Accurate Liability Estimation Substantially Improves Power in Ascertained Case Control Studies

Omer Weissbrod[1,*], Christoph Lippert[2], Dan Geiger[1] and David Heckerman[2,**]

[1]Computer Science Department, Technion - Israel Institute of Technology, Haifa 32000, Israel

[2]eScience Group, Microsoft Research, Los Angeles 90024, USA

* Correspondence: omerw@cs.technion.ac.il

** Correspondence: heckerma@microsoft.com

Running Title: Liability Estimation Improves Case Control GWAS

Abstract

Future genome-wide association studies (GWAS) of diseases will include hundreds of thousands of individuals in order to detect risk variants with small effect sizes. Such samples are susceptible to confounding, which can lead to spurious results. Recently, linear mixed models (LMMs) have emerged as the method of choice for GWAS, due to their robustness to confounding. However, the performance of LMMs in case-control studies deteriorates with increasing sample size, resulting in reduced power. This loss of power can be remedied by transforming observed case-control status to liability space, wherein each individual is assigned a score corresponding to severity of phenotype. We propose a novel method for estimating liabilities, and demonstrate that testing for associations with estimated liabilities by way of an LMM leads to a substantial power increase. The proposed framework enables testing for association in ascertained case-control studies, without suffering from reduced power, while remaining resilient to confounding. Extensive experiments on synthetic and real data demonstrate that the proposed framework can lead to an average increase of over 20 percent for test statistics of causal variants, thus dramatically improving GWAS power.

Introduction

In recent years, genome-wide association studies (GWAS) have uncovered thousands of risk variants for genetic traits[1]. Despite this success, case-control GWAS suffer from several difficulties.

The first difficulty is the small fraction of disease variance that is explained by discovered variants[2, 3]. One widely accepted explanation is that existing studies are underpowered to identify the vast majority of risk variants, because they only exert a small influence on genetic traits[4]. To identify such variants, future studies will need to include hundreds of thousands of individuals.

A second difficulty is sensitivity to confounding due to population structure and family relatedness[5], leading to spurious results and increased type I error rate. As sample sizes continue to increase, this difficulty becomes even more severe, because larger samples are more likely to include individuals with a different genetic ancestry or related individuals.

Recently, linear mixed models (LMMs) have emerged as the method of choice for GWAS, due to their robustness to diverse sources of confounding[6]. LMMs gain resilience to confounding by testing for association conditioned on pairwise kinship coefficients between study subjects. These kinship coefficients are typically estimated using genetic variants, such as single nucleotide polymorphisms (SNPs). Although designed to test for association with continuous phenotypes, LMMs have been successfully used in several large case-control GWAS[7-9], because alternative methods cannot capture diverse sources of confounding[6].

A third difficulty in GWAS concerns the use of LMMs in ascertained case-control studies, wherein cases are oversampled relative to the disease prevalence. It has recently been discovered that LMM performance in such studies deteriorates with increasing sample size, leading to loss of power compared to alternative methods[10]. Thus, the use of LMMs resolves the second difficulty of sensitivity to confounding, but leads to a different difficulty instead.

The loss of power of LMMs in ascertained case-control studies stems from violation of several of their modelling assumptions. First, LMMs assume that variants used to estimate kinship are independent of tested variants. However, several recent studies have demonstrated that causal variants tend to become correlated under ascertainment, because cases of rare diseases are likely to carry excessive dosages of risk alleles in multiple causal variants[10-15]. Second, LMMs assume that genetic and environmental disease factors are independent. However, it has recently been demonstrated that these factors become correlated under ascertainment as well[16]. Third, LMMs assume that variants have an additive effect on the phenotype, which is obviously not true for case-control phenotypes. Combined, these violations can lead to downward-biased estimates of the effect size of tested causal variants[10], leading to loss of power to identify causal variants (Supplementary note).

Yang *et al.* have demonstrated that the severity of power loss in LMMs increases with the ratio between the sample size and the effective number of common genetic variants used to estimate kinship[10]. This ratio increases with sample size, regardless of genotyping density, because the effective number of common variants is bounded and is smaller than the number of genotyped variants, owing to linkage disequilibrium[17]. Thus, although the absolute power of LMMs increases with sample size, the increase is expected to be small compared to alternative methods, owing to model misspecification. However, alternative methods may not be robust to confounding, as discussed above.

A possible remedy is testing for associations with a modelling framework that directly models the case-control phenotype and the ascertainment scheme, such as a generalized LMM (GLMM) with an appropriate link function[18], or a full retrospective model (methods). The probabilistic models underlying such modelling frameworks assume that observed case-control phenotypes are generated by an unobserved stochastic process with a well defined distribution. One prominent example is the well known liability threshold model,[19] which associates individuals with a latent normally distributed variable called the *liability*, such that cases are individuals whose liability exceeds a given cutoff. Despite their elegance, such modelling frameworks are extremely computationally expensive, rendering whole genome association tests infeasible in most circumstances.

As an alternative, we propose approximating such modelling frameworks by first estimating a latent liability value that is conditional on phenotypes, genotypes and disease prevalence for every individual, and then testing for association with the estimated liabilities via an LMM (see methods). This proposal is motivated by the observation that cases of rare diseases have a sharply peaked liability distribution (Figure 1), leading to improved liability estimation. Intuitively, liabilities of cases of rare diseases are approximately equal to the liability cutoff, which decreases the number of degrees of freedom of the estimation (see Supplementary Note for a full derivation). Testing for association with liabilities via LMMs is more powerful than the common approach of testing for association with case-control status, because variants then have an additive effect on the liability, which resolves one of the three model violations discussed above (Supplementary note). These arguments demonstrate that liability estimation can help increase power in the study of rare diseases, whereas naive use of LMMs decreases power in such settings, compared to alternative methods such as logistic regression[10]. However, alternative methods may not properly control for type I error, as discussed above.

