

Mixed Model with Correction for Case-Control Ascertainment Increases Association Power

Tristan Hayeck^{1,2,*}, Noah A. Zaitlen³, Po-Ru Loh^{2,4}, Bjarni Vilhjalmsen^{2,4}, Samuela Pollack^{2,4}, Alexander Gusev^{2,4}, Jian Yang^{5,6}, Guo-Bo Chen^{5,6}, Michael E. Goddard⁷, Peter M. Visscher^{5,6}, Nick Patterson², Alkes L. Price^{1,2,4*}

¹Department of Biostatistics, Harvard School of Public Health, Boston, MA

²Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA

³Department of Medicine, Lung Biology Center, University of California, San Francisco, CA

⁴Department of Epidemiology, Harvard School of Public Health, Boston, MA

⁵Queensland Brain Institute, University of Queensland, Brisbane, Australia

⁶University of Queensland Diamantina Institute, University of Queensland, Brisbane, Australia

⁷Faculty of Land and Food Resources, University of Melbourne, Melbourne, Australia

*Correspondence should be addressed to T.H. (tjh488@mail.harvard.edu) or A.L.P. (aprice@hsph.harvard.edu).

We introduce a Liability Threshold Mixed Linear Model (LTMLM) association statistic for ascertained case-control studies that increases power vs. existing mixed model methods, with a well-controlled false-positive rate. Recent work has shown that existing mixed model methods suffer a loss in power under case-control ascertainment, but no solution has been proposed. Here, we solve this problem using a chi-square score statistic computed from posterior mean liabilities (PML) under the liability threshold model. Each individual's PML is conditional not only on that individual's case-control status, but also on every individual's case-control status and on the genetic relationship matrix obtained from the data. The PML are estimated using a multivariate Gibbs sampler, with the liability-scale phenotypic covariance matrix based on the genetic relationship matrix (GRM) and a heritability parameter estimated via Haseman-Elston regression on case-control phenotypes followed by transformation to liability scale. In simulations of unrelated individuals, the LTMLM statistic was correctly calibrated and achieved higher power than existing mixed model methods in all scenarios tested, with the magnitude of the improvement depending on sample size and severity of case-control ascertainment. In a WTCCC2 multiple sclerosis data set with >10,000 samples, LTMLM was correctly calibrated and attained a 4.1% improvement ($P=0.007$) in chi-square statistics (vs. existing mixed model methods) at 75 known associated SNPs, consistent with simulations. Larger increases in power are expected at larger sample sizes. In conclusion, an increase in power over existing mixed model methods is available for ascertained case-control studies of diseases with low prevalence.

Introduction

Mixed model association statistics are a widely used approach to correct for population structure and cryptic relatedness in genome-wide association studies (GWAS)¹⁻¹¹. However, recent work shows that existing mixed model association statistics suffer a loss in power relative to standard logistic regression in ascertained case-control studies¹¹. It is widely known that appropriate modeling of case-control ascertainment can produce substantial increases in power for case-control studies with fixed-effect covariates¹²⁻¹⁴, but such increases in power have not yet been achieved with models that include random effects.

We developed an association score statistic based on a liability threshold mixed linear model (LTMLM). The LTMLM statistic relies on the posterior mean liability (PML) of each individual; the PML is calculated using a multivariate Gibbs sampler¹⁵. The PML of each individual is conditional on the genetic relationship matrix (GRM), the case-control status of every individual, and the disease prevalence. Existing methods use a univariate prospective model to compute association statistics, but here we use a multivariate retrospective model.

The LTMLM statistic provides an increase in power in simulations based on either simulated or real genotypes. In a WTCCC2 multiple sclerosis data set with >10,000 samples, LTMLM was correctly calibrated and attains a 4.1% improvement ($P=0.007$) in chi-square statistics (vs. existing mixed model methods) at 75 known associated SNPs, consistent with simulations.

Materials and Methods

Overview of Method

We improve upon standard mixed model methods¹¹ using a retrospective association score statistic (LTMLM) computed from posterior mean liabilities (PML) under the liability

threshold model. The improvement over previous approaches comes from appropriate modeling of case-control ascertainment. We consider all individuals simultaneously, incorporating prevalence information.

Our method consists of three steps. First, the genetic relationship matrix (GRM) is calculated and a corresponding heritability parameter is estimated, modeling the phenotype covariance of all individuals (see *Estimation of Heritability Parameter*). The heritability parameter is estimated using Haseman-Elston (H-E) regression on the observed scale followed by transformation to liability scale. Second, Posterior Mean Liabilities (PML) are estimated using a truncated multivariate normal Gibbs sampler (see *Posterior Mean Liabilities*). The PML of each individual is conditional on that individual's case-control status, on every other individual's case-control status, and on disease prevalence and liability-scale phenotypic covariance. Third, a chi-square (1 d.o.f) association score statistic is computed based on the association between the candidate SNP and the PML (see *LTMLM Association Statistic*).

The toy example in Figure 1 provides an illustration of how genetic relatedness to a disease case can increase an individual's PML. In Figure 1a and 1b, we plot the distribution of liabilities in 10,000 unrelated individuals with random ascertainment and case-control ascertainment (for a disease with prevalence 0.1%), respectively. In Figure 1c and 1d, we plot the same distributions conditional on an individual having genetic relatedness of 0.5 to a disease case, assuming liability-scale heritability of 1.0. In each case, the posterior distribution of liabilities (and hence the PML) is shifted upwards. (The magnitude and direction of this effect would be different for an individual having a genetic relatedness of 0.5 to a control.) Our main focus below is on much lower levels of genetic relatedness (identity-by-state) among many unrelated samples, but the same principles apply.

Estimation of Heritability Parameter

Mixed model association statistics rely on the estimation of a heritability parameter. We note that this heritability parameter, which Kang et al.⁴ referred to as “pseudo-heritability”, is generally lower than the total narrow-sense heritability (h^2) in data sets not dominated by family relatedness, but may be larger than the heritability explained by genotyped SNPs (h_g^2)¹⁶ in data sets with population structure or family relatedness. However, for ease of notation, we use the symbol h^2 to represent this heritability parameter. A list of all notation used below is provided in Table S1.

The goal is to test for association between a candidate SNP and a phenotype. We first consider a quantitative trait:

$$\boldsymbol{\varphi} = \boldsymbol{\beta}\boldsymbol{x} + \boldsymbol{u} + \boldsymbol{e} \quad . \quad (1)$$

The phenotypic data (transformed to have mean 0 and variance 1) may be represented as a vector $\boldsymbol{\varphi}$ with values for each individual i . Genotype values of candidate SNP are transformed to a vector \boldsymbol{x} with mean 0 and variance 1, with effect size $\boldsymbol{\beta}$. The quantitative trait value depends on the fixed effect of the candidate SNP ($\boldsymbol{\beta}\boldsymbol{x}$), the genetic random effect excluding the candidate SNP (\boldsymbol{u}), and the environmental component (\boldsymbol{e}). We extend to case-control traits via the liability threshold model, in which each individual has an underlying, unobserved normally distributed trait called the liability. An individual is a disease case if the liability exceeds a specified threshold t , corresponding to disease prevalence¹⁷ (Figure S1).

