requires genetic relatedness matrices from the full data but is n times faster than the standard formula used in MoM (equations 15, 16 and 18). The second (CI1) and the third (CI2) options use only summary statistics (equations 20 and 21). CI1 is exact but requires additional summary statistics besides the marginal z-scores equation (equation 25). CI2 follows the approach of LDSC [45] and requires only permutation of the marginal z-scores. However, CI2 assumes SNP independence and is approximate. Thus, CI2, like LDSC, does not yield calibrated confidence intervals when the LD pattern is complicated. Examples where CI2 does not apply include ascertained case control studies [48, 49], admixture populations [50], and related individuals (defined loosely as individuals who are not far away in time from their most recent common ancestor, MRCA [51]).

We summarize the key features of several variance component estimation methods and their computational complexity in Table 1.

Simulations: Point Estimates and Confidence Intervals

We perform simulations to compare the performance of several different methods on variance component estimation. We use two real genotype data sets for simulations: an Australian data set with $n = 3,925$ individuals and $p = 4,352,968$ imputed SNPs [24], and a Finish data set with $n = 5,123$ individuals

and $p = 319,148$ genotyped SNPs [52]. We choose these two data sets not only because both consist of Caucasian individuals, but also because the two differ in LD pattern: the Finland data displays longer LD than the Australia data (Figure S4). The pattern of long LD in the Finland data makes it easy to validate some of our expectations. The Finland data displays such long LD pattern presumably because individuals from the Finland data are more closely related to each other than individuals from the Australia data (notice that the Finland study is not a family study however). For each data set, the real genotypes are used to compute genetic relatedness matrices, with which we simulate phenotypes based on LMMs.

We compare five different methods: (1) REML uses individual-level phenotypes and genotypes; (2) HE regression uses individual-level phenotypes and genotypes; (3) LDSC uses z-scores computed from the full data and LD scores estimated based on 1 MB window from the full data; (4) MQS-HEW uses z-scores computed from the full data, and $\hat{\mathbf{S}}$ estimated from $m = 400$ randomly selected individuals; and (5) MQS-LDW uses z-scores computed from the full data, LD scores estimated based on 1 MB window from genotypes of $m = 400$ randomly selected individuals, and $\hat{\mathbf{S}}$ estimated from the same $m = 400$ selected individuals. Notice that for MQS-HEW and MQS-LDW, a different set of $m = 400$ individuals are used in each replicate.

We first simulate phenotypes under LMMs with $k = 1$. We check three scenarios: $h^2 = 0$, 0.25, or 0.5 (notice that most quantitative traits have a chip heritability below 0.5 [53]). For each scenario, we perform 1,000 replicates. We obtain the variance component estimates from different methods. Figures 1A and 1C show boxplots of these estimates. Figures 1B and 1D show the accuracy of these estimates by plotting the mean squared error (MSE) relative to REML. The simulations validate a few of our expectations.

First, because MQS-HEW with $\mathbf{S}$ is identical to HE and because $\hat{\mathbf{S}}$ is an accurate estimate of $\mathbf{S}$, estimates from MQS-HEW with $\hat{\mathbf{S}}$ are similar to those from HE. In addition, because the variance of the MQS estimates depends on a product of $V(\hat{\mathbf{S}})$ and heritability, estimates from MQS-HEW are more similar to HE for a small h^2 than for a large h^2 (Figure S5).

Second, in practice LD scores are estimated via a sliding window based approach and are inevitably under-estimated. Because LDSC uses these under-estimated LD scores to approximate $\mathbf{S}$, LDSC under-estimates $\mathbf{S}$ and over-estimates the variance components. Such upward bias is more obvious in the Finland data than in the Australia data because of longer LD in the Finland data. For instance, when $h^2 = 0.5$, the LDSC estimates on average are 8.0% higher than expected in the Australia data and 47.4% higher in the Finland data. Estimates from all other methods are unbiased. Because of the bias, the estimates from LDSC are the least accurate ones in all scenarios.

Third, MQS-HEW/HE and MQS-LDW approximate MINQUE in different ways, but both approximations are only accurate when the variance component is small and/or when individuals are unrelated. Thus, both MQS-HEW/HE and MQS-LDW are more accurate for a small h^2 than for a large h^2, and more accurate in the Australia data than in the Finland data. In addition, in the case of $h^2 = 0$, both MQS-HEW/HE and MQS-LDW are more accurate than REML because they effectively assume $h^2 = 0$ *a priori* and are locally optimal.

Fourth, presumably because MQS-LDW uses estimates from MQS-HEW as initial values, and presumably because MQS-LDW uses the extra information of LD scores to compute the SNP weights, MQS-LDW is more accurate than MQS-HEW in the Australia data. However, because long LD reduces the accuracy of LD score estimates, MQS-LDW is of similar accuracy as MQS-HEW in the Finland data.

Next, we simulate phenotypes under LMMs with $k = 6$. We first annotate the genome using six categories as in [29]. The six categories include coding, untranslated region (UTR), promoter, DNAse hyper-sensitivity regions (DHS), intronic and intergenic regions. SNPs are classified into these six categories based on genomic location. We then compute a genetic relatedness matrix for each category, based on which we simulate phenotypes using LMMs with multiple variance components. We examine three different scenarios: (I) a null scenario where all variance components are zero ($h_i^2 = 0$

for $i = 1, \cdots, 6$); (II) an alternative scenario with total chip heritability $h^2 = \sum_{i=1}^{6} h_i^2 = 0.5$, and with each category explaining an equal proportion ($h_i^2 = 0.5/6$ for $i = 1, \cdots, 6$); (III) a more realistic alternative scenario with $h^2 = 0.5$ and with DHS explaining a large proportion of heritability ($h_1^2 = 0.04, h_2^2 = 0.02, h_3^2 = 0.02, h_4^2 = 0.4, h_5^2 = 0.01, h_6^2 = 0.01$). For each scenario, we perform 1,000 replicates. Figure 2 show the estimates and estimation accuracy for scenario III. Figures S6 and S7 show the estimates and estimation accuracy for scenario I and scenario II, respectively. The conclusions from these $k = 6$ simulations are largely consistent with $k = 1$ simulations. For example, all estimates except for LDSC are unbiased. Estimates from LDSC for the total heritability are upward biased. In particular, in scenario III, the total chip heritabilty estimates on average are 6.0% higher than expected in the Australia data and 45.9% higher than expected in the Finland data. However, $k = 6$ simulations also revealed new insights. First, because LD scores in different categories are estimated with different accuracy, individual variance component estimates from LDSC sometimes are upward biased and sometimes downward biased. Second, MQS-LDW is more accurate than MQS-HEW in scenario II with the Australia data, is worse than MQS-HEW in scenario I with the Australia data, and is of comparable accuracy as MQS-HEW in other cases. The comparison between MQS-LDW and MQS-HEW suggests that MQS-LDW is not always more accurate than MQS-HEW even in the Australia data.

Finally, we check whether different methods produce calibrated confidence intervals. To do so, we compute the coverage probabilities of the 95% confidence intervals from different methods using the 1,000 replicates from both $k = 1$ simulations and $k = 6$ simulations (Figure 3). If the confidence interval is calibrated, then the coverage probability is expected to be 0.95, with 99% chance in the range of (93.1%, 96.7%). We use CI to compute standard errors for HE and we use CI1 or CI2 to compute standard errors for MQS-HEW and MQS-LDW. The results are as expected: REML, CI for HE, CI1 for both MQS-HEW and MQS-LDW all produced calibrated confidence intervals. CI2 works well in the Australia data with short range LD, but works poorly in the Finland data with long range LD. Like CI2, LDSC produced calibrated confidence intervals in the Australia data but not in the Finland data. The failure of LDSC confidence intervals in the Finland data is in part due to the estimation bias and in part due to the approximate standard error computation method.

Real Data Applications

To obtain further insights into the differences between various methods, we apply all five methods to estimate chip heritability for 18 phenotypes in three human GWAS data sets. The first GWAS data is the Australian data that contains height measurements for Australian. The second GWAS data is the Finland data that contains 10 quantitative traits, including C-reactive protein (CRP), glucose, insulin, total cholesterol (TC), high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides (TG), body mass index (BMI), systolic blood pressure (SysBP) and diastolic blood pressure (DiaBP). The third GWAS data is the WTCCC data [54], which includes about 14,000 cases from 7 common diseases and about 3,000 shared controls, all typed on a common set of 458,868 SNPs. The 7 common diseases are bipolar disorder (BD), coronary artery disease (CAD), Crohns disease (CD), hypertension (HT), rheumatoid arthritis (RA), type 1 diabetes (T1D) and type 2 diabetes (T2D). We select the WTCCC data because this is a case control study and we expect long range LD in case control studies due to ascertainment [48, 49]. We apply the five methods to these data in the same way as described in the simulations. For LDSC, we also use 10 MB sliding window in addition to the 1 MB window to estimate the LD scores.

The chip heritability estimates are presented in Table 2. For case control studies, we present estimates on the observed scale, which can be easily converted to the liability scale with a known disease prevalence in the population [21,33,55]. For example, for a disease with a disease prevalence of 0.5% in the population and a case proportion of 50% in the case control study, then the scaling factor is 0.47. Because of the small scaling factor, the heritability estimates for some diseases are above one. The heritability estimates in the Finland data are largely consistent with a previous study [56]. The estimates in the WTCCC are

also consistent with a previous study [33]. The heritability estimate for height in the Australia data is slightly smaller than that from previous studies [21, 24]. This is because we used imputed data here. Using imputed data is known to disrupt LD pattern and reduce heritability estimates [57].

The real data results are consistent with what we see from the simulations. First, MQS-HEW estimates and the standard errors (CI1) are almost identical to that of HE (CI). Second, LDSC estimates are larger than the other estimates, consistent with the upward bias in the simulations. The bias is also more obvious in the Finland and the WTCCC data than in the Australia data. The bias can be reduced by using 10 MB estimation window instead of 1 MB but cannot be completely removed. In contrast, consistent with its known downward bias in case control studies [32, 33], REML estimates are consistently smaller in 7 disease phenotypes. Third, MQS-LDW estimates are largely similar to MQS-HEW for quantitative traits and for diseases with a low chip heritability estimate (e.g. CAD, HT, RA, T2D). However, MQS-LDW estimates are noticeably different from MQS-HEW for diseases with a high chip heritability estimate (e.g. BD, CD, T1D). Because we do not know if MQS-LDW provides unbiased estimates in case control studies, we do not know whether such difference was due to a large variance or a potential bias of MQS-LDW in case control studies. Fourth, consistent with simulations, CI2 often produces overly narrow standard errors when compared with CI1 (Table S2). However, for two disease phenotypes (RA and T1D), the standard errors from CI2 are extremely large, suggesting that the calibration issue with CI2 may not always favor one direction. The standard errors of LDSC are similar to that of CI2 but are different from the exact method CI1 for most traits.