Our proposed framework, called LEAP (Liability Estimator As a Phenotype), computes accurate liability estimates for GWAS-sized data sets, conditional on phenotypes, whole-genome genotypes and disease prevalence. The computation can be carried in a few minutes, by reusing computations that are also employed by LMMs as a preprocessing step. These liability estimates are then used as observed phenotypes for association testing via an LMM. The proposed framework enables testing for association in ascertained case-control studies without suffering from power loss, while remaining resilient to confounding. LEAP thus successfully addresses the three difficulties described above.

LEAP bears similarities to several recently developed methods for estimating the portion of the liability that is explained by a given set of explanatory variables[11, 15]. These methods are designed to prevent power loss in ascertained case-control studies using covariates, which arises due to induced correlations between tested and conditioned variables, as discussed above. However, these methods estimate the liability explained by a small set of covariates, whereas LMMs implicitly condition association tests on the entire genome, owing to the well known equivalence between LMMs and linear regression[20, 21]. These methods are therefore not designed to estimate liabilities using whole genome information. A second key difference is that the aforementioned methods test variants for association with the residuals of the estimated liabilities, after regressing out the influence of the explanatory variables. In contrast, LEAP directly tests variants for association with the estimated liabilities, while effectively conditioning on all genome-wide variants. The use of genome-wide variants in both the liability estimation and in the association testing stage prevents spurious results due to confounding.

Very recently, Hayeck *et al.* proposed an alternative framework, called LTMLM, for association testing under ascertained case-control sampling in the presence of confounding[22]. Both LTMLM and LEAP first estimate latent liability values and then test for association with these estimates. However, LTMLM tests for association with the posterior mean of the liabilities in a score test framework, whereas LEAP tests for association with the maximum a posteriori (MAP) of the liabilities. We found that utilizing the MAP estimator results in improved accuracy over the posterior mean estimator under a wide range of scenarios and at a substantially reduced computational cost (methods and Supplementary note).

Through extensive analysis of synthetic and real disease data sets, we demonstrate that LEAP substantially improves both power and type I error control over a standard LMM, and remains robust to confounding even in the presence of extreme population structure and family relatedness. The power gains of LEAP over a standard LMM increase with sample size, heritability, and ascertainment. In real data sets, LEAP obtained an average increase of over 8% in test statistics of SNPs known to be associated with the phenotype. In synthetic data sets, LEAP obtained an average increase of over 20% in test statistics of causal SNPs, and over a 5% gain in power.

Results

We evaluated the performance of LEAP using synthetic and real data sets. For comparison, we evaluated the following methods: (a) LEAP, (b) A standard LMM, (c) A linear regression test using 10 principal component (PC) covariates[23] (denoted Linreg+PCs), and (d) a univariate linear regression test (Linreg) without PC covariates, used as a baseline measure.

The fixed effects models use the linear link function to prevent evaluation bias due to using a different link function for different methods. Experiments using logistic regression yielded very similar results (Supplementary Figure S1). Experiments where the LEAP liability estimates are tested for association via linear regression methods were not conducted, because this can lead to test statistic inflation, and consequently to lower power (Supplementary Figure S2).

Several recent papers proposed improving the power of LMMs by estimating kinship via a subset of variants that account for a large fraction of the phenotype variance[24-27]. This strategy can work well for continuous phenotypes. However, under ascertained case-control sampling, this strategy renders the problem of power loss under LMMs even more severe, because it increases the ratio between the number of individuals and the number of variants used by the LMM. Larger values of this ratio increase the severity of power loss, as shown in ref.[10] and in the results below. We empirically verified that under variant selection, the severity of power loss increases as prevalence decreases (Supplementary Figure S3 and Supplementary Note). We therefore opted to not follow this strategy in our experiments. We further address this issue in the discussion.

Sensitivity to Confounding

Sensitivity to confounding was evaluated by measuring the type I error rate for synthetic data sets (see methods for the data generation procedure). We evaluated various combinations of population structure (quantified via the F_{ST} measure[28]) and family relatedness (measured via the fraction of sib-pairs in the study). Specifically, we evaluated F_{ST} levels of 0, 0.01 and 0.05, and sib-pair fractions of 0%, 3% and 30%. Each data set contained 6,000 individuals, and 10 data sets were generated for each evaluated combination of settings.

The type I error rate was measured by comparing the expected and empirical type I error rates associated with different P values. The results, shown in Supplementary Figures S4-S6, demonstrate that the linear regression tests could not properly control type I error in the presence of population structure (without using PC covariates) and family relatedness (even when using PC covariates), which is consistent with previous findings[6]. In contrast, LEAP had type I error control within or below the expected rate in all cases, demonstrating robustness to confounding. Supplementary Figures S4-S6 indicate that LEAP is slightly conservative if there is only a small amount of genetic relatedness, indicating that it may be slightly underpowered in such settings. Nevertheless, LEAP is substantially more powerful than a standard LMM even under such settings, as shown below.

To further asses sensitivity to confounding, we measured the actual type I error rates at p=0.05 and at $p=10^{-5}$, and the genomic control inflation factor λ_{GC}[29], defined as the ratio between the observed and expected test statistic median in χ^2 space (Supplementary Table S1). LEAP properly controlled type I error at the tail of the distribution, and had a lower λ_{GC} value than a standard LMM in all cases, indicating robustness to confounding.

Power Evaluations

The power of the methods was evaluated according to the distribution of test statistics of causal variants[11, 15]. Any evaluation of power must take sensitivity to confounding into account. Otherwise, Linreg and Linreg+PCs may falsely appear to be more powerful than the other methods, because they yield inflated P values in the presence of confounding. Consequently, in the analysis of simulated data, we computed the empirical type I error rate associated with each P value under each method, and then computed power as a function of the type I error rate (Supplementary Note).