Standard mixed model association methods generally estimate h^2 from a genetic relationship matrix (GRM) and phenotypes using restricted maximum likelihood (ReML)^{4; 11}. Genotypic data is used to build a GRM (excluding the candidate SNP¹¹):

$$\hat{\boldsymbol{\theta}} = \frac{\mathbf{X}^T \mathbf{X}}{M}, \quad (2)$$

where \mathbf{X} is a matrix of non-candidate SNPs normalized to mean 0 and variance 1 and M is the number of SNPs. We estimate h^2 using Haseman-Elston (H-E) regression followed by a transformation to liability scale. The H-E regression estimate is obtained by regressing the product of the case-control phenotypes on the off diagonal terms of the GRM¹⁸⁻²⁰:

$$\widehat{h_{\text{HE}}^2} = \frac{\sum_{i \neq k} \pi_i \pi_k \hat{\boldsymbol{\theta}}_{ik}}{\sum_{i \neq k} \hat{\boldsymbol{\theta}}_{ik}^2}, \quad (3)$$

where π_i denotes the case-control status of individual i and $\hat{\boldsymbol{\theta}}_{ik}$ is the genetic relatedness of individuals i and k . This gives an estimate on the observed scale which is then transformed to the liability scale²¹:

$$\widehat{h_{\text{HE},l}^2} = \widehat{h_{\text{HE}}^2} \frac{[K(1-K)]^2}{z^2(P(1-P))}, \quad (4)$$

where z is the height of the standard normal density ($\frac{1}{\sqrt{2\pi}} e^{-t^2/2}$) at the liability threshold t , K is disease prevalence, and P is the proportion of cases in the sample²¹.

Then, the variance between the individuals is modeled as the phenotypic covariance

$$\mathbf{V} = h^2 \hat{\boldsymbol{\theta}} + (I - h^2) \mathbf{I}, \quad (5)$$

where $\hat{\boldsymbol{\theta}}$ is the N by N GRM, \mathbf{V} is the phenotypic covariance, h^2 is the heritability parameter, and \mathbf{I} is the identity matrix.

Using the phenotypic covariance matrix \mathbf{V} , the liability is modeled as a multivariate normal distribution:

$$L(\boldsymbol{\varphi}) = (2\pi)^{\frac{-n}{2}} |\mathbf{V}|^{-1/2} \exp\left(\frac{-1}{2} (\boldsymbol{\varphi})^T (\mathbf{V})^{-1} (\boldsymbol{\varphi})\right) \quad (6)$$

We note that we observe the case-control phenotypes of the individuals and not the continuous liabilities.

Posterior Mean Liabilities

We first consider the univariate PML (PML_{uni}), constructed independently for each individual; we generalize to the multivariate setting below. As described in equations 11 and 12 of ref. ²¹, these correspond to the expected value of the liability conditional on the case control status:

$$\begin{aligned} PML_{uni,case} &= E[\varphi|\pi_i = 1] = z/K \\ PML_{uni,control} &= E[\varphi|\pi_i = 0] = -z/(1 - K) \end{aligned} \quad (7)$$

These values are calculated analytically in the univariate setting, and can be thought of as the mean of a truncated normal above or below the liability threshold t depending on case control status²¹.

We now consider the multivariate PML (PML_{multi}), estimated jointly across individuals. The PML_{multi} for each individual is conditional on that individual's case-control status, on every other individual's case-control status, and on their phenotypic covariance. The PML_{multi} is estimated using a Gibbs sampler, analogous to previous work¹⁵ (which focused on family relatedness and did not consider association statistics). The Gibbs sampler is an iterative algorithm that generates random variables from conditional distributions in order to avoid the difficult task of explicitly calculating the marginal density for each random variable.

For each individual in turn, the conditional distribution of the liability is calculated based on all of the other individuals and a new value is generated. The algorithm is:

$$\begin{aligned} &\text{Initialization: for each individual } j, \\ &\quad \varphi_j = PML_{uni,case} \text{ if } \pi_j=1 \text{ or } \varphi_j = PML_{uni,control} \text{ if } \pi_j=0 \\ &\text{For each MCMC iteration } n \\ &\quad \text{For each individual } i \\ &\quad \quad \text{Sample } \varphi_i \text{ from the constrained conditional univariate normal distribution} \\ &\quad \quad L(\varphi_i) \sim \exp(-\varphi^T V^{-1} \varphi / 2) \text{ and constraint } \varphi_i \geq t \text{ if } \pi_i=1, \varphi_i < t \text{ if } \pi_i=0 \\ &\quad \quad \text{(where } \varphi_{\neq i} \text{ are fixed)} \end{aligned} \quad (8)$$

We use 100 burn-in iterations followed by 1,000 additional MCMC iterations. We estimate the PML_{multi} by averaging over MCMC iterations. We reduce the number of MCMC iterations needed via Rao-Blackwellization, which averages (across iterations n) the *posterior means* of the distributions from which each φ_i is sampled.

LTMLM Association Statistic

The LTMLM association statistic is calculated using PML_{multi} . For simplicity, we first consider the case where the liability is known. We jointly model the liability and the genotypes using a retrospective model, enabling appropriate treatment of sample ascertainment. We concatenate the two vectors (φ, \mathbf{x}) and derive the joint likelihood for these combined terms. The covariance of φ and \mathbf{x} between individual i and k is:

$$Cov(\varphi_i, \mathbf{x}_k) = E[\varphi_i, \mathbf{x}_k] - E[\varphi_i]E[\mathbf{x}_k] = E[\varphi_i, \mathbf{x}_k] = E[\beta \boldsymbol{\theta}_i, \mathbf{x}_k] = \beta \boldsymbol{\theta}_{i,k}, \quad (9)$$

where $\boldsymbol{\theta}$ is the true underlying genetic relatedness matrix from which genotypes are sampled.

(We note that $\boldsymbol{\theta}$, which is unobserved, is different from the GRM $\hat{\boldsymbol{\theta}}$ estimated from the data.)

The variance of (φ, \mathbf{x}) as a function of effect size β is:

$$\mathbf{C}(\beta) = \begin{pmatrix} \mathbf{V} & \beta \boldsymbol{\theta} \\ \beta \boldsymbol{\theta}^T & \boldsymbol{\theta} \end{pmatrix}, \quad (10)$$

thus

$$\mathbf{C}(\beta)^{-1} = \begin{pmatrix} \mathbf{V}^{-1} & -\beta \mathbf{V}^{-1} \\ -\beta (\mathbf{V}^{-1})^T & \boldsymbol{\theta}^{-1} \end{pmatrix} + O(\beta^2), \quad (11)$$

where both of these matrices are $2N$ by $2N$. The joint likelihood of the liability and genotypes are distributed as a multivariate normal $N(0, \mathbf{C}(\beta))$, and thus

$$L(\mathbf{x}, \boldsymbol{\varphi}|\boldsymbol{\beta}) = (2\pi)^{\frac{-n}{2}} |\mathbf{C}(\boldsymbol{\beta})|^{-1/2} \exp\left(\frac{-1}{2} (\boldsymbol{\varphi}, \mathbf{x})^T \mathbf{C}(\boldsymbol{\beta})^{-1} (\boldsymbol{\varphi}, \mathbf{x})\right). \quad (12)$$

Taking the derivative of the log likelihood results in the score equation. The determinant of the matrix \mathbf{V} does not have any terms linear in $\boldsymbol{\beta}$, so the terms with \mathbf{V} alone drop out when we take the derivative:

$$\begin{aligned} S(\mathbf{x}, \boldsymbol{\varphi}|\boldsymbol{\beta}) &= \frac{d}{d\boldsymbol{\beta}} \ln L(\mathbf{x}, \boldsymbol{\varphi}|\boldsymbol{\beta}) = \frac{d}{d\boldsymbol{\beta}} \frac{-1}{2} (\boldsymbol{\varphi}, \mathbf{x})^T \mathbf{C}(\boldsymbol{\beta}) (\boldsymbol{\varphi}, \mathbf{x}) \\ &= \frac{d}{d\boldsymbol{\beta}} (\boldsymbol{\varphi}, \mathbf{x})^T \begin{pmatrix} \mathbf{V}^{-1} & -\boldsymbol{\beta} \mathbf{V}^{-1} \\ -\boldsymbol{\beta} (\mathbf{V}^{-1})^T & \boldsymbol{\Theta}^{-1} \end{pmatrix} (\boldsymbol{\varphi}, \mathbf{x}) = \mathbf{V}^{-1} \boldsymbol{\varphi} \mathbf{x} \end{aligned} \quad (13)$$

The marginal score statistic tests the null hypothesis that the fixed effect of the candidate SNP is zero ($H_0: \boldsymbol{\beta} = 0$) vs. the alternative hypothesis ($H_A: \boldsymbol{\beta} \neq 0$). The denominator of the score statistic is the variance of the score evaluated under the null. :

$$\text{Var}(S(\mathbf{x}, \boldsymbol{\varphi}|\boldsymbol{\beta})) = (\mathbf{V}^{-1} \boldsymbol{\varphi})^T \boldsymbol{\Theta} (\mathbf{V}^{-1} \boldsymbol{\varphi}) \quad (14)$$

This leads to the score statistic:

$$\text{Score statistic} = \frac{(\mathbf{x}^T \mathbf{V}^{-1} \boldsymbol{\varphi})^2}{(\mathbf{V}^{-1} \boldsymbol{\varphi})^T \boldsymbol{\Theta} (\mathbf{V}^{-1} \boldsymbol{\varphi})}, \quad (15)$$

where $\boldsymbol{\Theta}$, the true underlying genetic relatedness of the individuals, can be approximated by the identity matrix in data sets of unrelated individuals.