Our method requires genotypes from a random sub-sample of the study to estimate S. When such a subset of individuals is not available, we can use a reference panel to estimate S, so long as individuals in the reference panel can be viewed as a sub-sample of the study. However, a mismatch between the reference panel and the study sample can cause estimation bias. In addition, using a separate reference panel prevents us from using the exact method CI1 to compute the standard errors. Here, we explore the use of genotype data from the 1,000 genomes project [47] for chip heritability estimation in the three GWASs. Specifically, instead of using 400 randomly selected individual from the study sample, we use 503 individuals of European ancestry from the 1,000 genomes project to estimate S. The chip heritability estimates for all traits from the three data sets are shown in Table S2. For both the Australia and WTCCC data sets, using the 1,000 genomes data as a reference panel produce similar results. However, for the Finland data set, the estimates from using the 1,000 genomes data are much larger, suggesting a potential over estimation. The results suggest that a match between the reference panel and study sample is critical for accurate estimation. In addition, because we can only use CI2 to compute the standard errors, the standard errors suffer from the same drawback as detailed in the previous paragraph.

Finally, we apply MQS-HEW and MQS-LDW methods to analyze 8 phenotypes from four consortium studies. These phenotypes include BMI ($n = 120,569$), height (HT, $n = 129,945$) from the GIANT consortium [58, 59], HDL ($n = 88,754$), LDL ($n = 84,685$), TC ($n = 89,005$) and TG ($n = 85,691$) from the Global Lipids Genetics Consortium [60], fasting glucose (FG, $n = 58,074$) from the MAGIC consortium [61], and Crohn's disease (CD, $n = 21,447$) from the International Inflammatory Bowel Disease Genetics Consortium [62]. The data have been pre-processed by a previous study [63]. We further select a common set of $p = 5,014,740$ SNPs among these phenotypes for analysis. We partition SNPs into the same six functional categories (coding, UTR, promoter, DHS, intronic and else) as before [29]. Because only z-scores are available for these phenotypes, we have to use CI2 to compute the standard errors and use genotypes from 503 individuals of European ancestry in the 1000 genomes project [47] as a reference panel to estimate $\mathbf{S}$. In addition, following previous approaches [30], to contrast the importance of different categories, we focus on estimating the relative value instead of the absolute value of variance components. Specifically, as in [30], we construct a fold enrichment parameter, defined as the ratio between the per-SNP variance in one category and the per-SNP variance in all categories, to quantify the relative importance of different functional categories (Supplementary Text).

Figure 4 shows the enrichment parameters for six categories in 8 phenotypes estimated by either

MQS-HEW or MQS-LDW. The results from both MQS-HEW and MQS-LDW are consistent with what we expect [30]: for most phenotypes (with the notable exception of BMI), the per-SNP variance in the coding region is the largest, followed by the UTR, promoter and the DNS regions. The per-SNP variance for both the intronic and intergenic regions are close to zero. The enrichment estimates between MQS-HEW and MQS-LDW are similar overall, though the enrichment of the coding region is estimated to be larger in MQS-LDW than in MQS-HEW for the lipid phenotypes.

Discussion

We have presented a novel framework, MQS, for variance component estimation with summary statistics. MQS is computationally efficient for large data and produces unbiased estimates with calibrated standard errors. MQS is also flexible and can be used with other methods to model uneven linkage disequilibrium [25, 64] or model the effect size dependency on minor allele frequencies [21]. In addition, MQS can be extended to model multiple correlated phenotypes [13,46] and/or incorporate overlapping SNP functional annotations. With simulations and applications to 33 phenotypes from 7 GWASs, we have shown the benefits of our method.

We have focused on two variations of MQS in the present study. Both variations, MQS-HEW and MQS-LDW, yield unbiased estimates but with varying degrees of accuracy. Although we cannot tell in advance which method is more accurate for a particular data set, simulations suggest that, when $k = 1$, MQS-LDW may be more accurate than MQS-HEW for quantitative traits and unrelated individuals. The superior accuracy presumably stems from the fact that MQS-LDW uses LD score information and relies on *a priori* set of estimates from MQS-HEW. However, MQS-LDW is not always more accurate than MQS-HEW. In real data applications, the estimates from MQS-LDW are often very similar to that from MQS-HEW. MQS-LDW is also comparable to and sometimes even worse than MQS-HEW for related individuals or for $k > 1$. In addition, MQS-LDW suffers from two important drawbacks. First, because MQS-LDW requires an iterative procedure, computing the standard errors using CI1 for MQS-LDW is less convenient than MQS-HEW. Inconvenience in computing the exact standard errors can affect the usage of MQS-LDW in consortium studies. Second, because of the iterative procedure and the non-linear dependence on the MQS-HEW estimates, it is unclear whether MQS-LDW can produce unbiased estimates in case control studies as MQS-HEW is known to do [32, 33]. Therefore, at this stage, we recommend the use of MQS-HEW as a default choice for both quantitative traits and case control studies. However, exploring other variations of MQS by using other SNP weighting matrices, especially non-diagonal ones, will be an interesting avenue for future research. Our derivation of MQS-LDW provides some possible non-diagonal weight matrices for exploration. It would be ideal to identify a weighting matrix that can lead to estimates that are consistently more accurate than MQS-HEW while remaining unbiased in case control studies.

MQS uses a small random subset of individuals to estimate $\mathbf{S}$. Using $\hat{\mathbf{S}}$ instead of $\mathbf{S}$ reduces much of the computational cost while yielding estimates that are almost as accurate as if the full data were used. For instance, in both our simulations and real data applications, we have used $\sim 10\%$ of the data to estimate $\mathbf{S}$. Using $\sim 10\%$ of the data incurs minimal loss of accuracy but results in an effective ~ 100 fold speed gain (because computational complexity scales with m^2). Our sub-sampling approach of using $\hat{\mathbf{S}}$ instead of $\mathbf{S}$ is motivated by recent genetic studies that make use of a reference panel for genotype imputation [65–68], and more recently, for multi-loci analysis [45,69] (including LDSC [45]): when the full data is not completely observed, these studies rely on a reference panel to impute the missing pieces to construct a complete data. Our approach, however, differs from the previous approaches in two important ways: we actively use a subset of data to estimate certain quantities even when the full data is completely observed (i.e. in line with the idea of stochastic approximation method [70]); and we account for the extra uncertainty introduced by using a smaller subset of data. Importantly, in the present study, we provide an initial set of statistical reasoning to justify our sub-sampling approach in MQS. However, even

with our guidelines, it often remains difficult to choose the right number of sub-samples, m, for practical analysis. A large m would not save much computational time while a small m could be insufficient to produce unbiased estimates. In practice, the optimal choice of m will likely depend on both the number of categories k and the effective number of independent SNPs in each category; thus we caution against the use of an m that is too small when k is large. Despite this small concern, however, we believe the sub-sampling strategy allows MQS to achieve an appealing balance between computational efficiency and statistical efficiency. With increasing data sizes, exploring the benefits of sub-sampling strategy in other statistical methods for large-scale GWASs – as well as other big data applications – is likely to yield fruitful results in the future.

Material and Methods

The Model

Although our method can be reasonably general, we introduce it by considering a particular application: partitioning heritability by different functional categories. To do so, we assume that variants have been pre-classified into k different, non-overlapping functional categories. Our goal is to estimate the proportion of phenotypic variance explained (PVE) by all variants in each category. We consider the following model

$$\mathbf{y} = \sum_{i=1}^{k} \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{MVN}(0, \sigma_{k+1}^2 \mathbf{M}), \tag{2}$$

where $\mathbf{y}$ is an n-vector of phenotypes for n individuals; $\mathbf{X}_i$ is an n by p_i genotype matrix for p_i variants in ith category; $\boldsymbol{\beta}_i$ is a p_i-vector of corresponding effect sizes; $\boldsymbol{\epsilon}$ is a n-vector of residual errors; σ_{k+1}^2 is the residual error variance; $\mathbf{M} = \mathbf{I} - \mathbf{1}_n \mathbf{1}_n^T / n$ is the projection matrix onto the null space of the intercept; and MVN denotes a (degenerate) multivariate normal distribution. Notice that we ignore the intercept and use $\mathbf{M}$ instead of the usual identity matrix $\mathbf{I}$ to reflect the fact that both $\mathbf{y}$ and every column of $\mathbf{X}$ have been centered to have mean zero. Because of the centering, the residual errors follow a multivariate normal distribution with a low-rank covariance matrix $\mathbf{M}$ that constrains the errors to sum to zero. Centering does not affect results but simplifies the algebra. Generalization of the model to incorporate other covariates in addition to the intercept is straightforward, requiring projecting $\mathbf{y}$, every column of $\mathbf{X}$ and the residual errors to the null space of the covariates.

Following previous approaches [21, 24], for each effect size vector $\boldsymbol{\beta}_i$, we specify a normal prior with mean zero and variance σ_i^2/p_i, or $\boldsymbol{\beta}_i \sim \text{MVN}(0, \sigma_i^2/p_i \mathbf{I})$. This normality assumption of effect size leads to an alternative but equivalent form – a LMM with k variance components [21]

$$\mathbf{y} = \sum_{i=1}^{k} \mathbf{g}_i + \boldsymbol{\epsilon}, \quad \mathbf{g}_i \sim \text{MVN}(0, \sigma_i^2 \mathbf{K}_i), \quad \boldsymbol{\epsilon} \sim \text{MVN}(0, \sigma_{k+1}^2 \mathbf{M}), \tag{3}$$

where $\mathbf{g}_i = \mathbf{X}_i \boldsymbol{\beta}_i$ is the combined genetic effects of ith category; $\mathbf{K}_i = \mathbf{X}_i \mathbf{X}_i^T / p_i$ is an n by n genetic relatedness matrix computed from SNPs in ith category; $\boldsymbol{\sigma}^2 = (\sigma_1^2, \cdots, \sigma_k^2)$ are the variance components.