For both simulated and real data, we employed an additional evaluation that facilitates direct comparison between methods, and provides easily interpretable results. This evaluation compares the distribution of test statistics of causal variants, normalized according to the distribution of test statistics of all variants (Supplementary Note). The ratio of the test statistics means is closely related to the relative increase in sample size needed to obtain equivalent power[30]. Both proposed evaluation approaches assess empirical power given the true type I error rate. However, the first measure simply counts the number of test statistics exceeding the significance cutoff (which is dependent on sample size), whereas the second one is sensitive to systematic differences in the distribution of such test statistics.

To evaluate the effects of sample size and ascertainment on power, we generated ascertained case-control data sets with disease prevalence of 0.1%, 1% and 10%, and sample sizes of 2,000, 4,000, 6,000, 8,000 and 10,000. Ten data sets were generated for every combination of settings. Unless otherwise noted, in all simulations there is population structure of F_{ST}=0.01, and 30% of the individuals in one of the two populations are sib-pairs. This was done to create a challenging scenario that is suitable for future large studies, which may include individuals with a different genetic ancestry and related individuals. The effect of these factors on power is examined below.

The results indicate that the advantage of LEAP over a standard LMM increases with sample size and with ascertainment (Figure 2 top row and Supplementary Figures S7-S8). In simulated samples with 0.1% prevalence and 10,000 individuals, LEAP gained an average increase of over 20% in test statistics of causal SNPs (Figure 2, top left pane) and a 3% gain in average power, where power was averaged over all significance levels (Figure S7). Moreover, LEAP gained a power increase of over 5% for significance levels smaller than 5×10^{-5} (Figure S7). Linreg+PCs also gained an advantage over an LMM as sample size increased, but it remains sensitive to confounding, as demonstrated above. We also evaluated the performance of LEAP under more complex ascertainment schemes, where it remained more powerful than a standard LMM (Supplementary Figure S19 and Supplementary note).

To gather some intuition into the advantage of LEAP over a standard LMM, we compared estimated and true liabilities. This comparison was applied only for controls, because liabilities of cases are trivial to estimate, as they are tightly clustered near the liability cutoff (Figure 1). In balanced case-control studies, estimation accuracy increased as prevalence decreased (Figure 3). In the supplementary note, we demonstrate that accurate liability estimation leads to increase in power. These results indicate that while the performance of standard LMMs decrease with decreasing prevalence, the opposite effect holds for LEAP.

Accurate liability estimation depends on the fraction of liability variance that is driven by genetic factors, called the narrow-sense heritability[2, 3]. A higher heritability is expected to improve estimation accuracy, because more of the liability signal can be inferred from observed variants. We empirically verified that the advantage of LEAP over an LMM increased with heritability, with noticeable power gains for disease with heritability greater or equal to 25% (Figure 2 bottom row and Supplementary Figure S9). It is estimated that many rare genetic diseases have narrow-sense heritability greater than 25%[16], indicating that LEAP is relevant for the study of real diseases.

We evaluated the effects of different population structure and family relatedness levels on LEAP performance. To this end, we verified that LEAP outperformed the other methods under various population structure settings, and held its advantage even under unusually large F_{ST} levels (Supplementary Figures S10-S11). We also observed that the advantage of LEAP increased with the number of related individuals in the sample, because alternative methods are more susceptible to power loss or to an inflation of P values in the presence of relatedness (Supplementary Table S1 and Supplementary Figures S12-S13). We conclude that LEAP outperforms other methods in the presence of diverse sources of confounding.

We continued by evaluating the performance of the methods under different polygenicity levels, defined as the number of causal SNPs driving the disease (Supplementary Figures S14-S15). All methods gradually lost power as polygenicity increased, because the effect size of each causal SNP became weaker. Nevertheless, LEAP outperformed the other methods under all evaluated settings.

Finally, we evaluated the methods in the presence of covariates. Naive inclusion of covariates in ascertained case-control studies can lead to power loss, owing to induced correlations between covariates and tested variants[11-14]. However, power can be gained by explicitly including covariates in the liability estimation, and then regressing their effect out of the estimated liabilities (Supplementary Note). This approach led to an average increase of over 2% in test statistics of causal SNPs, over standard use of LEAP that ignored the covariates (Supplementary Figures S16-S17). The relatively modest increase may be attributed to the fact that a significant portion of the covariate signal is already captured by genotyped SNPs due to induced correlations, and is thus already accounted for in standard use of LEAP.

Analysis of Real Data

We analyzed nine disease data sets from the Wellcome Trust case control consortium (WTCCC)[8, 31, 32]. Measuring power for real data sets is an inherently difficult task, because the identities of true causal SNPs are unknown. Evaluating type I error control for real data is also a difficult task, because inflation of P values may stem from either sensitivity to confounding, or from high polygenicity of the studied trait[33].

As an approximate measure for type I error, we verified that the proportion of SNPs having $p<0.05$ and $p<10^{-5}$, and that are not within 2M base pairs of SNPs reported to be associated with the disease in previous studies, is comparable under LEAP and under a standard LMM. As an approximate measure for power, we computed normalized test statistics for known associated SNPs from the NHGRI catalog[34] as a bronze standard.

LEAP demonstrated robustness to confounding, and was significantly more powerful than a standard LMM in five out of the six rare phenotypes, with prevalence smaller than 1% (Figure 4 and Supplementary Tables S2-S3). As expected from the simulations, the advantage of LEAP over an LMM increased with sample size and with confounding. Thus, only a small advantage was observed in the WTCCC1 data sets, which contain about 4,500 individuals per data set and little population structure or family relatedness, whereas a significantly greater advantage was observed in the larger and more confounded multiple sclerosis (MS) and ulcerative colitis (UC) data sets.