We now consider a case-control trait, with unobserved liability, and derive the score function using the observed case-control status of each individual, $\boldsymbol{\pi}$. Returning to the score function and conditioning on case control status:

$$S(\mathbf{x}, \boldsymbol{\varphi}|\boldsymbol{\beta}, \boldsymbol{\pi})_{\boldsymbol{\beta}=0} = \frac{d}{d\boldsymbol{\beta}} \ln L(\mathbf{x}, \boldsymbol{\varphi}|\boldsymbol{\beta}, \boldsymbol{\pi})_{\boldsymbol{\beta}=0} = \frac{\frac{dL(\mathbf{x}, \boldsymbol{\varphi}|\boldsymbol{\beta}, \boldsymbol{\pi})_{\boldsymbol{\beta}=0}}{d\boldsymbol{\beta}}}{L(\mathbf{x}, \boldsymbol{\varphi}|\boldsymbol{\beta}, \boldsymbol{\pi})_{\boldsymbol{\beta}=0}} \quad (16)$$

Introducing the unobserved quantitative liability, φ , the score function can be rewritten in terms of the probability density of the liability:

$$\frac{\frac{dL(\mathbf{x}, \boldsymbol{\varphi}|\beta)_{\beta=0}}{d\beta}}{L(\mathbf{x}, \boldsymbol{\varphi}|\beta)_{\beta=0}} = \frac{\int P(\varphi) \frac{dL(\mathbf{x}, \boldsymbol{\varphi}|\beta)_{\beta=0}}{d\beta} d\varphi}{L(\mathbf{x}, \boldsymbol{\varphi}|\beta)_{\beta=0}}$$

$$S(\mathbf{x}, \boldsymbol{\varphi}|\beta, \boldsymbol{\pi}) = C \int P(\varphi) S(\varphi|\boldsymbol{\pi}) d\varphi = S(E[\varphi|\boldsymbol{\pi}]), \quad (17)$$

where $P(\boldsymbol{\varphi})$ is the probability density of the liability and $E[\boldsymbol{\varphi}|\boldsymbol{\pi}]$ is the PML. It follows that an appropriate score statistic is

$$LTMLM \text{ score statistic} = \frac{(\mathbf{x}^T \mathbf{V}^{-1} \mathbf{PML}_{multi})^2}{(\mathbf{V}^{-1} \mathbf{PML}_{multi})^T \boldsymbol{\theta} (\mathbf{V}^{-1} \mathbf{PML}_{multi})} \quad (18)$$

Again $\boldsymbol{\theta}$ can be approximated by the identity matrix in data sets of unrelated individuals; we note that this choice affects only a constant calibration factor (since the denominator is the same for each candidate SNP), and that other calibration options are available (see below). As with other association statistics, the LTMML score statistic generalizes to non-normally distributed genotypes²²⁻²⁴. The overall computational cost of computing the LTMML statistic is $O(MN^2)$ when $M > N$ (Table S2).

We calculate the GRM via Leave One Chromosome Out (LOCO) analysis, i.e. for each candidate SNP on a given chromosome the GRM is calculated using all of the other chromosomes. This prevents deflation due to double counting the candidate SNP as both a fixed effect and random effect in the mixed model^{4; 6; 11}.

Simulated Genotypes and Simulated Phenotypes

We performed simulations both using simulated genotypes and simulated phenotypes, and using real genotypes and simulated phenotypes (see below). Quantitative liabilities for each individual were generated from SNP effects and an environmental component. The proportion of causal SNPs was set to 0.01. The quantitative liabilities were then dichotomized based on the liability threshold to categorize each individual as a case or control. Case-control ascertainment

was performed, simulating 50% cases and 50% controls. We compared ATT, MLM, and LTMMLM statistics (see Table 1). MLM statistics were computed using the GCTA-LOCO statistic described in ref. ¹¹, with the heritability parameters estimated using the GCTA software²⁵. All simulations used M SNPs to calculate the GRM and M additional SNP as candidate SNPs being tested for association (so that candidate SNPs are not included in the GRM).

In the primary analyses, we simulated individuals without population structure or LD, with $N = 1K$ or $5K$ samples, $M = 1K$, $5K$ or $50K$ SNPs, and prevalence $K = 50\%$, 10% , 1% or 0.1% . Genotypes were sampled from independent binomials with allele frequencies uniform on $[0.1, 0.9]$. In secondary analyses, we simulated population structure by simulating two populations with an F_{ST} of 0.01 , whose allele frequencies were drawn from beta distributions with parameters $p(1 - F_{ST})/F_{ST}$ and $(1 - p)(1 - F_{ST})/F_{ST}$, based on ancestral allele frequency p which is uniform on $[0.1, 0.9]$.

WTCCC2 Genotypes and Simulated Phenotypes

We also conducted simulations using real genotypes from WTCCC2 to incorporate LD and realistic population structure. The WTCCC2 data contained 360,557 SNPs and 15,633 samples, as described previously¹¹. Since the goal of the power study is demonstrate a comparison of the statistics under case-control ascertainment, we used $N = 1000$ samples (500 cases and 500 controls), with simulated phenotypes having prevalence of 50%, 25%, 10%. The prevalence was restricted to a lower bound of 10% because of the limitation of only 15,633 WTCCC2 samples for simulating case-control ascertainment. We computed ATT, MLM and LTMMLM statistics as described above.

WTCCC2 Genotypes and MS Phenotypes

Finally, we analyzed WTCCC2 individuals with ascertained case-control phenotypes for MS¹¹, a disease with a prevalence of around 0.1%. We computed ATT, MLM and LTMLM statistics as described above. Although the underlying MS study was appropriately matched for ancestry²⁶, the data made available to researchers included only pan-European cases and UK controls. Thus, the WTCCC2 data set shows a severe mismatch in ancestry of cases and controls; this severe mismatch between cases and controls is not representative of a typical GWAS. We thus restricted our primary analysis to 10,034 samples with only a moderate mismatch in ancestry, but analyses of unmatched and stringently matched data sets were also performed (Figure S2). The unmatched data set contained 10,204 case and 5,429 controls. Matching was performed by first calculating 20 PCs in the full cohort and weighing the contribution of each PC based on the variance in phenotype it explained in a multiple regression. A Euclidean distance over these 20 weighted dimensions was then computed for all pairs of individuals, and each case was greedily assigned the nearest unmatched control until no matched case-control pairs could be identified. Finally, any matched case-control pairs that were not within 6 standard deviations of the mean pairwise distance were removed as outliers, yielding the 5,017 cases and 5,017 matched controls used in our primary analysis. Stringent matching was performed by additionally removing any matched case-control pairs that were not within 2 standard deviations of the mean pairwise distance, yielding 4,094 cases and 4,094 matched controls used in our stringently matched analysis.

We compared association statistics at 75 published SNPs associated to MS¹¹. We used a jackknife approach to assess the statistical significance of differences in association statistics, by excluding each of the 75 published SNPs in turn.