As usual [24, 71, 72], we standardize every column of $\mathbf{X}$ to have variance one. Unlike centering, standardizing the $\mathbf{X}$ columns will affect the results because it changes our prior assumption [21]. Specifically, standardizing the $\mathbf{X}$ columns corresponds to making an assumption that rarer variants tend to have larger effects than common variants, and that marker effect sizes depend on the minor allele frequencies (MAFs) in a particular mathematical form (see [21] for relevant discussion). We do not, however, standardize $\mathbf{y}$, and the phenotype variance is $s_y^2 = \mathbf{y}^T \mathbf{y}/(n-1)$.

At this point, it is useful to define the scaled version of the variance components: $\mathbf{h}^2 = (h_1^2, \cdots, h_k^2) = (\sigma_1^2/s_y^2, \cdots, \sigma_k^2/s_y^2) = \boldsymbol{\sigma}^2/s_y^2$ and $h_{k+1}^2 = \sigma_{k+1}^2/s_y^2$. As we will show below, when the phenotype variance s_y^2 is unknown, $\boldsymbol{\sigma}^2$ and σ_{k+1}^2 are not estimable from summary statistics alone, but $\mathbf{h}^2$ and h_{k+1}^2 are. With these modeling assumptions, PVE by ith category is $\text{PVE}_i = \sigma_i^2/(\sum_{j=1}^{k+1} \sigma_j^2) = h_i^2/(\sum_{j=1}^{k+1} h_j^2)$, $i \in \{1, \cdots, k\}$ [21].

Finally, although we will focus on the specific model defined in equation 3, we note that it is straightforward to generalize our model to incorporate other modeling assumptions. For example, we can generalize the model to incorporate other prior assumptions on the dependence of effect sizes on the MAFs (and other variant information), as long as the dependency is linear in σ_i^2 (i.e. in the form of $g(f_l)\sigma_i^2/p_i$, where f_l is the MAF for the lth variant in the ith category). Such generalization can be achieved by replacing the usual genetic relatedness matrix $\mathbf{K}_i$ with a corresponding weighted genetic relatedness matrix with variant weights depending on $g(f_l)$. Similarly, we can generalize our model to incorporate overlapping categories. If lth variant belongs to k_l different categories, we can assume its effect size variance to be

a weighted function of the variances of each category, or $\sum_{i=1}^{k_l} \sigma_i^2/(p_i k_l)$. With this assumption, we can again generalize our model by replacing the genetic relatedness matrix $\mathbf{K}_i$ with a corresponding weighted genetic relatedness matrix with variance weights $1_{l\in i}/k_l$, where the indicator function $1_{l\in i}$ equals one when the lth variant belongs to the ith category and equals zero otherwise.

MoM and MINQUE

Our goal is to estimate the variance components $\boldsymbol{\sigma}^2$ and the residual error variance σ_{k+1}^2 (or the scaled version $\mathbf{h}^2$ and h_{k+1}^2). The estimated parameters can then be used to compute PVE. To estimate the variance components, we consider the method of moments, which is based on the following set of quadratic equations (e.g. [34])

$$E(\mathbf{y}^T \mathbf{A}_j \mathbf{y}) = \sum_{i=1}^{k} tr(\mathbf{A}_j \mathbf{K}_i)\sigma_i^2 + tr(\mathbf{A}_j)\sigma_{k+1}^2, \tag{4}$$

where each $\mathbf{A}_j$ is a symmetric non-negative definite matrix and tr denotes matrix trace. For estimation with MoM, we replace the expectation on the left hand side (LHS) of equation 4 with the realized value $\mathbf{y}^T\mathbf{A}_j\mathbf{y}$. We then solve the equations to obtain estimates for $\boldsymbol{\sigma}^2$ and σ_{k+1}^2. Because there are $k+1$ parameters in our model, we only need $k+1$ different $\mathbf{A}_j$ to obtain the estimates. Though unnecessarily, we can also use more than $k+1$ $\mathbf{A}_j$ to set up an over-determined linear system, and use the ordinary least squares (OLS) to obtain unbiased estimates. For example, the naive LDSC equation [45] is based on using a set of p different $\mathbf{A}_j$, where $\mathbf{A}_j = \mathbf{X}\boldsymbol{\Lambda}_j\mathbf{X}^T$ and $\boldsymbol{\Lambda}_j$ is a rank one matrix with jjth diagonal element being one and all other elements being zero. (Notice that we call this naive LDSC to distinguish it from LDSC. LDSC does not use OLS estimates based on the naive LDSC equation because of poor performance. Rather, LDSC introduces two extra weights to improve estimation accuracy. See Supplementary Text for details.)

Any choice of $\mathbf{A}_j$ will yield unbiased estimates for $\boldsymbol{\sigma}^2$ and σ_{k+1}^2, but different choices of $\mathbf{A}_j$ affect estimation accuracy. The optimal choice of $\mathbf{A}_j$ is found based on the Minimal Norm Quadratic Unbiased Estimation (MINQUE) criterion [34, 35, 73, 74] and takes the following form

$$\hat{\mathbf{A}}_j = \mathbf{H}^{-1}\mathbf{K}_j\mathbf{H}^{-1}, \tag{5}$$

where $j = 1, \cdots, k+1$, $\mathbf{K}_{k+1} = \mathbf{M}$ and $\mathbf{H} = \sum_{i=1}^{k+1} \sigma_i^2 \mathbf{K}_i$. Notice that we have loosely used the matrix inverse notation to denote a generalized inverse, and we have also loosely used hat on top of $\mathbf{A}_j$ to not denote an estimate but to denote an optimal choice. Under normality assumptions, MINQUE is also referred to as the best quadratic unbiased estimation (BQUE) [75] and/or the minimal variance quadratic unbiased estimation (MIVQUE) [73, 74].

This optimal $\hat{\mathbf{A}}_j$ depends on a set of variance components that are unknown *a priori*. Thus, we cannot use the optimal $\hat{\mathbf{A}}_j$ directly in practice. Two options are available to obtain MINQUE estimates in practice. The first option is to apply an iterative procedure on equation 4: in each iteration t we plug in the current estimates $\hat{\boldsymbol{\sigma}}_t^2, \hat{\sigma}_{k+1,t}^2$ in $\mathbf{H}$ and $\mathbf{A}_j$ to obtained the updated estimates $\hat{\boldsymbol{\sigma}}_{t+1}^2, \hat{\sigma}_{k+1,t+1}^2$. The resulting algorithm is often referred to as an iterative MINQUE, or I-MINQUE, which, not surprisingly, is REML [36]. The second option is to use a pre-determined set of parameters $\tilde{\boldsymbol{\sigma}}_i^2, \tilde{\sigma}_{k+1}^2$ to construct $\tilde{\mathbf{H}} = \sum_{i=1}^{k+1} \tilde{\sigma}_i^2 \mathbf{K}_i$ and then use $\tilde{\mathbf{A}}_j = \tilde{\mathbf{H}}^{-1}\mathbf{K}_j\tilde{\mathbf{H}}^{-1}$. For example, MINQUE(0) sets $\mathbf{h}^2 = \mathbf{0}$, $h_{k+1}^2 = 1$ to obtain estimates of reasonably accuracy [76, 77]. However, both these two options have the same drawbacks mentioned in the introduction.

First Approximation: SNP Weights and Approximation to MINQUE

We aim to develop an approximation form of the optimal $\hat{\mathbf{A}}_j$ that allows the use of summary statistics, facilitates computation, and is reasonably accurate and thus maintains the statistical efficiency comes

with the optimal $\hat{\mathbf{A}}_j$. The particular approximation we consider takes the following form

$$\tilde{\mathbf{A}}_j = \mathbf{X}_j \mathbf{W}_j \mathbf{X}_j^T / p_j \ (j = 1, \cdots, k), \quad \tilde{\mathbf{A}}_{k+1} = \mathbf{M} \tag{6}$$

where $\mathbf{W}_j = diag(w_{j1}, \cdots, w_{jp_j})$ is a pre-specified p_j by p_j diagonal matrix of SNP weights; and $\mathbf{X}_i$ again is an n by p_i genotype matrix for p_i variants in ith category. $\hat{\mathbf{A}}_j$ $(j \neq k+1)$ is effectively a weighted genetic relatedness matrix, while $\tilde{\mathbf{A}}_{k+1}$ approximates $\hat{\mathbf{A}}_{k+1}$ by assuming that $\mathbf{H}$ is approximately $\mathbf{M}$. Notice that $\tilde{\mathbf{A}}_{k+1}$ ensures that $\sum_{i=1}^{k+1} \sigma_i^2 = s_y^2$ and $\sum_{i=1}^{k+1} h_i^2 = 1$. Because a scale transformation of $\mathbf{W}_j$ does not affect results, to simplify the algebra, we constrain $tr(\mathbf{W}_j) = p_j$ or equivalently that the average SNP weight equals one.

With this set of $\tilde{\mathbf{A}}_j$, the set of $k+1$ estimating equations from equation 4 become

$$\mathbf{y}^T \tilde{\mathbf{A}}_j \mathbf{y} = \sum_{i=1}^{k} tr(\mathbf{K}_i \tilde{\mathbf{A}}_j) \sigma_i^2 + tr(\tilde{\mathbf{A}}_j) \sigma_{k+1}^2, \tag{7}$$

$$\mathbf{y}^T \mathbf{y} = \sum_{i=1}^{k} tr(\mathbf{K}_i) \sigma_i^2 + tr(\mathbf{M}) \sigma_{k+1}^2. \tag{8}$$

where $tr(\mathbf{K}_i) = tr(\mathbf{M}) = tr(\tilde{\mathbf{A}}_j) = n - 1$. We refer to the above equations as MQS estimating equations.

We are yet to specify the weighting matrices $\mathbf{W}_j$. In the present study, we consider two particular choices. The two choices represent different ways of approximating the optimal $\hat{\mathbf{A}}_j$ and are related to the HE regression [32, 39–44] and the LDSC [30, 45, 46], respectively.

HE Weights

The first set, which we refer to as the HE weights, assigns equal weights for all SNPs, or

$$w_{jl} = 1. \tag{9}$$

The HW weights are derived based on an approximation to ensure $\tilde{\mathbf{A}}_j \approx \hat{\mathbf{A}}_j$ (Supplementary Text). The approximation becomes exact for unrelated individuals or a trait with zero hertiablity.