In the highly confounded MS data set[8], LEAP obtained a mean increase of more than 8% over an LMM in test statistics of tag SNPs, and an even greater advantage over other methods, while demonstrating robustness to confounding. All genome-wide significant loci identified by LEAP and LMM, having $p<5\times10^{-8}$, have previously been reported to be associated with MS in meta-analyses. In contrast, Linreg+PCs and Linreg identified 2 and 508 previously unidentified significant loci, respectively. These results indicate that LEAP is more powerful than the other methods, while remaining robust to confounding.

In the WTCCC1 data sets, LEAP, Linreg+PCs and Linreg all had a similar advantage over a standard LMM, consistent with the simulations that show similar performance for Linreg+PCs and Linreg under no population structure (Supplementary Figure S10), and for LEAP and Linreg+PCs under no family relatedness (Supplementary Figure S12).

Discussion

We presented a novel framework called LEAP for liability estimation and association testing, and demonstrated that it can lead to substantial improvements over existing methods. The core idea of LEAP is that liabilities can be accurately estimated under severe ascertainment, by inferring the overall effect of genotypes on case-control status. The advantage of LEAP over existing methods increases with sample size and with increasing levels of heritability, ascertainment and confounding in the data. GWAS sample sizes are expected to greatly increase in the near future, necessitating efficient association testing methods that can retain high power in ascertained case-control studies of unbounded size, while remaining resilient to confounding. LEAP is a promising framework for such large studies.

LEAP is derived from the well-known liability threshold model[19], which may only partially reflect the underlying mechanism of real diseases. The effect of power loss under ascertainment, and the accuracy of liability estimators, may be different under alternative disease models. Despite the purely theoretical background of the liability threshold model, many recent studies have demonstrated its usefulness and applicability[3, 11, 13, 16, 35].

LEAP approximates models that fully account for dichotomous phenotypes and the presence of ascertainment. Such models can in principle more accurately fit case-control data and thus yield improved results. However, such models are intractable due to the need to integrate over a high dimensional subspace of liabilities[36]. We argue that the advantage of such models over using a single liability estimator is attenuated with decreasing prevalence levels, because liabilities can be more accurately estimated with decreasing prevalence (Figure 3 and Supplementary Note). We further note that liabilities follow a truncated multivariate normal distribution, and thus their likelihood cannot be computed by an LMM without model misspecification, even if they are perfectly estimated. A potential future direction is to modify the objective function of LEAP so that its estimated liabilities more closely follow a multivariate normal distribution, similarly to ref.[37].

Finally, LEAP models polygenicity by assuming a Gaussian prior on the effect size of all genotyped variants, in accordance with refs.[2, 3, 16]. In recent years, researchers have argued that polygenicity can be more accurately modelled by either selecting phenotype-specific variants[24-27], or by using more elaborate models that can express richer interaction patterns[27, 38-42]. Adapting LEAP to such models merits investigation.

Online Methods

LEAP Overview

The LEAP procedure is composed of four parts, which are now briefly overviewed, with detailed explanations following below.

1. Heritability estimation: The heritability of a trait quantifies the degree to which it is driven by genetic factors[2, 3]. Several methods for heritability estimation in case-control studies have been proposed recently[3, 15]. We adopt the method of ref.[16], which directly models the ascertainment procedure.
2. Fitting a Probit model: Using the heritability estimate, we fit a regularized Probit model, to estimate the effect size of each genetic variant on the liability.
3. Liability estimation: Using the fitted Probit model, a liability estimate is computed for every individual.
4. Association testing: The liability estimate is used as an observed phenotype in a GWAS context. Genetic variants are tested for association with this estimate via a standard LMM. The LMM is fitted using the heritability estimate, as described below.

Our main contribution lies at stages 2-3 of the procedure, described in detail below. To motivate the use of LEAP, we begin by introducing the liability threshold model and its relation to LMMs.

The Liability Threshold Model

LEAP originates from the statistical framework of the liability threshold model[19], which is briefly presented here. A key assumption behind this model is that every individual i carries a latent normally distributed liability variable $l_i \sim \mathrm{N}(0,1)$. Cases are individuals whose liability exceeds a given cutoff t, i.e. $l_i \geq t$. The cutoff t can be inferred given the disease prevalence K as $t = \Phi^{-1}(1-K)$, where $\Phi^{-1}(\cdot)$ is the inverse cumulative probability density of the standard normal distribution.

The liability l_i can be decomposed into two additive terms corresponding to the genetic and environmental effects affecting a trait, denoted as g_i and e_i:

$$l_i = g_i + e_i. \qquad (1)$$

Without loss of generalization, we assume that g_i and e_i are independently drawn from zero-mean normal distributions with variances σ_g^2 and σ_e^2, respectively, and thus $\sigma_g^2 + \sigma_e^2 = 1$. The genetic term g_i for an individual is given by a linear combination of genetic variants and their corresponding effect sizes,

$$g_i = \sum_{j=1}^{m} v_{ij}\beta_j \qquad (2)$$

where β_j is the effect size of variant j and v_{ij} is the value of variant j for individual i, standardized to have zero mean and unit variance. The effect sizes are assumed to be drawn iid from a normal distribution,

$$\beta_j \sim \mathrm{N}\left(0, \sigma_g^2/m\right). \qquad (3)$$

When the identities of truly causal variants are unknown, a commonly used assumption is that all genotyped variants have an effect size drawn from this normal distribution[2].

The genetic and environmental effects g_i and e_i are deeply related to the narrow-sense heritability[2] of a trait, defined as $h^2 = \sigma_g^2/\left(\sigma_g^2 + \sigma_e^2\right)$. This term is used to quantify the degree to which a given trait is affected by genetic factors. Recently, methods for estimating the underlying heritability of the liability of case-control traits have been proposed[3, 16]. These methods can be used to estimate the heritability, and consequently the variances σ_g^2 and σ_e^2.