Results

Simulations: Simulated Genotypes and Simulated Phenotypes

We first conducted simulations using simulated genotypes and simulated ascertained case-control phenotypes (see Materials and Methods). Our main simulations involve unrelated individuals with no population structure, but the impact of population structure is explored below. We evaluated the power of ATT, MLM and LTMLM (see Table 2). Results for additional values of #SNPs (M) and #samples (N) are displayed in Table S3. The LTMLM statistic consistently outperforms the ATT and MLM statistics, particularly at low values of disease prevalence. For LTMLM vs. MLM at disease prevalence of 0.1%, 3% and 24% improvements were observed in simulations with 5,000 SNPs and 50,000 SNPs respectively. Smaller improvements were observed at higher disease prevalences. Test statistics were well-calibrated at null markers. Simulations at other values of M and N indicate that the magnitude of the improvement depends on the value of N/M (Table S3). Simulations with population structure demonstrate similar results, but with inflation in the ATT statistic as expected (Table S4).

The MLM statistics were calculated using an h^2 parameter estimated using Restricted Maximum Likelihood Methods (ReML)⁴, but the LTMLM statistics were calculated using an h^2 parameter estimated via Haseman-Elston (H-E) regression on case-control phenotypes followed by transformation to liability scale^{18; 21} (see Materials and Methods). As case-control ascertainment becomes more severe the H-E regression estimate of the h^2 remains unbiased, whereas the variance component estimate is severely downwardly biased even after transformation to the liability scale (Table 3 and Table S5), consistent with previous work (see ref.¹⁹ and Supp Table 9 of ref.¹¹). Population structure resulted in bias of both ReML and HE-regression estimates of h^2 , but consistently higher bias for the ReML estimates (Table S6). These biases do not inflate LTMLM or MLM statistics under the null (Table S4). We note that

previous work has shown that running MLM using the correct h^2 parameter does not ameliorate the loss in power for MLM¹¹.

Simulations: WTCCC2 Genotypes and Simulated Phenotypes

We next conducted simulations using real WTCCC2 genotypes and simulated ascertained case-control phenotypes (see Materials and Methods)^{11, 26}. For a given value of M (M SNPs to calculate the GRM and M candidate SNPs, for a total of $2M$ SNPs), we used the first $M/2$ SNPs from each of the first four chromosomes. The GRM was calculated using SNPs on chromosomes 3 and 4, with SNPs on chromosomes 1 and 2 treated as the candidate SNPs. The simulated phenotypes were generated from chromosome 1 and 3, where 1% of the SNPs were randomly selected as being causal. Results are reported for causal SNPs on chromosome 1 and null SNPs on chromosome 2, which were not used to build the GRM.

Results for 1,000 and 10,000 SNPs (M) are displayed in Table 4 and Table S7, with sample size fixed at 500 cases and 500 controls; formal power calculations produce similar results (Table S8). Once again, the LTMLM statistic outperforms ATT and MLM as case-control ascertainment becomes more severe. (A limitation of these simulations is that performing case-control ascertainment on a fixed set of individuals limits case-control sample size; thus, these simulations were restricted to a disease prevalence of 10% or higher. It is reasonable to infer that for rarer diseases with more extreme case-control ascertainment the LTMLM statistic would achieve even higher power gains, as was demonstrated in simulations with simulated genotypes.)

The h^2 parameter estimates for simulations using real genotypes are displayed in Table S9. The H-E regression estimates are unbiased, but the ReML estimates are again downwardly biased at lower prevalence and large N/M .

WTCCC2 Multiple Sclerosis data set

We analyzed the WTCCC2 genotypes together with multiple sclerosis (MS) case-control phenotypes: 5,172 MS cases and 5,172 controls genotyped on Illumina chips^{11; 26} (see Materials and Methods). We compared ATT, ATT with 5 PCs (PCA)²³, MLM and LTMLM. We evaluated calibration using the average χ^2 over all SNPs; we note that the average χ^2 over all SNPs is expected to be greater than 1 due to polygenic effects^{11; 27}, and all methods can be correctly calibrated via LD Score regression²⁸.

We evaluated power using the average χ^2 over the 75 published SNPs. The results are displayed in Table 5. The LTMLM method performed best, with a 4.1% improvement vs. MLM (jackknife $P=0.007$; see Materials and Methods) and an even larger improvement versus ATT and PCA, consistent with simulations (Table 2). Similar results are obtained when calibrating association statistics via LD Score regression²⁸ (Table S10). A perfectly matched data set with 4,094 MS cases and 4,094 controls yielded a similar improvement for LTMLM vs. MLM (Table S11). We also applied LTMLM to the full unmatched data set of 10,204 MS cases and 5,429 controls, where there is a severe mismatch in ancestry between cases and controls that is not representative of a typical GWAS. The LOCO estimates of h^2 demonstrate inflation before controlling for population structure (Table S12). In this analysis, the H-E regression estimate of the h^2 produces an unrealistic value of 7.3 on the observed scale (corresponding to 2.8 on the liability scale), which is outside the plausible 0-1 range suggesting severe population stratification or other severe problems with the data. We do not recommend the use of LTMLM on unmatched samples when such severe problems are detected. For completeness, we report the results of running LTMLM, which results in a loss in power (Table S11).

Discussion

We have shown that controlling for case-control ascertainment using the LTMLM statistic can lead to significant power improvements in ascertained case-control studies of diseases of low prevalence. This was demonstrated via simulations using both simulated and real genotypes, and in WTCCC2 MS case-control data.

The LTMLM statistic should not be used if the inferred liability-scale h^2 parameter is outside the plausible 0-1 bound, as this is indicative of severe population stratification or other severe problems with the data (this can also be assessed via PCA; see Figure S2). In such settings, either matching based on ancestry should first be performed, or other statistics should be used.

Several limitations of LTMLM remain as directions for future study. First, previous work has shown that using the posterior mean liabilities in conjunction with fixed effects such as BMI, age, or known associated SNPs will further increase power^{12; 21}. The incorporation of fixed-effect covariates into the LTMLM statistic is not considered here, and remains as a future direction. Second, the calibration of our statistic in unrelated samples relies on an approximation that works well in the WTCCC2 data analyzed, but may not work well in all data sets. Here, calibration via LD Score regression offers an appealing alternative²⁸. Third, we did not consider ascertained case-control studies in family data sets, which also represents a future direction. Fourth, LTMLM requires running time $O(MN^2)$, analogous to standard mixed model association methods. This may be computationally intractable in very large data sets. We are developing much faster mixed model methods²⁹, but those methods do not consider case-control ascertainment and should not be applied to ascertained case-control data for diseases of low prevalence. The incorporation of the ideas we have described here into those methods is an

open question. Finally, our methods could potentially be extended to multiple traits^{7; 30; 31}.

Acknowledgements

We thank D. Golan and S. Rosset for helpful discussions. This research was funded by NIH grant R01 HG006399.

This study 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 <http://www.wtccc.org.uk/>.