We refer to the variation of MQS under the HE weights as MQS-HEW. Surprisingly, MQS-HEW is equivalent to both MINQUE(0) [76, 77] and the HE regression (Supplementary Text). Because of the equivalence between MQS-HEW and HE, MQS-HEW is expected to provide unbiased estimates for case control studies [33]. Compared with HE and MINQUE(0), however, our MQS formulation allows the use of summary statistics to compute both the point estimates and the standard errors (see below).

LD Weights

The second set, which we refer to as the LD weights, takes the following form

$$w_{jl} = (\sum_i (n-1)/p_i l_{(jl \sim i)} \tilde{h}_i^2 + 1)^{-2} / c_j, \tag{10}$$

where $\tilde{h}_i^2$ is a pre-specified estimate of h_i^2; $l_{(jl \sim i)}$ is the LD score of SNP jl with respect to all SNPs (including itself if $j = i$) in ith category; and c_j is a normalizing constant of jth category to ensure that $\sum_l w_{jl} = p_j$. In practice, we use the variance component estimates from MQS-HEW as $\tilde{h}_i^2$, but restricted them to be between 0 and 1 for algorithm stability. The LD score here is defined to be the summation of squared correlations between the jlth SNP and all SNPs in the ith category minus the expectation under the null, or $l_{(jl \sim i)} = \sum_{m=1}^{p_i} (r_{jl,im}^2 - 1/(n-1))$. The LD weights are derived based on

two approximations to ensure $\tilde{\mathbf{A}}_j \approx \hat{\mathbf{A}}_j$ (Supplementary Text). The approximations become exact also for unrelated individuals or a trait with zero hertiablity.

We refer to the variation of MQS under the LD weights as MQS-HEW. When $k = 1$ and when the LD scores are exact, MQS-LDW is mathematically equivalent to LDSC [45] (Supplementary Text). However, LD scores in practice are estimated via a sliding window based approach [45] and are inevitably under-estimated. Under-estimation of the LD scores have different impacts on MQS-LDW and LDSC. For MQS-LDW, the estimates with estimated LD scores are still unbiased but are less accurate than that based on exact LD scores. However, because LDSC uses estimated LD scores to approximate $tr(\mathbf{K}_i\tilde{\mathbf{A}}_j)$ instead of computing it as in MQS-LDW, LDSC is expected to under-estimate $tr(\mathbf{K}_i\tilde{\mathbf{A}}_j)$ and thus over-estimate the variance components.

When $k > 1$, MQS-LDW is not longer equivalent to the stratified LDSC [30] even with exact LD scores. Unlike $k = 1$, it is no longer possible to convert MQS-LDW estimating equations to a set of linear equations that relate per-SNP z-scores to per-SNP LD scores. Similar to the case of $k = 1$, the stratified LDSC estimates are expected to be biased. This time, because the LD scores for SNPs in different categories are inevitably under-estimated to a different degree, the direction and degree of bias in the stratified LDSC are expected to vary across multiple variance components.

Point Estimates and Confidence Intervals

With the MQS estimating equations 7 and 8, it is straightforward to obtain variance component estimates. To do so, we subtract equation 8 from each equation in 7, solve the resulting k equations to estimate $\boldsymbol{\sigma}^2$, and plug in the estimated $\boldsymbol{\sigma}^2$ to equation 8 to estimate σ^2_{k+1}. The resulting MQS estimates are in a simple, closed-form solution

$$\hat{\boldsymbol{\sigma}}^2 = \mathbf{S}^{-1}\mathbf{q}, \tag{11}$$

$$\hat{\sigma}^2_{k+1} = s^2_y - \sum_{i=1}^{k}\hat{\sigma}^2_i, \tag{12}$$

where the elements in the k-vector $\mathbf{q}$ and the k by k matrix $\mathbf{S}$ are

$$\mathbf{q}_i = \mathbf{y}^T(\tilde{\mathbf{A}}_i - tr(\tilde{\mathbf{A}}_i)/tr(\mathbf{M}))\mathbf{y}/(n-1)^2 = \mathbf{y}^T(\tilde{\mathbf{A}}_i - \mathbf{I})\mathbf{y}/(n-1)^2, \tag{13}$$

$$\mathbf{S}_{ij} = (tr(\tilde{\mathbf{A}}_j\mathbf{K}_i) - tr(\tilde{\mathbf{A}}_j)tr(\mathbf{K}_i)/tr(\mathbf{M}))/(n-1)^2 = tr(\tilde{\mathbf{A}}_j\mathbf{K}_i)/(n-1)^2 - 1/(n-1). \tag{14}$$

The variance for the estimates are

$$V(\hat{\boldsymbol{\sigma}}^2) = \mathbf{S}^{-1}V(\mathbf{q})\mathbf{S}^{-1}, \tag{15}$$

$$V(\hat{\sigma}^2_{k+1}) = \mathbf{1}_k^T\mathbf{S}^{-1}V(\mathbf{q})\mathbf{S}^{-1}\mathbf{1}_k, \tag{16}$$

where $\mathbf{1}_k$ is a k-vector of 1s and the k by k covariance matrix $V(\mathbf{q})$ is

$$V(\mathbf{q})_{ij} = V(\mathbf{y}^T(\tilde{\mathbf{A}}_i - \mathbf{I})\mathbf{y}, \mathbf{y}^T(\tilde{\mathbf{A}}_j - \mathbf{I})\mathbf{y}) = 2tr(\mathbf{H}(\tilde{\mathbf{A}}_i - \mathbf{I})\mathbf{H}(\tilde{\mathbf{A}}_j - \mathbf{I}))/(n-1)^4, \tag{17}$$

since $V(\mathbf{y}) = \mathbf{H}$.

The above form of $V(\mathbf{q})$ requires cubic operations to compute. To speed up computation, we instead consider a novel approximation to $V(\mathbf{q})$

$$V(\mathbf{q})_{ij} \approx 2\mathbf{y}^T(\tilde{\mathbf{A}}_i - \mathbf{I})\mathbf{H}(\tilde{\mathbf{A}}_j - \mathbf{I})\mathbf{y}/(n-1)^4. \tag{18}$$

We refer to the above method of computing standard errors as CI. The above approximation is based on replacing the expectation $H = E(\mathbf{y}\mathbf{y}^T)$ with its realized value $\mathbf{y}\mathbf{y}^T$. The approximation is motivated

by the average information (AI) algorithm [78] and effectively uses a realized information matrix in place of the expected information matrix. Our approximation not only makes computation n times faster than the usual MoM, but also allows the use of summary statistics in computing standard errors (see below). We will show later in the simulations that the confidence intervals constructed based on this approximation are indeed calibrated. (As a side note, when the sample size is not large enough and the asymptotic normality does not kick in, then we cannot use $V(\mathbf{q})$ for hypothesis tests. Instead, we need to use a mixture of chi-square distributions to obtain more accurate p-values [79]. We do not explore this issue here as we typically have large sample size in GWASs.)

Second Approximation: Estimating S via Sub-sampling

Up to now, the total computational complexity of MQS scales quadratically with respect to the sample size and linearly with respect to the number of SNPs, or on the order of $O(pn^2)$. In particular, it takes $O(pn)$ time to compute $\mathbf{q}$, $O(pn^2)$ time to compute $\mathbf{S}$, and $O(n^2)$ time to compute $V(\mathbf{q})$. Because the most computationally expensive part is the computation of $\mathbf{S}$, we consider estimating $\mathbf{S}$ instead of computing it. Specifically, we consider using m randomly selected individuals from the full sample to estimate $\mathbf{S}$, or

$$\hat{\mathbf{S}}_{ij} = (tr(\tilde{\mathbf{A}}'_j\mathbf{K}'_i) - tr(\tilde{\mathbf{A}}'_j)tr(\mathbf{K}'_i)/tr(\mathbf{M}'))/(m-1)^2, \quad \tilde{\mathbf{A}}'_j = \mathbf{X}'_j\mathbf{W}\mathbf{X}'^T_j/p_j, \quad \tilde{\mathbf{K}}'_i = \mathbf{X}'_i\mathbf{X}'^T_i/p_i. \tag{19}$$

where $\mathbf{X}'_i$ is the standardized genotype matrix for the subset of individuals and $\mathbf{M}' = \mathbf{I} - \mathbf{1}_m\mathbf{1}_m/m$ is the projecting matrix onto the null space of the intercept for this subset of individuals.

Using the estimated $\hat{\mathbf{S}}$ in place of $\mathbf{S}$ reduces computational complexity from $O(pn^2)$ to $O(pm^2)$, resulting in an overall computational complexity of $O(pn + pm^2)$ for MQS. In addition, using $\hat{\mathbf{S}}$ allows us to apply MQS to data from many consortium studies: we can pair $\mathbf{q}$ computed from the complete data with $\hat{\mathbf{S}}$ estimated from a random sub-sample of the study. When such a random sub-sample of the study is not available, we can also use a reference panel, such as the 1,000 genomes project [47], to estimate $\mathbf{S}$, so long as individuals in the reference panel can be viewed as a sub-sample of the study (e.g. of the same ethnic origin).

To account for the extra variance introduced by using $\hat{\mathbf{S}}$ instead of $\mathbf{S}$, we adjust the variance for the estimates by the Delta method

$$V(\hat{\boldsymbol{\sigma}}^2) = \hat{\mathbf{S}}^{-1}(V(\mathbf{q}) + \mathbf{D}(\hat{\boldsymbol{\sigma}}^2)V(\hat{\mathbf{S}})\mathbf{D}(\hat{\boldsymbol{\sigma}}^2))\hat{\mathbf{S}}^{-1}, \tag{20}$$

$$V(\hat{\sigma}^2_{k+1}) = \mathbf{1}^T\hat{\mathbf{S}}^{-1}(V(\mathbf{q}) + \mathbf{D}(\hat{\boldsymbol{\sigma}}^2)V(\hat{\mathbf{S}})\mathbf{D}(\hat{\boldsymbol{\sigma}}^2))\hat{\mathbf{S}}^{-1}\mathbf{1}, \tag{21}$$

where $\mathbf{D}(\hat{\boldsymbol{\sigma}}^2)$ denotes a k by k diagonal matrix with ith diagonal element $\hat{\sigma}^2_i$. We compute $V(\hat{\mathbf{S}})$ via Jackknife [80]. Specifically, we first compute $\mathbf{K}'_i$, $\tilde{\mathbf{A}}'_i$ and $\hat{\mathbf{S}}$. Then, we remove one individual at a time and use the corresponding sub-matrix $\mathbf{K}'_{m-1}$ and $\tilde{\mathbf{A}}'_{m-1}$ to compute $\hat{\mathbf{S}}_{m-1}$. Finally, we estimate the element-wise variance of $\hat{\mathbf{S}}$ as $V(\hat{\mathbf{S}})$. Because the second step of computing $\hat{\mathbf{S}}_{m-1}$ re-uses many quantities that are available from computing $\hat{\mathbf{S}}$, the overall complexity of estimating $V(\hat{\mathbf{S}})$ is only $O(m^3)$.