Linear Mixed Models

To motivate LEAP, we first present the LMM framework. For a given sample of individuals, LMMs assume that an observed phenotypes vector y follows a multivariate normal distribution

$$y \sim \mathrm{N}\left(\mu, \sigma_g^2 C + \sigma_e^2 I\right) \qquad (4)$$

where μ is the distribution mean, I is the identity matrix, C is a covariance matrix encoding genetic correlations between individuals, and σ_g^2, σ_e^2 are the variances of the genetic and environmental components of the covariance, respectively. This model naturally encodes the assumption that genetically similar individuals are more likely to share similar phenotypes. The genetic covariance matrix C is often estimated from genotypes variants as $C = \frac{1}{m} X X^T$, where X is a design matrix of genotyped variants, standardized so that all columns have zero mean and unit variance. Association testing for a given variant v can be carried out by assigning $\mu = \mu_0 + v\alpha_v$, where α_v is the variant effect size, and attempting to reject the null hypothesis $\alpha_v = 0$ by fitting the model via restricted maximum likelihood[43].

A close relation between the LMM and the liability threshold model is revealed by considering the relation between an LMM and linear regression, wherein effect sizes are drawn from $\mathrm{N}\left(0, \sigma_g^2/m\right)$. Denoting $\varphi(\, y;\ \mu, \Sigma)$ as the probability density of the multivariate normal distribution, and using basic properties of the normal distribution, the LMM can be rewritten as follows.

$$\varphi\left(\mathrm{y};\ \mu, \sigma_\mathrm{g}^2 \mathrm{C} + \sigma_\mathrm{e}^2 \mathrm{I}\right) = \varphi\left(y;\ \mu, \frac{\sigma_\mathrm{g}^2}{\mathrm{m}} \mathrm{XX}^\mathrm{T} + \sigma_\mathrm{e}^2 \mathrm{I}\right) = \int \varphi(\mathrm{y};\ \mu + \mathrm{X}\beta, \sigma_\mathrm{e}^2 \mathrm{I})\, \varphi\left(\beta;\ 0, \frac{\sigma_\mathrm{g}^2}{\mathrm{m}} \mathrm{I}\right) \mathrm{d}\beta. \qquad (5)$$

The phenotypes distribution under LMMs is therefore equivalent to the liability distribution under the liability threshold model, after integrating the effect sizes out.

This interpretation of LMMs provides a straightforward way to extend them to handle binary phenotypes. Given a disease with prevalence K and a corresponding liability cutoff t, the likelihood for a given case-control status vector p conditional on K is given by

$$\mathrm{P}\big(p \mid \mathrm{X}; \mu, \sigma_g^2, \sigma_e^2, \mathrm{K}\big) = \int \left[\begin{array}{c} \varphi\left(\beta;\ 0, \frac{\sigma_g^2}{m}\mathrm{I}\right) \prod_{\mathrm{i}\in\mathrm{controls}} \Phi\big(\mathrm{t} - \mu - \mathrm{X}_\mathrm{i}^\mathrm{T}\beta;\ 0, \sigma_e^2\big) \\ \prod_{\mathrm{i}\in\mathrm{cases}} \Big(1 - \Phi\big(\mathrm{t} - \mu - \mathrm{X}_\mathrm{i}^\mathrm{T}\beta;\ 0, \sigma_e^2\big)\Big) \end{array} \right] d\beta \qquad (6)$$

where $\Phi(\mathrm{y};\ \mu, \Sigma)$ is the cumulative probability density of the normal distribution, and $\mathrm{X}_\mathrm{i}^\mathrm{T}$ is the i^{th} row of X. The relation to the liability threshold model can be made clearer by rewriting this likelihood as

$$\mathrm{P}\big(p \mid X; \mu, \sigma_g^2, \sigma_e^2, \mathrm{K}\big) = \int_\mathrm{V} \mathrm{P}(\mathrm{l} \mid \mathrm{X})\mathrm{dl} \qquad (7)$$

where $\mathrm{l} = \mu + \mathrm{X}\beta + \mathrm{e}$ is the underlying liability, V is the subspace wherein $\mathrm{l_i} \geq \mathrm{t}$ for cases and $\mathrm{l_i} < t$ for controls, and

$$\mathrm{P}(\mathrm{l} \mid \mathrm{X}) = \int \varphi(\mathrm{l};\ \mu + \mathrm{X}\beta, \sigma_e^2\mathrm{I})\varphi\left(\beta;\ 0, \frac{\sigma_g^2}{\mathrm{m}}\mathrm{I}\right) \mathrm{d}\beta = \varphi\left(\mathrm{l};\ \mu, \sigma_g^2 \frac{1}{\mathrm{m}}\mathrm{XX}^\mathrm{T} + \sigma_e^2\mathrm{I}\right) \qquad (8)$$

is the liability density. Recall that $C = \frac{1}{\mathrm{m}}\mathrm{XX}^\mathrm{T}$ is the LMM genetic covariance matrix. Thus, computing the likelihood of a case-control phenotype is equivalent to integrating the underlying liability over its support.

The above derivation suggests a natural way to perform association testing in the presence of case-control phenotypes. However, this requires fitting the parameters and performing a sampling procedure over liability values for every tested variant, resulting in excessively expensive computations that are infeasible in most circumstances. We note that in order to accurately model ascertainment, an even more complex model should be used, which either directly models the ascertainment scheme, or uses a retrospective likelihood (Supplementary note).

Liabilities Estimation

As discussed above, testing for associations under the liability threshold model requires integrating the underlying liability vector over its support. Motivated by this observation, we propose approximating such association testing by selecting a liability estimator and treating it as the observed phenotype vector. A good liability estimator has values close to the true, unobserved, underlying liability. Thus, the problem is equivalent to inferring the value of an unknown continuous variable with a known distribution.