Web Resources

The URLs for data presented herein are as follows:

1. Liability Threshold Mixed Linear Model (LTMLM) software will be provided at <http://www.hsph.harvard.edu/alkes-price/software/>
2. GCTA (Genome-wide Complex Trait Analysis) software, <http://www.complextaitgenomics.com/software/gcta/>

References

1. Yu, J., Pressoir, G., Briggs, W.H., Vroh Bi, I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* 38, 203-208.
2. Zhao, K., Aranzana, M.J., Kim, S., Lister, C., Shindo, C., Tang, C., Toomajian, C., Zheng, H., Dean, C., Marjoram, P., et al. (2007). An Arabidopsis example of association mapping in structured samples. *PLoS Genet* 3, e4.
3. Kang, H.M., Zaitlen, N.A., Wade, C.M., Kirby, A., Heckerman, D., Daly, M.J., and Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709-1723.
4. Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42, 348-354.
5. Svishcheva, G.R., Axenovich, T.I., Belonogova, N.M., van Duijn, C.M., and Aulchenko, Y.S. (2012). Rapid variance components-based method for whole-genome association analysis. *Nat Genet* 44, 1166-1170.
6. Listgarten, J., Lippert, C., Kadie, C.M., Davidson, R.I., Eskin, E., and Heckerman, D. (2012). Improved linear mixed models for genome-wide association studies. *Nat Methods* 9, 525-526.
7. Korte, A., Vilhjalmsen, B.J., Segura, V., Platt, A., Long, Q., and Nordborg, M. (2012). A

- mixed-model approach for genome-wide association studies of correlated traits in structured populations. *Nat Genet* 44, 1066-1071.
8. Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 44, 821-824.
9. Segura, V., Vilhjalmsen, B.J., Platt, A., Korte, A., Seren, U., Long, Q., and Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* 44, 825-830.
10. Listgarten, J., Lippert, C., and Heckerman, D. (2013). FaST-LMM-Select for addressing confounding from spatial structure and rare variants. *Nat Genet* 45, 470-471.
11. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M., and Price, A.L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* 46, 100-106.
12. Zaitlen, N., Lindstrom, S., Pasaniuc, B., Cornelis, M., Genovese, G., Pollack, S., Barton, A., Bickeboller, H., Bowden, D.W., Eyre, S., et al. (2012). Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genet* 8, e1003032.
13. Zaitlen, N., Pasaniuc, B., Patterson, N., Pollack, S., Voight, B., Groop, L., Altshuler, D., Henderson, B.E., Kolonel, L.N., Le Marchand, L., et al. (2012). Analysis of case-control association studies with known risk variants. *Bioinformatics* 28, 1729-1737.
14. Mefford, J., and Witte, J.S. (2012). The Covariate's Dilemma. *PLoS Genet* 8, e1003096.
15. Campbell, D.D., Sham, P.C., Knight, J., Wickham, H., and Landau, S. (2010). Software for generating liability distributions for pedigrees conditional on their observed disease states and covariates. *Genet Epidemiol* 34, 159-170.
16. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 42, 565-569.
17. Falconer, D.S. (1967). The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann Hum Genet* 31, 1-20.
18. Haseman, J.K., and Elston, R.C. (1972). The investigation of linkage between a quantitative trait and a marker locus. *Behav Genet* 2, 3-19.
19. Golan, D., and Rosset, S. (2013). Narrowing the gap on heritability of common disease by direct estimation in case-control GWAS. *ArXiv e-prints*, <http://arxiv.org/abs/1305.5363>
20. Chen, G.-B. (2014). Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. *Frontiers in Genetics* 5.
21. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 88, 294-305.
22. Thornton, T., and McPeck, M.S. (2007). Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *Am J Hum Genet* 81, 321-337.
23. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38, 904-909.
24. Heyde, C.C. (1997). *Quasi-likelihood and its application: a general approach to optimal parameter estimation.*(Springer).
25. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88, 76-82.
26. Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C.C., Patsopoulos, N.A., Moutsianas, L.,

- Dilthey, A., Su, Z., Freeman, C., Hunt, S.E., et al. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 476, 214-219.
27. Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A.V., Ingelsson, E., O'Connell, J.R., Mangino, M., et al. (2011). Genomic inflation factors under polygenic inheritance. *Eur J Hum Genet* 19, 807-812.
28. Bulik-Sullivan, B., Loh, P.-R., Finucane, H., Ripke, S., Yang, J., Schizophrenia Working Group Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. LD Score Regression Distinguishes Confounding from Polygenicity. <http://biorxiv.org/content/early/2014/02/21/002931short>.
29. Loh, P.-R., Tucker, G., Bulik-Sullivan, B.K., Vilhjalmsdottir, B.J., Finucane, H.K., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., Patterson, N., et al. (2014). Efficient Bayesian mixed model analysis increases association power in large cohorts. <http://biorxiv.org/biorxiv/early/2014/08/09/007799fullpdf>.
30. Lee, S.H., Decandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 44, 247-250.
31. Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods* 11, 407-409.

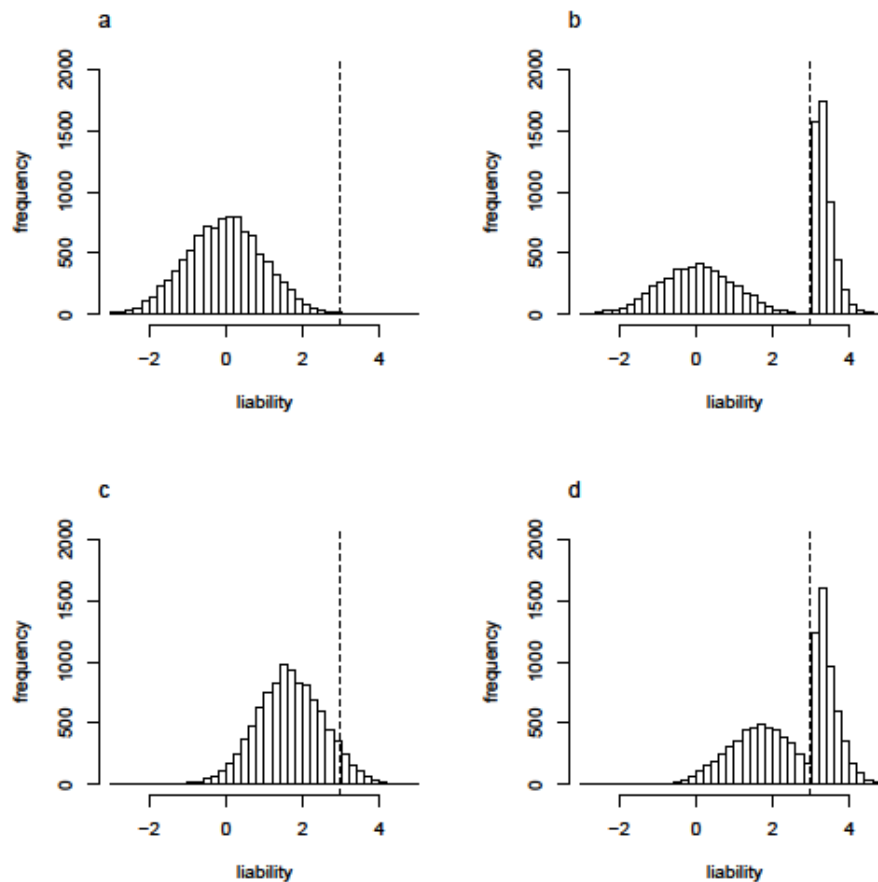


Figure 1. Genetic relatedness to a disease case can increase an individual's PML. In (a) and (b), we plot distributions of liabilities for a set of 10,000 individuals under (a) random ascertainment or (b) case-control ascertainment for a disease with prevalence 0.1% (see Figure 2 of Lee et al.¹⁷). In (c) and (d), we plot the same distributions conditional on an individual having genetic relatedness of 0.5 to a disease case, assuming a heritability of 1 on the liability scale.

	ATT	MLM	LTMLM
Quant vs Case control Trait	Both	Both	CC
Herit Param Est.	None	ReML	H-E regression
Prospective vs. Retrospective	Pro	Pro	Retro
Equation	$\frac{(x^T \pi^*)^2}{x^T x}$	$\frac{(x^T \pi^* V^{-1})^2}{x^T V^{-1} x} = \frac{(x^T V^{-1} PML_{uni})^2}{x^T V^{-1} x}$	$\frac{(x^T V^{-1} PML_{multi})^2}{(V^{-1} PML_{multi})^T I (V^{-1} PML_{multi})}$
Corrects for Confounding?	No	Yes	Yes
Models Case-Control Asc.	No	No	Yes

Table 1. List of association statistics. We list properties of the Armitage Trend Test (ATT), standard mixed model association statistic (MLM), and proposed statistic (LTMLM). π^* is normalized case-control status (mean 0, variance 1), x are normalized genotypes, PML_{uni} is the univariate PML conditional on the case-control status of a single individual, PML_{multi} is the multivariate PML conditional of the case-control status of all individuals, I is the identity matrix, V is the phenotypic covariance (on the observed scale for MLM, and on the liability scale for LTMLM).