Our reasoning behind estimating $\mathbf{S}$ stems from the following alternative representation of the MQS solution

$$\mathbf{q}_i = \frac{1}{p_i}\sum_{l=1}^{p_i} w_{il}r^2_{il,y} - \frac{s^2_y}{n-1}, \tag{22}$$

$$\mathbf{S}_{ij} = \frac{1}{p_ip_j}\sum_{l=1}^{p_i}\sum_{l'=1}^{p_j} w_{il}r^2_{il,jl'} - \frac{1}{n-1}, \tag{23}$$

where $r_{il,y} = \mathbf{x}^T_{jl}\mathbf{y}/(n-1)$ is the correlation between phenotype and the genotype of lth SNP in ith category, and $r_{il,jl'} = \mathbf{x}^T_{il}\mathbf{x}_{jl'}/(n-1)$ is the genotype correlation between lth SNP in ith category and

l'th SNP in jth category. Intuitively, each element of $\mathbf{S}$ is a weighted average of $p_i p_j$ different terms of $\hat{r}^2_{il,jl'}$ while each element of $\mathbf{q}$ is a weighted average of only p_i different terms of $\hat{\beta}^2_{il}$. Therefore, $\mathbf{S}$ is much easier to be estimated accurately than $\mathbf{q}$.

Besides the intuitive explanation, we also provide two formal arguments to support the sub-sampling approach for MQS. We use MQS-HEW for illustration. Our first argument is that the variance of $\mathbf{S}$ estimated by using m individuals, or $V(\mathbf{S})$, is often small compared with $\mathbf{S}$ itself. Because of this, the Delta method approximation will be accurate and $\mathbf{S}^{-1}$ can be estimated well by using $\hat{\mathbf{S}}^{-1}$. Specifically, when $k = 1$, for independent SNPs, S is expected to be $1/p$ while $\sqrt{V(\hat{S})}$ is expected to be $2/(mp)$ under the null, where p is the number of SNPs (Supplementary Text). Simulations with independently and binomially distributed SNPs confirm these relationship (Figure S1 and Table S1). For SNPs with LD, we can use the effective number of independent SNPs in place of p to provide approximate forms $E(S) \approx 1/p'$ and $\sqrt{E(V(\hat{S}))} \approx 2/(mp')$. Simulations with real data confirm these approximate relationship (Figures S2, S3 and Table S1).

Our second argument is that $V(\hat{\mathbf{S}})$ is also small compared with $V(\mathbf{q})$. Because of this, the variance of $\hat{\boldsymbol{\sigma}}^2$ and $\hat{\sigma}^2_{k+1}$ is dominated by $V(\mathbf{q})$ and estimating $\mathbf{S}$ does not introduce much extra variance. Specifically, when $k = 1$, for independent SNPs, $\sqrt{E(V(q))} = \sqrt{2}/(n\sqrt{p})$ under the null (Supplementary Text), which is $\sqrt{2p}m/n$ times larger than $\sqrt{V(\hat{S})}$. Simulations with independently and binomially distributed SNPs confirm this relationship (Figure S1). For SNPs with LD, we use the effective number of independent SNPs in place of p to provide an approximate form $\sqrt{E(V(q))} \approx \sqrt{2}/(n\sqrt{p'})$. Simulations with real data again confirm the approximate form (Figures S2, S3 and Table S1).

The two arguments above represent the first attempt to understand the behavior of sub-sampling strategy and the reference panel idea that has been widely used in genetics [45,65–69]. However, we note that the two arguments are by no means complete. For example, we have focused on $k = 1$ here. For $k > 1$, the number of variance components k as well as the effective number of independent SNPs in each category will both play a role in determining the size of m. A larger m is likely required to ensure accurate estimation of $\mathbf{S}^{-1}$ as well as a small $V(\hat{\mathbf{S}})$ with respect to $V(\mathbf{q})$. In addition, we have focused only on MQS-HEW. MQS-LDW uses an iterative procedure that makes it harder for a thorough investigation. However, we will provide simulations as well as real data applications to show that estimating $\mathbf{S}$ does work well for either $k = 1$ or $k > 1$ and for both MQS-HEW and MQS-LDW.

Estimation with Summary Statistics

We are now ready to describe the details of MQS estimation when we only have summary statistics. In this case, we require marginal z-scores from the complete data and individual-level genotypes from a subset of individuals (or from a separate reference panel whose individuals can be thought of as a subset of the study sample). When we only have summary statistics and when the phenotype variance s^2_y is unknown, $\boldsymbol{\sigma}^2$ and σ^2_{k+1} are not estimable but their scaled versions $\mathbf{h}^2$ and h^2_{k+1} are.

To estimate $\mathbf{h}^2$ and h^2_{k+1}, we first use the marginal z-scores to approximate $\mathbf{q}$

$$\mathbf{q}_i/s^2_y \approx \frac{1}{n-1}\frac{1}{p_i}\sum_{l=1}^{p_i} w_{il}(\hat{z}^2_{il} - 1). \tag{24}$$

The approximate assumes that the marginal variant effect size is small, which holds well for most GWASs. Next, we use the individual-level genotype data from a sub-sample of data or a separate reference panel to compute $\hat{\mathbf{S}}$ (equation 19). With $\mathbf{q}/s^2_y$ and $\hat{\mathbf{S}}$, we estimate $\mathbf{h}^2$ and h^2_{k+1} with equations 11 and 12.

To compute the variance for the estimates, we consider two alternative strategies. The first strategy, which we refer to as CI1, is accurate but requires summary statistics in addition to the marginal z scores.

This strategy is based on an alternative expression of the equation 18

$$\begin{aligned} V(\mathbf{q})_{ij}/s_y^2 \approx 2\sum_{l=1}^{k} \hat{h}_l^2 (&\mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i^T\mathbf{X}_l\mathbf{X}_l^T\mathbf{X}_j\mathbf{W}_j\mathbf{z}_j/(p_ip_jp_l) - \mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i^T\mathbf{X}_l\mathbf{z}_l/(p_ip_l) \\ &- \mathbf{z}_l^T\mathbf{X}_l\mathbf{X}_j\mathbf{W}_j\mathbf{z}_j/(p_jp_l) + \mathbf{z}_l^T\mathbf{z}_l/p_l)/(n-1)^3 \\ &+ 2\hat{h}_{k+1}^2(\mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i^T\mathbf{X}_j\mathbf{W}_j\mathbf{z}_j/p_ip_j - \mathbf{z}_i^T\mathbf{W}_i\mathbf{z}_i/p_i - \mathbf{z}_j^T\mathbf{W}_j\mathbf{z}_j/p_j + 1)/(n-1)^3 \end{aligned} \quad (25)$$

where $\mathbf{z}_i$ is a p_i vector of marginal z-scores for SNPs in ith category. Thus, if we can compute additional summary statistics – specifically the n by k matrix of $\mathbf{X}_i\mathbf{z}_i$, the n by k matrix of $\mathbf{X}_i\mathbf{W}_i\mathbf{z}_i$, and the p by k matrix of $\mathbf{X}^T\mathbf{X}_i\mathbf{W}_i\mathbf{z}_i$ – then we can compute the standard errors. These additional summary statistics are easy to compute; in consortium studies, they can be computed within each sub-study and then combined across studies without sharing individual level data. Importantly, for MQS-HEW, we only need to compute two of these three matrices, $\mathbf{X}_i\mathbf{z}_i$ and $\mathbf{X}^T\mathbf{X}_i\mathbf{z}_i$. These two matrices do not require $\mathbf{W}$, which is a function of variance components in MQS-LDW. Thus, MQS-HEW can be much more convenient than MQS-LDW. Finally, we note that, although it is tempting to use quantities computed from a subset of individuals to estimate the confidence interval, we find that $\mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i'^T\mathbf{X}_l'\mathbf{X}_l'^T\mathbf{X}_j'\mathbf{W}_j\mathbf{z}_j/((m-1)^2)$ from m individuals is not a good estimate of $\mathbf{z}_i^T\mathbf{W}_i\mathbf{X}_i^T\mathbf{X}_l\mathbf{X}_l^T\mathbf{X}_j\mathbf{W}_j\mathbf{z}_j/((n-1)^2)$.

The second strategy, which we refer to as CI2, does not require extra summary statistics. It is based on the strategy used in LDSC [45]. It works well when SNPs are approximately independent but can work poorly otherwise. This second strategy is based on the observation that the covariance function $V(\mathbf{q})$ as a function of $\mathbf{y}$ can be written as a function of the marginal z-scores $\mathbf{z} = (\mathbf{z}_1, \cdots, \mathbf{z}_k)^T$. $\mathbf{z}_i \approx \mathbf{X}_i^T\mathbf{y}/(s_y\sqrt{n-1})$ follows approximately a degenerate multivariate normal distribution $\mathbf{z}_i \sim \text{MVN}(0, \mathbf{X}_i^T\mathbf{H}\mathbf{X}_i/(s_y^2(n-1)))$. When SNPs are independent (i.e. $\mathbf{X}_i^T\mathbf{H}\mathbf{X}_i$ is diagonal) or when $\mathbf{X}_i^T\mathbf{H}\mathbf{X}_i$ is block-diagonal, we can use block-wise permutation of z-scores to estimate $V(\mathbf{q})$ as in LDSC [45]. However, as we show in the Results, when LD pattern is complicated, CI2, like LDSC, can yield untrustworthy confidence intervals.

Overall, we recommend the use of CI1. However, we recognize that CI1 requires additional summary statistics that may not be readily available from many studies at the moment. Thus, we have also implemented CI2 as a useful practical option.