Recall that the liabilities vector l is given by $l = g + e$, where g and e are the genetic and environmental components of the liability, respectively. Further recall that g is given by $g = X\beta$, where X is the genotypes matrix and β is a vector of effect sizes. We consider two closely related liability estimators: The joint maximum a posteriori estimate (MAP), and the genetic MAP. The first quantity jointly estimates the posterior mode of β and e, conditional on the observed phenotypes, genotypes and the disease prevalence. The second quantity first estimates $\hat{\beta}$, the MAP of β, by considering e as a nuisance parameter that is integrated out, and then finds the MAP of e given $\hat{\beta}$. Although the first quantity has a clearer interpretation, the second quantity has favourable properties that render it superior in practice (detailed below), and will be used in LEAP.

Another natural estimator of l is its posterior mean, which can be obtained via sampling[22, 35]. Our experiments have demonstrated that, in the presence of population structure, the genetic MAP estimator employed by LEAP obtains similar or greater liability estimation accuracy, at a significantly reduced computational cost (Supplementary note, Supplementary Figure S18 and Supplementary Table S4).

We now describe the derivation and computation of both MAP quantities in detail. Importantly, while the derivations below do not explicitly take the case-control sampling scheme into account, they yield identical results to derivations that do take the ascertainment procedure into account (Supplementary Note). Furthermore, while the optimization problems derived below are extremely high dimensional, they can readily be formulated as lower-dimensional problems with dimensionality equal to the sample size, as described in the next section. Inclusion of covariates is described in the Supplementary Note.

Joint MAP Estimator

The joint MAP maximizes the joint posterior likelihood of β and e, conditional on the phenotypes, genotypes and the disease prevalence. Denoting p as the vector of observed case-control phenotypes and K as the disease prevalence, the likelihood to maximize can be written as

$$P(\beta, e \mid X, p; K) \propto P(\beta)P(e)P(p \mid \beta, e, X; K). \qquad (9)$$

where the proportionality sign indicates that the likelihood is scaled by a constant that is independent of β and e. Equation 9 makes use of the fact that β and e are marginally independent of X, K and of each other. The probability $P(p \mid \beta, e, X; K)$ is a delta function that is equal to one if all cases/controls have liabilities greater/smaller than the cutoff $t = \Phi^{-1}(1 - K)$, and zero otherwise. Therefore, using the definitions of β and e, computing the MAP is equivalent to solving the optimization problem

$$argmax_{\beta,e}\, \varphi\left(\beta;\ 0, \frac{\sigma_g^2}{m}\mathrm{I}\right)\varphi(e;\ 0, \sigma_e^2\mathrm{I}) \quad \mathrm{s.t.}\ \ \mathrm{p}(\mathrm{X}\beta + \mathrm{e}) \le \mathrm{pt} \qquad (10)$$

where we encode $\mathrm{p_i} = 1$ for controls and $\mathrm{p_i} = -1$ for cases, and the inequality is evaluated component-wise. Taking the logarithm, transforming the maximization to a minimization, and using the definition of the normal distribution, we obtain the equivalent problem:

$$argmin_{\beta,e} \frac{1}{2\sigma_g^2/m}\sum_j \beta_j^2 + \frac{1}{2\sigma_e^2}\sum_i e_i^2 + \mathrm{W} \quad \mathrm{s.t.}\ \ \mathrm{p}(\mathrm{X}\beta + \mathrm{e}) \le \mathrm{pt} \qquad (11)$$

where W is a quantity that does not depend on β or e, and can thus be ignored. This is a standard quadratic optimization problem, amenable to exact solution using standard convex optimization techniques[44]. Given the joint MAP of β and e, $\hat{l}$ is given by $\hat{l} = X\hat{\beta} + \hat{e}$.

Genetic MAP Estimator

The MAP of the effect sizes β can be found by maximizing their posterior likelihood. Using the same derivation as before, with the exception that e is integrated out, the quantity to maximize is given by

$$\varphi\left(\beta;\ 0, \frac{\sigma_g^2}{m}\mathrm{I}\right) \prod_{\mathrm{i}\in\mathrm{controls}} \Phi\left(\mathrm{t} - \mathrm{X_i^T}\beta;\ 0, \sigma_e^2\right) \prod_{\mathrm{i}\in\mathrm{cases}} \left(1 - \Phi\left(\mathrm{t} - \mathrm{X_i^T}\beta;\ 0, \sigma_e^2\right)\right). \quad (12)$$

Taking the logarithm and using the normal distribution definition, the quantity to maximize is

$$\sum_{\mathrm{i}\in\mathrm{controls}} \log \Phi\left(\mathrm{t} - \mathrm{X_i^T}\beta;\ 0, \sigma_e^2\right) + \sum_{\mathrm{i}\in\mathrm{cases}} \log\left(1 - \Phi\left(\mathrm{t} - \mathrm{X_i^T}\beta;\ 0, \sigma_e^2\right)\right) - \frac{1}{2\sigma_g^2/m}\sum_j \beta_j^2 + W \qquad (13)$$

where W is a quantity that does not depend on β and can thus be ignored. This problem is equivalent to Probit regression[45] with L2 regularization and a pre-specified offset term, and can thus be solved using standard techniques (Supplementary Note). Unlike typical uses of such models, here the regularization parameter is known in advance, given a value for σ_g^2.

The MAP $\hat{g}$ is given by $\hat{g} = X\hat{\beta}$, where $\hat{\beta}$ is the MAP of β. Given the MAP $\hat{g}$, $\hat{l}$ is determined by setting the entries of all cases with $\hat{g_\iota} < t$, and all controls with $\hat{g_\iota} > t$, to be equal to t. All other entries in $\hat{l}$ are equal to the corresponding entry in $\hat{g}$. This follows because e has a zero-mean normal distribution.