M	N	PREVALENCE	SNP SET	ATT	MLM	LTMLM
5000	5000	50%	all	1.14	1.15	1.15
			causal	16.79	17.13	17.12
			null	0.99	0.99	0.99
		10%	all	1.23	1.23	1.24
			causal	24.59	24.79	25.18
			null	0.99	0.99	0.99
		1%	all	1.45	1.40	1.46
			causal	45.72	42.61	46.85
			null	1.00	0.99	1.00
		0.1%	all	1.71	1.51	1.73
			causal	71.99	59.53	74.05
			null	1.00	0.93	1.00
50000	5000	50%	all	1.02	1.02	1.02
			causal	2.68	2.68	2.69
			null	1.00	1.00	1.00
		10%	all	1.02	1.02	1.02
			causal	3.35	3.35	3.35
			null	1.00	1.00	1.00
		1%	all	1.04	1.04	1.04
			causal	5.47	5.42	5.48
			null	1.00	1.00	1.00
		0.1%	all	1.07	1.07	1.07
			causal	8.02	7.81	8.05
			null	1.00	1.00	1.00

Table 2. Results on simulated genotypes and simulated phenotypes. We report average χ^2 statistics. N is the number of individuals and M is the number of SNPs. SNP set indicates either all SNPs, the 1% causal SNPs, or the 99% null SNPs. The disease prevalence ranges from 50% (no case-control ascertainment) to 0.1%.

M	N	PREVALENCE	LIABILITY		OBSERVED	
			H-E	REML	H-E	REML
5000	5000	50%	0.248 (0.005)	0.247 (0.005)	0.158 (0.003)	0.157 (0.003)
		25%	0.232 (0.005)	0.228 (0.005)	0.167 (0.004)	0.164 (0.003)
		10%	0.249 (0.005)	0.239 (0.005)	0.236 (0.005)	0.227 (0.004)
		1%	0.253 (0.003)	0.212 (0.002)	0.459 (0.006)	0.384 (0.004)
		0.1%	0.243 (0.003)	0.157 (0.001)	0.691 (0.009)	0.447 (0.001)
50000	5000	50%	0.266 (0.009)	0.265 (0.009)	0.169 (0.006)	0.169 (0.006)
		25%	0.268 (0.007)	0.275 (0.007)	0.193 (0.005)	0.197 (0.005)
		10%	0.233 (0.007)	0.233 (0.007)	0.222 (0.007)	0.222 (0.007)
		1%	0.254 (0.004)	0.248 (0.004)	0.46 (0.007)	0.449 (0.006)
		0.1%	0.247 (0.002)	0.231 (0.002)	0.700 (0.007)	0.656 (0.006)

Table 3. Heritability parameter estimates on simulated genotypes and phenotypes. These results are from the same simulations used to generate Table 2. We report results on both liability and observed scales. The true h^2 explained by the SNPs used to build the GRM is 25% on the liability scale for all simulations.

M	PREVALENCE	SNP SET	ATT	MLM	LTMLM
1000	50%	all	1.61	1.64	1.64
		causal	17.07	17.96	17.90
		null	1.01	1.01	1.01
	25%	all	1.73	1.75	1.75
		causal	19.64	20.46	20.56
		null	1.02	1.01	1.01
	10%	all	1.90	1.89	1.93
		causal	24.88	25.15	26.11
		null	1.04	1.02	1.03
10000	50%	all	1.08	1.08	1.08
		causal	2.65	2.66	2.66
		null	1.00	1.00	1.00
	25%	all	1.10	1.10	1.10
		causal	2.90	2.91	2.92
		null	1.02	1.01	1.01
	10%	all	1.14	1.13	1.13
		causal	3.58	3.58	3.61
		null	1.03	1.02	1.02

Table 4. Results on real genotypes and simulated phenotypes. We report average χ^2 statistics. M is the number of SNPs, and sample size is fixed at 500 cases and 500 controls.

SNP SET	ATT	PCA	MLM	LTMLM
Genome Wide Average	1.38	1.16	1.14	1.17
Published SNPs Average	11.64	9.97	9.92	10.59
Published SNPs/Genome Wide Average	8.44	8.61	8.67	9.03

Table 5. Results on WTCCC2 MS data set. We report the genome-wide average χ^2 over 360,557 SNPs and the average across 75 published SNPs, before or after normalizing by the genome-wide average. All results are based on analysis of 10,034 individuals (see main text).

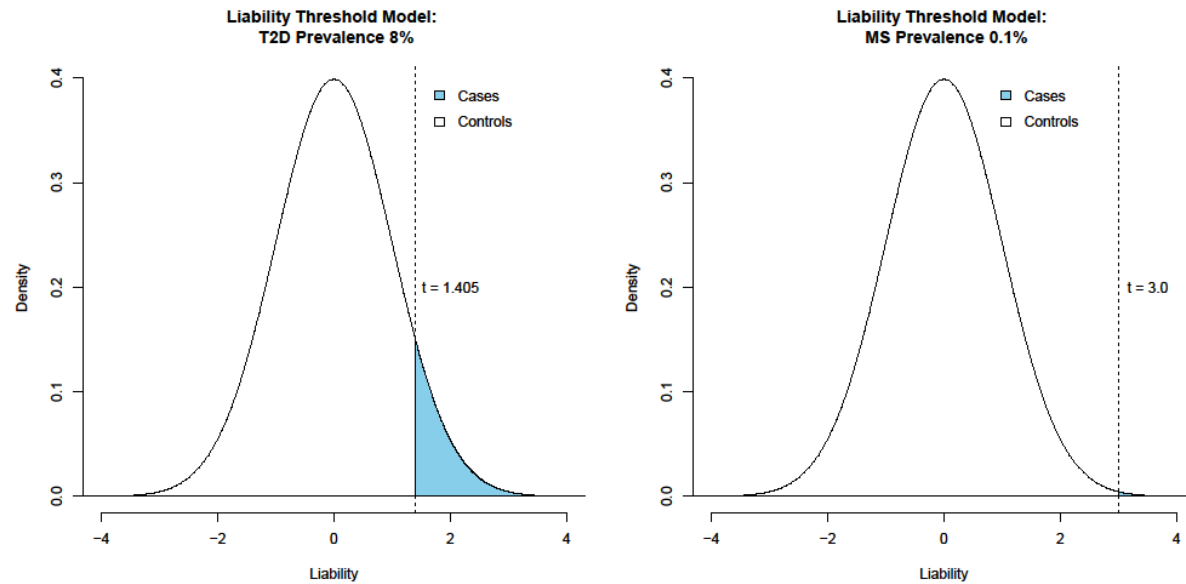
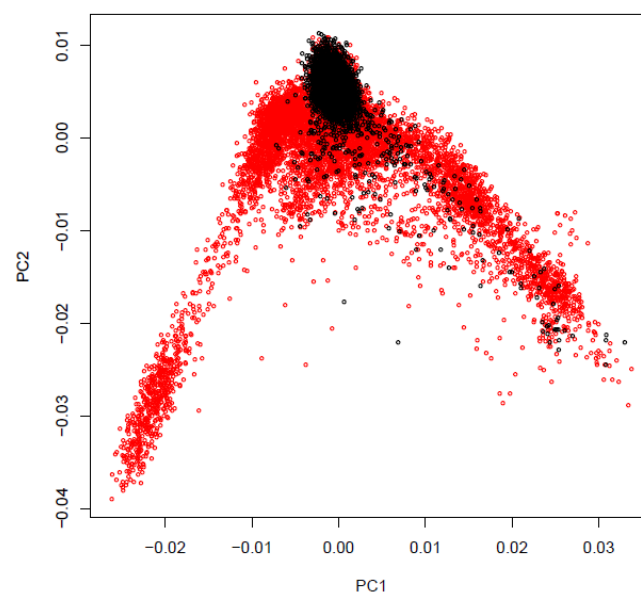


Figure S1. Liability Threshold Model. The liability threshold model performs a transformation based on disease prevalence. As ascertainment becomes more drastic so does the difference between the PML for cases versus controls. In Figure S1, the portion of the population above the threshold is a case (blue). For T2D, at a prevalence of 8% (blue), the threshold is set to 1.405. In this region, the expected value for the posterior liability is 1.85 and the expected value for the controls is -0.14. Comparing T2D to MS with disease prevalence around 0.1% and t around 3.00, the PML_{indiv} for a control is 0.00 and 3.33 for a case. As the disease prevalence goes down the difference in the PML_{indiv} for cases versus controls increases, the transformation plays a larger role for rare diseases and results in a power gain for the LTMLM.