Acknowledgment

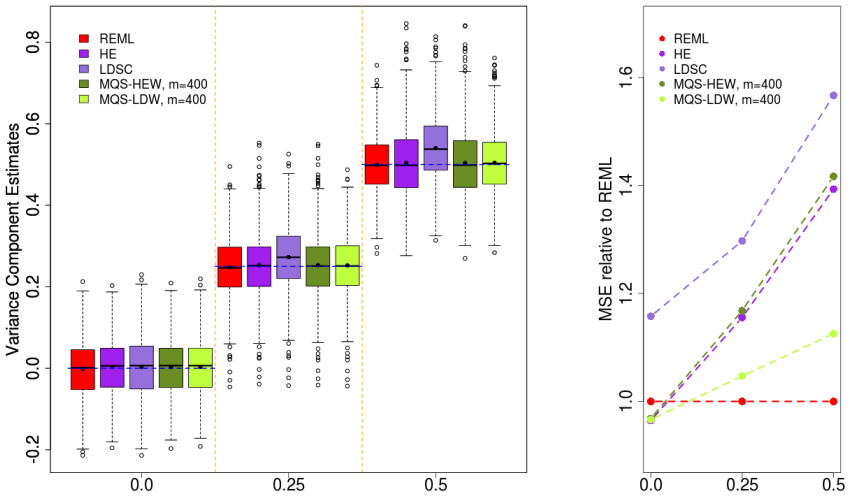
This research is supported by start up funds from the University of Michigan to XZ. We thank Nick Martin and the Queensland Institute of Medical Research for making the Australia height data available to us. We thank Joseph K. Pickrell for making the summary data from 8 quantitative human traits available to us. We thank the NFBC1966 Study Investigators for making the Finland NFBC1966 data available to us. The NFBC1966 study is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with the Broad Institute, University of California Los Angeles, University of Oulu, and the National Institute for Health and Welfare in Finland. This manuscript was not prepared in collaboration with investigators of the NFBC1966 study and does not necessarily reflect their views or those of their host institutions. This study also makes use of data generated by the Wellcome Trust Case Control Consortium (WTCCC). A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the WTCCC project was provided by the Wellcome Trust under award 076113 and 085475. We thank Chaolong Wang, William Wen, Ping Zeng, and Xiang Zhu for helpful comments on a previous version of the manuscript.

Table 1. Features and computational complexity of different methods for variance component estimation. The computational complexity includes time to compute the genetic relatedness matrices. n is the number of individuals, m is the number of randomly selected subset of individuals ($m < n$), k is the number of variance components, p is the number of genetic markers. AI: average information method. HE: Haseman-Elston regression. LDSC: LD score regression. w is the average number of variants used to estimate the LD scores. c is the number of blocks used for the block-wise Jackknife re-sampling algorithm. t is the number of iterations used for estimation: $t = 1$ for MQS-HEW, $t = 2$ (or $t > 2$) for MQS-LDW and LDSC. The confidence interval in MQS can be computed in three different ways: CI requires computed genetic relatedness matrices; CI1 requires summary statistics in addition to z-scores; CI2 follows LDSC, requires only z-scores but is an approximation.

Type	Methods	Computational Complexity		Use Summary Statistics	Unbiased	Calibrated Confidence Interval
		Point Estimate	Confidence Interval			
REML	AI	$pn^2 + t(n^3 + k^2n^2)$	k^2n^2	No	Yes	Yes
MoM	HE	$pn^2 + k^2n^2$	k^2n^3	No	Yes	Yes
	LDSC	$pn + wpm + tkp$	cp	Yes	No	No
	MQS	$pn + t(pm^2 + k^2m^2)$	CI: k^2n^2; CI1: kpn; CI2: cp	Yes	Yes	Yes (CI/CI1), No(CI2)

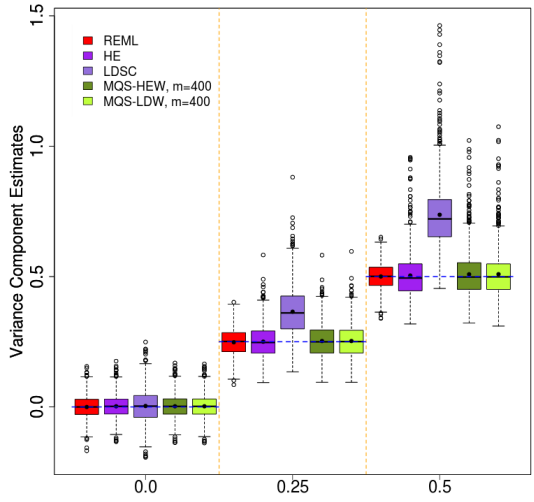
Table 2. Chip heritability estimates from different methods for 11 quantitative traits and 7 binary phenotypes from three GWASs. Values in parentheses are standard errors. For MQS-HEW and MQS-LDW, the standard errors are computed by CI1. The standard errors computed by CI2 are available in a supplementary table. The heritability estimates for the WTCCC data is presented at the observed scale. A small scaling factor is required to transform the estimates to liability scale.

Trait	Methods					
	REML	HE	LDSC		MQS ($m = 400$; CI1)	
			1 MB	10 MB	HEW	LDW
Australia, $n = 3,925$, $p = 4,352,968$						
Height	0.27(0.072)	0.25(0.072)	0.30 (0.080)	0.29 (0.076)	0.26(0.072)	0.28(0.072)
Finland, $n = 5,123$, $p = 319,148$						
CRP	0.043(0.049)	0.034(0.045)	0.054(0.066)	0.046(0.056)	0.037(0.045)	0.037(0.045)
Glucose	0.17(0.046)	0.21(0.069)	0.30 (0.067)	0.25 (0.057)	0.21(0.070)	0.21(0.068)
Insulin	-0.063(0.037)	-0.084(0.043)	-0.12 (0.057)	-0.1 (0.0476)	-0.082(0.044)	-0.081(0.044)
TC	0.29(0.043)	0.42(0.097)	0.63 (0.079)	0.54 (0.068)	0.42(0.098)	0.43(0.099)
HDL	0.34(0.043)	0.36(0.071)	0.50 (0.081)	0.42 (0.067)	0.36(0.072)	0.35(0.068)
LDL	0.38(0.041)	0.60(0.13)	0.91 (0.080)	0.77 (0.069)	0.61(0.13)	0.61(0.13)
TG	0.19(0.047)	0.17(0.053)	0.25 (0.070)	0.21 (0.059)	0.18(0.053)	0.17(0.053)
BMI	0.20(0.047)	0.18(0.053)	0.28 (0.063)	0.24 (0.054)	0.19(0.053)	0.19(0.054)
SysBP	0.19(0.045)	0.27(0.087)	0.41 (0.064)	0.35 (0.055)	0.27(0.088)	0.28(0.090)
DiaBP	0.071(0.046)	0.073(0.049)	0.11 (0.063)	0.093 (0.053)	0.075(0.049)	0.076(0.049)
WTCCC, $n = \sim 5,000$, $p = 458,868$						
BD	0.83(0.057)	1.05(0.095)	1.32(0.085)	1.31(0.084)	1.08(0.098)	1.16(0.10)
CAD	0.57(0.061)	0.58(0.071)	0.72(0.078)	0.71(0.078)	0.60(0.073)	0.63(0.074)
CD	0.71(0.060)	1.06(0.15)	1.32(0.11)	1.30(0.10)	1.08(0.16)	1.12(0.17)
HT	0.58(0.060)	0.62(0.072)	0.78(0.077)	0.77(0.076)	0.63(0.072)	0.66(0.073)
RA	0.69(0.059)	0.81(0.083)	0.95(0.32)	0.90(0.29)	0.84(0.085)	0.83(0.083)
T1D	0.97(0.052)	1.53(0.15)	1.65(0.74)	1.48(0.62)	1.60(0.15)	1.41(0.11)
T2D	0.61(0.060)	0.69(0.082)	0.87(0.080)	0.86(0.078)	0.71(0.084)	0.76(0.084)

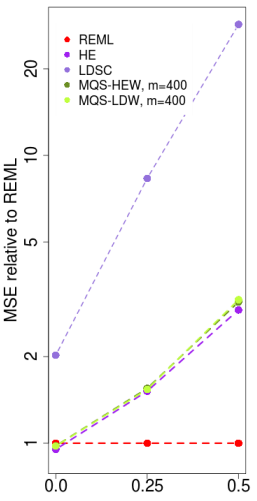


(a) Estimates (Australia)

(b) MSE (Australia)