We opted to use the genetic MAP estimator, rather than the joint MAP estimator, because it is more suitable for liability estimation in the presence of related individuals. This greater suitability comes from the way LEAP handles related individuals, which consists of first excluding them from the model fitting stage, and then estimating their liabilities via the fitted model (see further details below). The joint MAP estimator minimizes the in-sample estimation error of the liabilities, because it directly fits the environmental component e of individuals participating in the fitting stage. In contrast, the genetic MAP estimator attempts to minimize the out-of-sample estimation error, because it integrates the environmental component e out and only fits the effect sizes β. The effect sizes β are later used to estimate liabilities for all individuals, including those that did not participate in the model fitting stage. Therefore, the genetic MAP estimator is more suitable for the purposes of LEAP.

We verified empirically that the genetic MAP estimator often yields more accurate estimates than either the MAP or the posterior mean estimator (Supplementary Note, Supplemental Figure S18 and Supplementary Table S4).

Dimensionality Reduction

A straightforward solution of the optimization problems presented above is difficult due to their high dimensionality, which is equal to the number of genotyped variants. Fortunately, the problems can be reformulated as lower dimensional problems, with dimensionality equal to the number of individuals.

The equivalence stems from the fact that the genotypes matrix X can be represented in terms of the eigenvectors of its covariance matrix alone. To see this, we rewrite Expression 13 as follows

$$f(X\beta) - \frac{1}{2\sigma_g^2/m}\beta^T\beta \qquad (14)$$

where $f(X\beta)$ is a function that depends on β only through the product $X\beta$. Consider the singular value decomposition (SVD) of X, given by

$$X = USV^T \qquad (15)$$

where U is the matrix of the eigenvectors of XX^T, and V is orthonormal. Denote $Z = US$ and $\beta_Z = V^T\beta$. Due to the orthonormality of V, the following equations hold.

$$\beta_Z{}^T\beta_Z = \beta^T\beta. \qquad (16)$$

$$Z\beta_Z = X\beta. \qquad (17)$$

Therefore, Expression 14 can be rewritten as

$$f(Z\beta_Z) - \frac{1}{2\sigma_g^2/m}\beta_Z{}^T\beta_Z. \qquad (17)$$

Denoting the number of individuals and genotyped variants by n and m, respectively, and assuming $m > n$ and that the columns of Z are ordered according to the magnitude of their respective eigenvalues, then all columns of Z except for the leftmost n ones are equal to zero. Consequently, the vector $Z\beta_Z$ depends only on the top n entries of the vector β_Z, and thus all the other entries can be set to zero.

We conclude that the quantity in Expression 13 can be maximized by considering only the non-zero components of the matrix Z and the vector β_Z, which have dimensionalities $n \times n$ and n, respectively. In contrast, the original formulation of the problem uses the matrix X and the vector β, which have dimensionalities $n \times m$ and m, respectively. The original effect sizes are given by $\beta = V\beta_Z$. However, they are not needed in practice, since the liabilities estimator can be computed using β_Z directly.

Finally, we note that when performing GWAS, the matrix Z is typically computed regardless of whether LEAP is employed, and is thus available at no further computational cost. This results from the close relation between the SVD of X and the eigendecomposition of the matrix XX^T. Namely, given the eigendecomposition $XX^T = US^2U^T$, the matrix Z is given by $Z = US$, where S is the matrix of the componentwise square roots of the entries of S^2. In GWAS, the eigendecomposition of XX^T is computed both when using an LMM[46] and when performing regression using principal component covariates[23], and is thus available for use in LEAP at no further computational cost.

Use in GWAS

LEAP uses liability estimates by treating them as observed continuous phenotypes in an LMM. Three difficulties that must be dealt with are accurate fitting of the LMM parameters, avoiding testing SNPs for association with the liability estimator that they helped estimate, and dealing with family relatedness. We now describe solutions to these difficulties.

The difficulty of parameter estimation stems from the non-normality of the liability under case-control sampling. This non-normality arises because in rare diseases, the majority of cases share a similar liability close to the cutoff. Parameter estimation can be suboptimal in such settings. The most important parameter that is fitted in LMMs is the variances ratio $\delta = \sigma_e^2/\sigma_g^2$. Given this parameter, all other parameters can be evaluated via closed form formulas[46]. There is a close connection between this parameter and the narrow-sense heritability, $h^2 = \sigma_g^2/\left(\sigma_g^2 + \sigma_e^2\right)$, expressed via $\delta = 1/h^2 - 1$. We therefore fit this parameter by estimating the heritability using the method of ref.[16], as described in the supplementary note.

A second difficulty arises because SNPs should not be tested for association with a liability estimator that they helped estimate. Otherwise the test statistic for these SNPs will be inflated, because they can always account for some of the liability variance. Similarly, SNPs in linkage disequilibrium with a tested SNP should also not participate in the liability estimation. To prevent such inflation, we estimate liabilities on a per-chromosome basis. For every chromosome, the liability is estimated using all SNPs except for the ones on the chromosome. The SNPs on the excluded chromosome are then tested for association using this liability estimator. We note that LMM-based GWAS typically compute the eigendecomposition of the covariance matrix on a per-chromosome basis as well, in order to prevent a SNP from incorrectly affecting the null likelihood (the phenomenon termed *proximal contamination*[10, 24]). LEAP can make use of these available eigendeompositions for dimensionality reduction - thus incurring no computational cost other than the liability estimation procedure itself.

A third difficulty arises when the data is confounded by family relatedness. The presence of related individuals can lead to biased effect size estimates, and consequently to a biased liability estimator. We deal with this difficulty by excluding related individuals from the parameter estimation stage of the MAP computation. We employ a greedy algorithm, where at each stage we exclude the individual having the largest number of related individuals with correlation coefficient >0.05. After fitting the model, we estimate liabilities for the excluded individuals as well. We note that population structure does not present similar problems, because it is naturally captured by top principal components[6, 17, 23], which are fitted in the MAP computation.