A



B

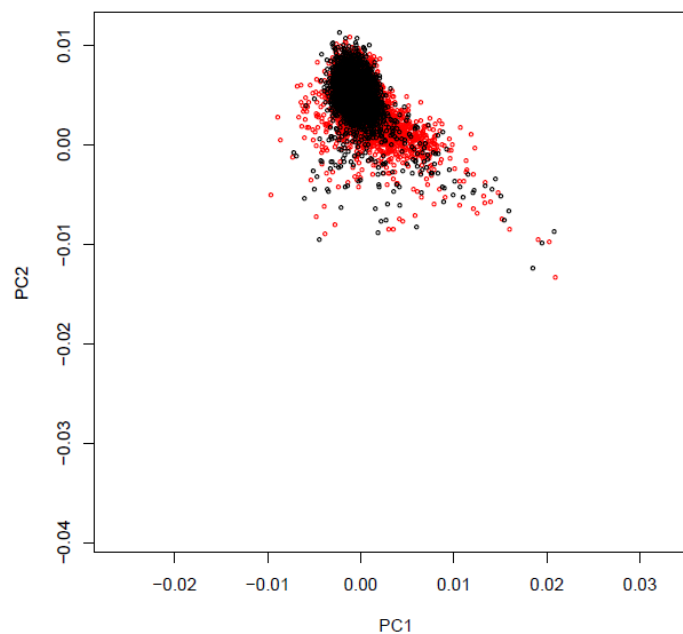


Figure S2. Mismatch in ancestry between MS cases and controls. We plot the first two principal components for (a) unmatched data with a severe mismatch (5,429 MS cases and 10,204 controls), (b) stringently matched data using the first 20 PC (4,094 MS cases and 4,094 controls). The controls are depicted in red and cases in black. After PC matching the remaining samples show considerably less population stratification differentiation between cases and controls.

Term	Description
ϕ	Quantitative liability, the unobserved trait
β	Effect Size of the SNP
x	Genotype values of candidate SNP, normalized to mean 0 variance 1
u	Genetic random effect excluding the candidate SNP
e	Environmental component
X	Matrix of genotype values of non-candidate SNPs, normalized to mean 0 and variance 1
π	Observed binary case control phenotype.
t	Threshold corresponding to the disease prevalence
K	Prevalence of the disease in the population
P	Proportion of cases in the sample
$\hat{\Theta}$	Genetic Relationship Matrix (GRM) computed from the data
Θ	True underlying Genetic Relationship Matrix (GRM)
V	Phenotypic covariance matrix
I	Identity matrix
h^2	Heritability parameter

Table S1: Description of notation used and a brief description of the terms.

Computation	ATT	MLM	LTMLM
GRM and V^{-1}	NA	$O(MN^2)$	$O(MN^2)$
PML	NA	NA	$O(N^3 + iter * N^2) = O(N^3)$
Assoc. Statistic	$O(MN)$	$O(MN)$ or $O(MN^2)$	$O(MN)$
Overall	$O(MN)$	$O(MN^2)$	$O(MN^2)$

Table S2. Computational cost. M is the number of SNPs and N is the number of individuals. We assume that $M > N$. The details of the computational costs of MLM are provided in Table 1 of ref¹¹.

M	N	PREVALENCE	SNP SET	LIABILITY		CASE CONTROL		
				ATT	MLM	ATT	MLM	LTMLM
5000	5000	50%	all	1.24	1.26	1.14	1.15	1.15
			causal	25.83	27.20	16.79	17.13	17.12
			null	0.99	0.99	0.99	0.99	0.99
		10%	all	1.34	1.35	1.23	1.23	1.24
			causal	35.41	36.83	24.59	24.79	25.18
			null	1.00	1.00	0.99	0.99	0.99
		1%	all	1.57	1.50	1.45	1.40	1.46
			causal	58.66	55.61	45.72	42.61	46.85
			null	1.00	0.95	1.00	0.99	1.00
		0.1%	all	1.84	1.55	1.71	1.51	1.73
			causal	85.47	70.68	71.99	59.53	74.05
			null	1.00	0.85	1.00	0.93	1.00
50000	5000	50%	all	1.03	1.03	1.02	1.02	1.02
			causal	3.56	3.57	2.68	2.68	2.69
			null	1.00	1.00	1.00	1.00	1.00
		10%	all	1.04	1.04	1.02	1.02	1.02
			causal	4.45	4.47	3.35	3.35	3.35
			null	1.00	1.00	1.00	1.00	1.00
		1%	all	1.06	1.06	1.04	1.04	1.04
			causal	6.75	6.72	5.47	5.42	5.48
			null	1.00	1.00	1.00	1.00	1.00
		0.1%	all	1.08	1.08	1.07	1.07	1.07
			causal	9.32	9.11	8.02	7.81	8.05
			null	1.00	1.00	1.00	1.00	1.00

M	N	PREVALENCE	SNP SET	LIABILITY		CASE CONTROL		
				ATT	MLM	ATT	MLM	LTMLM
1000	1000	50%	all	1.24	1.25	1.16	1.17	1.16
			causal	24.60	25.94	16.27	16.66	16.65
			null	1.00	1.00	1.01	1.01	1.01
		25%	all	1.28	1.29	1.18	1.18	1.19
			causal	28.07	29.34	18.66	19.02	19.08
			null	1.01	1.01	1.01	1.00	1.00
		10%	all	1.33	1.34	1.22	1.22	1.22
			causal	34.40	35.43	23.77	23.77	24.34
			null	1.00	0.99	0.99	0.99	0.99
		1%	all	1.56	1.52	1.43	1.39	1.45
			causal	56.78	54.64	44.01	41.11	46.16
			null	1.00	0.98	1.00	0.99	1.00
		0.1%	all	1.73	1.59	1.60	1.48	1.64
			causal	72.64	63.00	60.55	51.07	64.69
			null	1.01	0.97	1.01	0.98	1.00
5000	1000	50%	all	1.05	1.05	1.03	1.03	1.03
			causal	6.08	6.14	4.28	4.30	4.30
			null	1.00	1.00	1.00	1.00	1.00
		25%	all	1.06	1.06	1.04	1.04	1.03
			causal	6.64	6.69	4.69	4.69	4.70
			null	1.00	1.00	1.00	1.00	1.00
		10%	all	1.07	1.07	1.05	1.05	1.05
			causal	7.93	8.00	5.79	5.80	5.83
			null	1.00	1.00	1.00	1.00	1.00
		1%	all	1.11	1.11	1.09	1.09	1.09
			causal	12.39	12.27	9.89	9.71	10.01
			null	1.00	1.00	1.00	1.00	1.00
		0.1%	all	1.17	1.16	1.14	1.13	1.15
			causal	17.33	16.51	14.76	13.97	15.03
			null	1.01	1.00	1.01	1.00	1.01

Table S3. Complete results on simulated genotypes and simulated phenotypes. Results are analogous to Table 2, but are reported for other values of M and N . For completeness, we also report ATT and MLM statistics computed using the underlying liability, where we again observe a loss in power for MLM at lower prevalence.