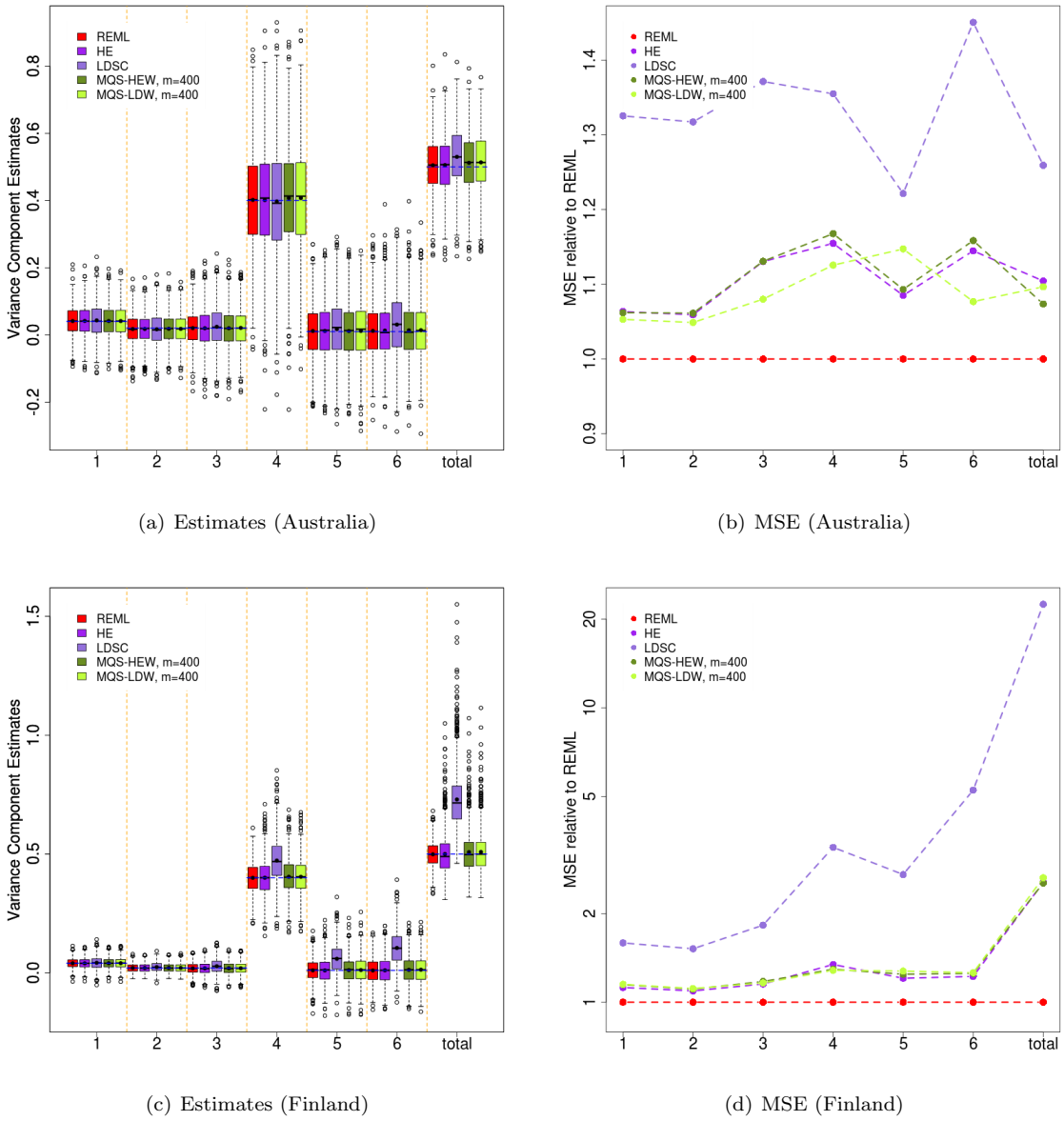


(c) Estimates (Finland)



(d) MSE (Finland)

Figure 1. Comparison of variance component estimates from REML (red), HE (purple), LDSC (light purple), MQS-HEW (green), and MQS-LDW (light green) for $k = 1$ simulations based on the Australian data (a, b) or the Finland data (c, d). (a) and (c) show boxplots while (b) and (d) show the MSE relative to REML. The true variance components from the three simulations (0, 0.25 and 0.5) are shown on the x-axis of all panels, and are also shown as three blue horizontal dashed lines in (a) and (c). The dot in the middle of the boxplots represents the mean of the estimates in (a) and (c). Note that the y-axis in (d) is on log scale.



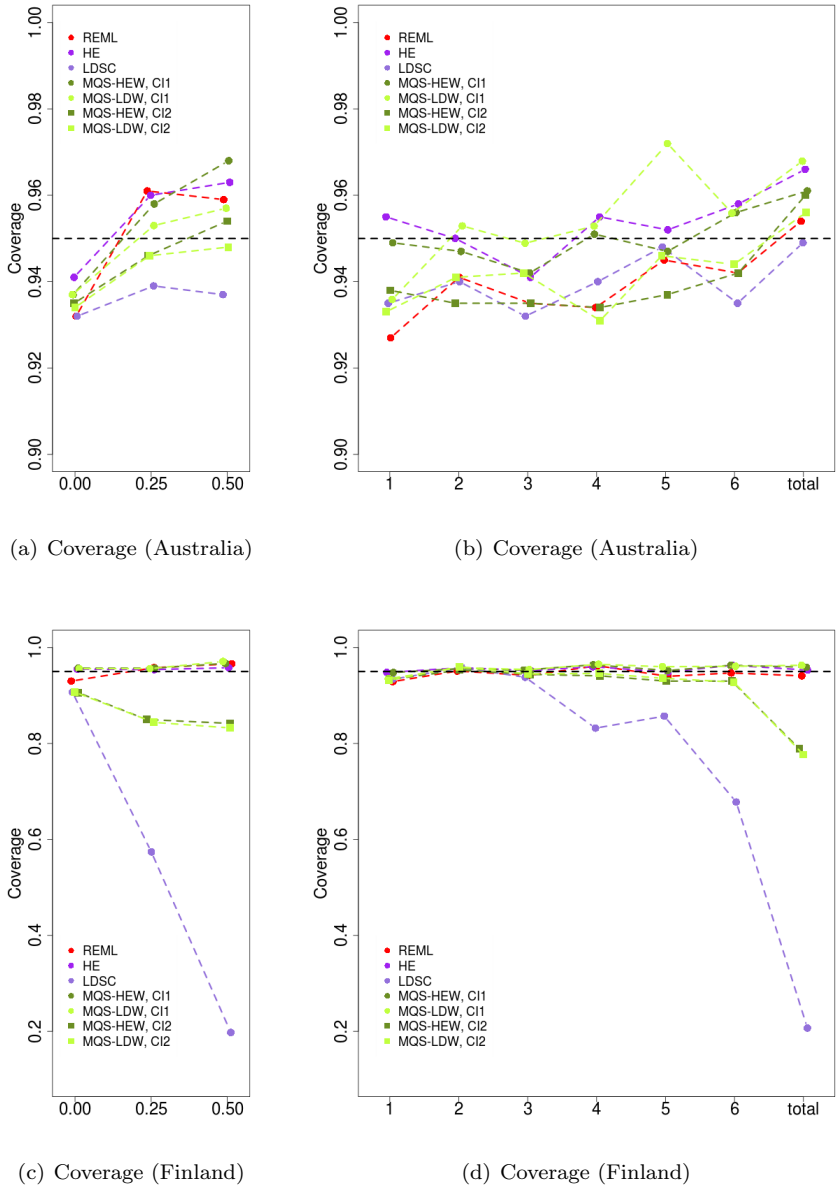
(a) Estimates (Australia)

(b) MSE (Australia)

(c) Estimates (Finland)

(d) MSE (Finland)

Figure 2. Comparison of variance component estimates from REML (red), HE (purple), LDSC (light purple), MQS-HEW (green), and MQS-LDW (light green) for $k = 6$ simulations based on the Australian data (a, b) or the Finland data (c, d) in scenario III of the $k = 6$ simulations. (a) and (c) show boxplots while (b) and (d) show the MSE relative to REML for all six variance components (1-6) as well as the total. The true variance components are shown as three blue horizontal dashed lines in (a) and (c). The dot in the middle of the boxplots represents the mean of the estimates in (a) and (c). Note that the y-axis in (d) is on log scale.



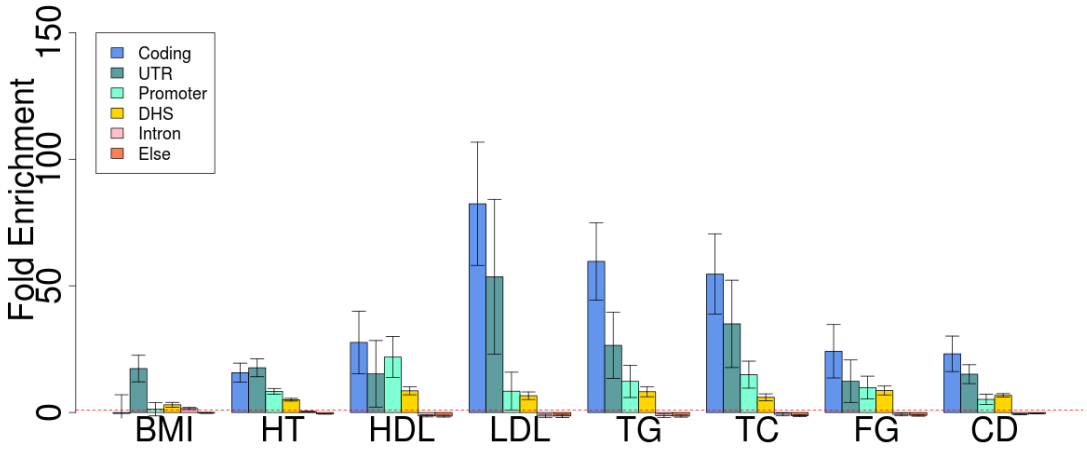
(a) Coverage (Australia)

(b) Coverage (Australia)

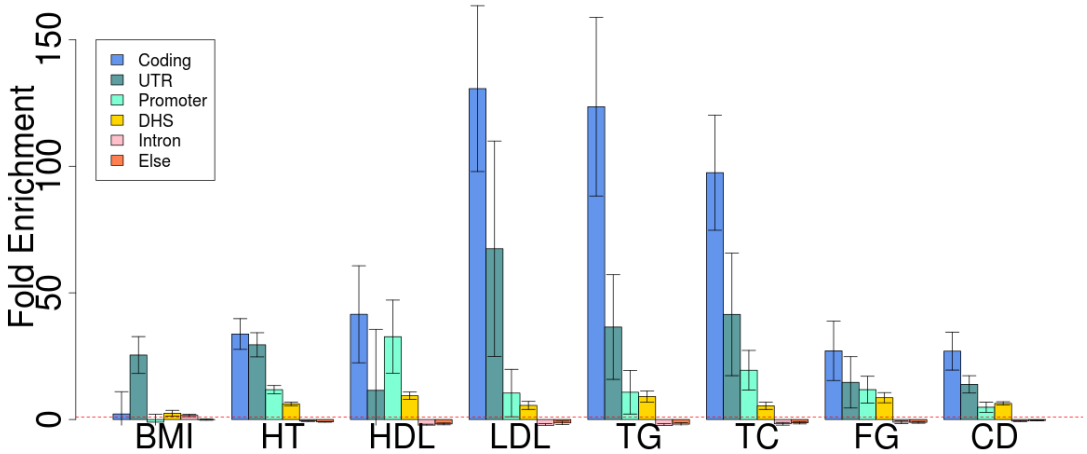
(c) Coverage (Finland)

(d) Coverage (Finland)

Figure 3. Comparison of the coverage probability of confidence intervals from REML (red), HE with CI (purple), LDSC (light purple), MQS-HEW (green), and MQS-LDW (light green) for simulations based on the Australian data (a, b) or the Finland data (c, d). Two different methods are used for MQS-HEW and MQS-LDW to compute the confidence intervals: CI1 (circle) and CI2 (square). The coverage probability is computed based on a 95% confidence interval. (a) and (c) show the coverage probability for the $k = 1$ simulations, where the true variance components (0, 0.25 and 0.5) are shown on the x-axis. (b) and (d) show the coverage probability for six variance components as well as the total heritability in the $k = 6$ simulations with scenario III. Notice that the scale of y-axis in (a) and (b) is different from that in (c) and (d).