M	N	PREVALENCE	SNP SET	ATT	MLM	LTMLM
1000	1000	10%	All	1.99	1.32	1.25
			Causal	24.28	16.59	16.03
			Null	1.54	1.01	0.95
			Causal/All	12.19	12.54	12.80
10000			All	1.13	1.04	1.03
			Causal	3.40	3.24	3.20
			Null	1.08	1.00	0.99
			Causal/All	3.02	3.11	3.10
20000			All	1.10	1.02	1.01
			Causal	2.26	2.16	2.13
			Null	1.07	1.00	0.99
			Causal/All	2.07	2.11	2.10

Table S4. Results on simulated genotypes and simulated phenotypes with population structure. We report average χ^2 statistics for simulations with population structure (see main text).

M	N	PREVALENCE	LIABILITY		OBSERVED	
			H-E	REML	H-E	REML
1000	1000	50%	0.251 (0.013)	0.253 (0.012)	0.16 (0.008)	0.161 (0.008)
		25%	0.244 (0.012)	0.238 (0.011)	0.175 (0.009)	0.171 (0.008)
		10%	0.234 (0.009)	0.228 (0.009)	0.223 (0.009)	0.217 (0.008)
		1%	0.245 (0.008)	0.207 (0.005)	0.444 (0.014)	0.376 (0.009)
		0.1%	0.245 (0.006)	0.176 (0.003)	0.695 (0.018)	0.499 (0.009)
5000	1000	50%	0.245 (0.016)	0.255 (0.014)	0.156 (0.01)	0.162 (0.009)
		25%	0.261 (0.016)	0.261 (0.015)	0.187 (0.011)	0.188 (0.011)
		10%	0.245 (0.013)	0.249 (0.013)	0.233 (0.012)	0.236 (0.012)
		1%	0.246 (0.006)	0.237 (0.006)	0.446 (0.011)	0.429 (0.011)
		0.1%	0.247 (0.005)	0.222 (0.004)	0.700 (0.014)	0.631 (0.012)

Table S5. Heritability parameter estimates on simulated genotypes and phenotypes. Results are analogous to Table 3, under different settings of M and N .

M	N	PREVALENCE	LIABILITY		OBSERVED	
			HE	REML	HE	REML
1000	10000	10%	0.49 (0.01)	0.442 (0.007)	0.42 (0.007)	0.465 (0.009)
10000			0.348 (0.024)	0.375 (0.014)	0.357 (0.013)	0.413 (0.041)
20000			0.374 (0.027)	0.384 (0.018)	0.365 (0.017)	0.457 (0.045)

Table S6. Heritability parameter estimates on simulated genotypes and phenotypes with population structure.

These results are from the same simulations used to generate Table S5. We report results on both liability and observed scales. The true h^2 explained by the SNPs used to build the GRM is 25% on the liability scale for all simulations. We report results on both liability and observed scales. The true h^2 explained by the SNPs used to build the GRM is 25% on the liability scale for all simulations.

M	PREVALENCE	SET	LIABILITY		CASE-CONTROL		
			ATT	MLM	ATT	MLM	LTMLM
1000	50%	all	1.95	2.07	1.61	1.64	1.64
		causal	26.35	29.53	17.07	17.96	17.90
		null	1.00	1.01	1.01	1.01	1.01
	25%	all	2.12	2.24	1.73	1.75	1.75
		causal	30.37	33.66	19.64	20.46	20.56
		null	1.03	1.02	1.02	1.01	1.01
	10%	all	2.29	2.37	1.90	1.89	1.93
		causal	35.61	38.29	24.88	25.15	26.11
		null	1.02	1.01	1.04	1.02	1.03
10000	50%	all	1.13	1.13	1.08	1.08	1.08
		causal	3.63	3.71	2.65	2.66	2.66
		null	1.01	1.00	1.00	1.00	1.00
	25%	all	1.15	1.15	1.10	1.10	1.10
		causal	3.91	3.99	2.90	2.91	2.92
		null	1.02	1.01	1.02	1.01	1.01
	10%	all	1.19	1.18	1.14	1.13	1.13
		causal	4.74	4.83	3.58	3.58	3.61
		null	1.02	1.01	1.03	1.02	1.02

Table S7. Complete results on real genotypes and simulated phenotypes. Results are analogous to Table 4, but we also report ATT and MLM statistics computed using the underlying liability.

PREVALENCE	ATT	MLM	LTMLM
50%	30.0%	29.7%	30.4%
25%	21.4%	22.2%	22.5%
10%	21.4%	22.2%	22.5

Table S8. Percentage of causal SNPs achieving genome-wide significance ($p < 5 \times 10^{-8}$). Results are based on 500 cases and 500 controls with real genotypes and simulated phenotypes, where $M = 1000$ SNPs.

M	PREVALENCE	LIABILITY		OBSERVED	
		HE	REML	HE	REML
1000	50%	0.241 (0.013)	0.243 (0.01)	0.153 (0.008)	0.155 (0.006)
	25%	0.237 (0.011)	0.24 (0.009)	0.170 (0.008)	0.173 (0.006)
	10%	0.24 (0.013)	0.227 (0.008)	0.228 (0.012)	0.216 (0.008)
10000	50%	0.262 (0.01)	0.276 (0.009)	0.167 (0.006)	0.176 (0.006)
	25%	0.258 (0.01)	0.274 (0.01)	0.185 (0.007)	0.197 (0.007)
	10%	0.263 (0.008)	0.273 (0.008)	0.250 (0.008)	0.260 (0.008)

Table S9. Heritability parameter estimates on real genotypes and simulated phenotypes. These results are from the same simulations used to generate Table 4. The true heritability explained by the SNPs used to build the GRM is 25% on the liability scale for all simulations.

SNP SET	ATT	PCA	MLM	LTMLM
GENOME WIDE AVERAGE	1.38	1.16	1.14	1.17
GENOME WIDE LD SCORE INTERCEPT	1.29	1.09	1.08	1.10
PUBLISHED SNPS AVERAGE	11.64	9.97	9.92	10.59
PUBLISHED SNPS/GENOME WIDE AVERAGE	8.44	8.61	8.67	9.03
PUBLISHED SNPS/LD SCORE INTERCEPT	9.06	9.17	9.20	9.66

Table S10: Results on WTCCC2 MS data set with calibration via LD Score regression. We report the genome wide χ^2 averages using 10,034 individuals over 360,557 SNPs and the average across 75 published SNPs standardized by the genome wide average and LD Score regression intercept.

SNP SET	N	ATT	MLM	LTMLM
GENOME WIDE AVERAGE	8188	1.16	1.11	1.14
PUBLISHED SNPS AVERAGE		8.94	8.26	8.76
PUBLISHED SNPS/GENOME WIDE AVERAGE		7.73	7.45	7.71
GENOME WIDE AVERAGE	10034	1.38	1.14	1.17
PUBLISHED SNPS AVERAGE		11.64	9.92	10.59
PUBLISHED SNPS/GENOME WIDE AVERAGE		8.44	8.67	9.03
GENOME WIDE AVERAGE	15633	3.95	1.23	1.08
PUBLISHED SNPS AVERAGE		18.54	11.30	5.76
PUBLISHED SNPS/GENOME WIDE AVERAGE		4.69	9.20	5.32

Table S11. Results on WTCCC2 MS data set at different levels of QC. We report results for stringently matched ($N = 8,188$), partially matched ($N = 10,034$) and unmatched ($N = 15,633$) data sets (see main text).

N	Liability		Observed	
	HE	ReML	HE	ReML
8188	0.363 (0.0017)	0.260 (0.001)	0.979 (0.005)	0.702 (0.003)
10034	0.704 (0.009)	0.279 (0.001)	1.901 (0.025)	0.753 (0.002)
15633	2.792 (0.010)	0.293 (0.001)	7.543 (0.0266)	0.792 (0.002)

Table S12: Heritability parameter estimates on WTCCC2 MS data set at different levels of QC. We report results for stringently matched ($N = 8,188$), partially matched ($N = 10,034$) and unmatched ($N = 15,633$) data sets (see main text).