(a) MQS-HEW



(b) MQS-LDW

Figure 4. MQS-HEW (a) and MQS-LDW (b) reveal the importance of six functional categories in 8 phenotypes from four GWAS data sets. y-axis shows the fold enrichment, computed as a ratio between the average SNP effect size in one category and the average SNP effect size across the whole genome. Both MQS-HEW and MQS-LDW use marginal z-scores together with genotypes of 503 individuals with European ancestry from the 1,000 genomes project. CI2 is used to construct the confidence intervals.

References

1. Amos CI (1994) Robust variance-components approach for assessing genetic linkage in pedigrees. American Journal of Human Genetics 54: 535-543.

2. Almasy L, Blangero J (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. American Journal of Human Genetics 62: 1198-1211.

3. Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. American Journal of Human Genetics 66: 279-292.

4. Diao G, Lin DY (2005) A powerful and robust method for mapping quantitative trait loci in general pedigrees. American Journal of Human Genetics 77: 97-111.

5. Visscher PM, Hill WG, Wray NR (2008) Heritability in the genomics era – concepts and misconceptions. Nature Reviews Genetics 9: 255-266.

6. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38: 203-208.

7. Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, et al. (2008) Efficient control of population structure in model organism association mapping. Genetics 178: 1709-1723.

8. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nature Genetics 42: 348-354.

9. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, et al. (2010) Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42: 355-360.

10. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, et al. (2011) FaST linear mixed models for genome-wide association studies. Nature Methods 8: 833-835.

11. Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44: 821-824.

12. Pirinen M, Donnelly P, Spencer CCA (2013) Efficient computation with a linear mixed model on large-scale data sets with applications to genetic studies. Annals of Applied Statistics 7: 369-390.

13. Zhou X, Stephens M (2014) Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature Methods : 407-409.

14. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL (2014) Advantages and pitfalls in the application of mixed-model association methods. Nature Genetics 46: 100-106.

15. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, et al. (2015) Efficient Bayesian mixed model analysis increases association power in large cohorts. Nature Genetics 47: 284-290.

16. Robinson GK (1991) That BLUP is a good thing: The estimation of random effects. Statistical Science 6: 15-32.

17. Hofer A (1998) Variance component estimation in animal breeding: a review. Journal of Animal Breeding and Genetics 115: 247-265.

18. Whittaker JC, Thompson R, Denham M (2000) Marker-assisted selection using ridge regression. Genetical Research 75: 249-252.

19. Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. Genetics Research 91: 47-60.

20. Makowsky R, Pajewski NM, Klimentidis YC, Vazquez AI, Duarte CW, et al. (2011) Beyond missing heritability: Prediction of complex traits. PLoS Genetics 7: e1002051.

21. Zhou X, Carbonetto P, Stephens M (2013) Polygenic modelling with Bayesian sparse linear mixed models. PLoS Genetics 9: e1003264.

22. Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, et al. (2013) Pitfalls of predicting complex traits from SNPs. Nature Review Genetics 14: 507-515.

23. Wu TT, Chen YF, Hastie T, Sobel E, Lange K (2009) Genome-wide association analysis by lasso penalized logistic regression. Bioinformatics 25: 714-721.

24. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. Nature Genetics 42: 565-569.

25. Speed D, Hemani G, Johnson MR, Balding DJ (2012) Improved heritability estimation from genome-wide SNPs. American Journal of Human Genetics 91: 1011-1021.

26. de los Campos G, Sorensen D, Gianola D (2015) Genomic heritability: What is it? PLOS Genetics 11: e1005048.

27. Yang J, Weedon MN, Purcell S, Lettre G, Estrada K, et al. (2011) Genomic inflation factors under polygenic inheritance. European Journal of Human Genetics 19: 807-812.

28. Kostem E, Eskin E (2013) Improving the accuracy and efficiency of partitioning heritability into the contributions of genomic regions. American Journal of Human Genetics 92: 558-564.

29. Gusev A, Lee SH, Trynka G, Finucane H, Vilhjlmsson BJ, et al. (2014) Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. American Journal of Human Genetics 5: 535-552.

30. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, et al. (2015) Partitioning heritability by functional category using GWAS summary statistics. Nature Genetics 47: 1228-1235.

31. Thompson EA, Shaw RG (1990) Pedigree analysis for quantitative traits: variance components without matrix inversion. Biometrics 46: 399-413.

32. Chen GB (2014) Estimating heritability of complex traits from genome-wide association studies using IBS-based HasemanElston regression. Frontiers in Genetics 5: 107.

33. Golan D, Lander ES, Rosseta S (2014) Measuring missing heritability: Inferring the contribution of common variants. Proceedings of the National Academy of Sciences, USA 111: E5272E5281.

34. Rao CR (1970) Estimation of heteroscedastic variances in linear models. Journal of the American Statistical Association 65: 161-172.

35. Rao CR (1971) Estimation of variance and covariance component – MINQUE theory. Journal of Multivariate Analysis 1: 257-275.

36. Brown KG (1976) Asymptotic behavior of minque-type estimators of variance components. Annals of Statistics 4: 746-754.

37. Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American Statistical Association 72: 320-338.

38. Zhu J, Weir B (1996) Mixed model approaches for diallele analysis based on a bio-model. Genetical Research 68: 233-240.

39. Haseman JK, Elston RC (1972) The investigation of linkage between a quantitative trait and a marker locus. Behavior Genetics 2: 3-19.

40. Drigalenko E (1998) How sib-pairs reveal linkage. American Journal of Human Genetics 63: 1242-1245.

41. Elston RC, Buxbaum S, Jacobs KB, Olson JM (2000) Haseman and Elston revisited. Genetic Epidemiology 19: 1-17.

42. Sham PC, Purcell S (2001) Equivalence between Haseman-Elston and variance-components linkage analyses for sib pairs. American Journal of Human Genetics 68: 1527-1532.

43. Sham PC, Purcell S, Cherny SS, Abecasis GR (2002) Powerful regression-based quantitative-trait linkage analysis of general pedigrees. American Journal of Human Genetics 71: 238-253.

44. Chen WM, Broman KW, Liang KY (2004) Quantitative trait linkage analysis by generalized estimating equations: unification of variance components and Haseman-Elston regression. Genetic Epidemiology 26: 265-272.

45. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, et al. (2015) LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics 47: 291-295.

46. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, et al. (2015) An atlas of genetic correlations across human diseases and traits. Nature Genetics 47: 1236-1241.

47. Consortium TGP (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56-65.

48. Hayes MG, del Bosque-Plata L, Tsuchiya T, Hanis CL, Bell GI, et al. (2005) Patterns of linkage disequilibrium in the type 2 diabetes gene calpain-10. Diabetes 54: 3573-3576.

49. Zaykin DV, Meng Z, Ehm MG (2006) Contrasting linkage-disequilibrium patterns between cases and controls as a novel association-mapping method. American Journal of Human Genetics 78: 737-746.

50. Price AL, Weale ME, Patterson N, Myers SR, Need AC, et al. (2008) Long-range LD can confound genome scans in admixed populations. American Journal of Human Genetics 1: 132-135.

51. Speed D, Balding DJ (2015) Relatedness in the post-genomic era: is it still useful? Nature Review Genetics 16: 33-33.

52. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, et al. (2008) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. Nature Genetics 41: 35-46.

53. Furlotte NA, Heckerman D, Lippert C (2014) Quantifying the uncertainty in heritability. Journal of Human Genetics 59: 269-275.

54. The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678.

55. Lee SH, Wray NR, Goddard ME, Visscher PM (2011) Estimating missing heritability for disease from genome-wide association studies. The American Journal of Human Genetics 88: 294-305.

56. Browning SR, Browning BL (2013) Identity-by-descent-based heritability analysis in the Northern Finland Birth Cohort. Human Genetics 132: 129-138.

57. Gusev A, Bhatia G, Zaitlen N, Vilhjalmsson BJ, Diogo D, et al. (2013) Quantifying missing heritability at known GWAS loci. PLoS Genetics 9: e1003993.

58. Allen HL, Estrada K, Lettre G, Berndt SI, Weedon MN, et al. (2010) Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832-838.

59. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, et al. (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nature Genetics 42: 937-948.

60. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, et al. (2010) Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466: 707-713.

61. Manning AK, Hivert MF, Scott RA, Grimsby JL, Bouatia-Naji N, et al. (2012) A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nature Genetics 44: 659-669.

62. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, et al. (2012) Hostmicrobe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491: 119-124.

63. Pickrell JK (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. American Journal of Human Genetics 94: 559-573.

64. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, et al. (2015) Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nature Genetics 47: 1114-1120.

65. Browning SR (2006) Multilocus association mapping using variable-length Markov chains. American Journal of Human Genetics 78: 903-913.

66. Guan Y, Stephens M (2008) Practical issues in imputation-based association mapping. PLoS Genetics 4: e1000279.

67. Wen X, Stephens M (2010) Using linear predictors to impute allele frequencies from summary or pooled genotype data. Annals of Applied Statistics 4: 1158-1182.

68. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR (2012) Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nature Genetics 44: 955-959.

69. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, et al. (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nature Genetics 44: 369-375.

70. Robbins H, Monro S (1951) A stochastic approximation method. Annals of Mathematical Statistics 22: 400.

71. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics 38: 904-909.

72. McVean G (2009) A genealogical interpretation of principal components analysis. PLoS Genetics 5: e1000686.

73. Rao CR (1971) Minimum variance quadratic unbiased estimation of variance components. Journal of Multivariate Analysis 1: 445-456.

74. Rao CR (1972) Estimation of variance and covariance components in linear models. Journal of the American Statistical Association 67: 112-115.

75. Rao CR (1973) Quadratic estimation of variance components. Biometrics 2: 311-330.

76. Hartley HO, Rao JNK, Lamotte LR (1978) A simple 'synthesis'-based method of variance component estimation. Biometrics 34: 233-242.

77. Swallow WH, Searle SR (1978) Minimum variance quadratic unbiased estimation (MIVQUE) of variance components. Technometrics 20: 265-272.

78. Gilmour AR, Thompson R, Cullis BR (1995) Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51: 1440-1450.

79. Wu MC, Lee S, Cai T, Li Y, Boehnke M, et al. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. American Journal of Human Genetics 89: 82-93.

80. Efron B, Stein C (1981) The Jackknife estimate of variance. Annals of Statistics 3: 586-596.