Data Simulation

All experiments reported in this paper are based on a uniform data generation procedure that can simulate different settings via a variety of parameters. In these simulations, each individual carried 60,100 SNPs that do not affect the phenotype, as well as 50-5000 causal SNPs with normally distributed effect sizes. Population structure was simulated via the Balding-Nichols model[47], which generates populations with genetic divergence measured via Wright's F_{ST}[28]. Family relatedness was simulated by generating various numbers of sib-pairs in one of the two populations, as in ref.[6]. To simulate ascertainment, we generated $3000/K$ individuals and a latent liability value for every individual, where K is the disease prevalence. We then determined the $1-K$ percentile of the liabilities, and generated new individuals until 50% of the sample had liabilities exceeding this cutoff[15]. A detailed description of the simulation procedure and its default parameters is provided in the supplementary note.

Availability

LEAP is available to download from http://bioinfo.cs.technion.ac.il/LEAP/leap.zip

Acknowledgements

This work was supported by the Israeli Science Foundation. This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113. The MS and UC data sets were filtered by Alexander Gusev.

Conflict of interest: C.L. and D.H. performed work on this manuscript while employed by Microsoft.

References

1. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* **42**, D1001-1006 (2014).
2. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-569 (2010).
3. Lee, S.H., Wray, N.R., Goddard, M.E. & Visscher, P.M. Estimating missing heritability for disease from genome-wide association studies. *American journal of human genetics* **88**, 294-305 (2011).
4. Gibson, G. Rare and common variants: twenty arguments. *Nat Rev Genet* **13**, 135-145 (2011).
5. Balding, D.J. A tutorial on statistical methods for population association studies. *Nat Rev Genet* **7**, 781-791 (2006).
6. Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* **11**, 459-463 (2010).
7. Fakiola, M. et al. Common variants in the HLA-DRB1-HLA-DQA1 HLA class II region are associated with susceptibility to visceral leishmaniasis. *Nat Genet* **45**, 208-213 (2013).
8. Sawcer, S. et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476**, 214-219 (2011).
9. Tsoi, L.C. et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet* **44**, 1341-1348 (2012).
10. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* **46**, 100-106 (2014).
11. Zaitlen, N. et al. Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genet* **8**, e1003032 (2012).
12. Pirinen, M., Donnelly, P. & Spencer, C.C. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nat. Genet.* **44**, 848-851 (2012).
13. Clayton, D. Link functions in multi-locus genetic models: implications for testing, prediction, and interpretation. *Genet Epidemiol* **36**, 409-418 (2012).
14. Mefford, J. & Witte, J.S. The Covariate's Dilemma. *PLoS Genet* **8**, e1003096 (2012).
15. Zaitlen, N. et al. Analysis of case-control association studies with known risk variants. *Bioinformatics* **28**, 1729-1737 (2012).
16. Golan, D. & Rosset, S. Narrowing the gap on heritability of common disease by direct estimation in case-control GWAS. *arXiv preprint arXiv:1305.5363* (2013).
17. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* **2**, e190 (2006).

18. McCulloch, C.E., Sciences, C.B.o.t.M. & Foundation, N.S. Generalized Linear Mixed Models. (Institute of Mathematical Statistics, 2003).

19. Dempster, E.R. & Lerner, I.M. Heritability of Threshold Characters. *Genetics* **35**, 212-236 (1950).

20. Rasmussen, C.E. & Williams, C.K.I. Gaussian Processes for Machine Learning. ((Massachusetts: MIT Press), 2005).

21. Hayes, B.J., Visscher, P.M. & Goddard, M.E. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res (Camb)* **91**, 47-60 (2009).

22. Hayeck, T. et al. Mixed Model with Correction for Case-Control Ascertainment Increases Association Power. *BioRxiv preprint* (2014).

23. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-909 (2006).

24. Listgarten, J. et al. Improved linear mixed models for genome-wide association studies. *Nature methods* **9**, 525-526 (2012).

25. Tucker, G., Price, A.L. & Berger, B. Improving the Power of GWAS and Avoiding Confounding from Population Stratification with PC-Select. *Genetics* **197**, 1045-1049 (2014).

26. Lippert, C. et al. The benefits of selecting phenotype-specific variants for applications of mixed models in genomics. *Scientific reports* **3**, 1815 (2013).

27. Widmer, C. et al. Further Improvements to Linear Mixed Models for Genome-Wide Association Studies. *Sci. Rep.* **4** (2014).

28. Wright, S. The genetical structure of populations. *Ann Eugenic* **15**, 323-354 (1949).

29. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).

30. Pritchard, J.K. & Przeworski, M. Linkage disequilibrium in humans: models and data. *American journal of human genetics* **69**, 1-14 (2001).

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

32. The Wellcome Trust Case Control Consortium Genome-wide association study of ulcerative colitis identifies three new susceptibility loci, including the HNF4A region. *Nat Genet* **41**, 1330-1334 (2009).

33. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *European journal of human genetics : EJHG* **19**, 807-812 (2011).

34. Hindorff, L.A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9362-9367 (2009).

35. Golan, D. & Rosset, S. Effective genetic-risk prediction using mixed models. *American journal of human genetics* **95**, 383-393 (2014).

36. Genz, A. & Bretz, F. Computation of Multivariate Normal and t Probabilities. (Springer, 2009).

37. Fusi, N., Lippert, C., Lawrence, N.D. & Stegle, O. Warped linear mixed models for the genetic analysis of transformed phenotypes. *Nat Commun* **5**, 4890 (2014).

38. Speed, D. & Balding, D.J. MultiBLUP: improved SNP-based prediction for complex traits. *Genome Res* **24**, 1550-1557 (2014).

39. Zhou, X., Carbonetto, P. & Stephens, M. Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genet* **9**, e1003264 (2013).

40. Loh, P.R. et al. Efficient Bayesian mixed model analysis increases association power in large cohorts. *BioRxiv preprint* (2014).

41. Segura, V. et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat Genet* **44**, 825-830 (2012).

42. Golan, D. & Rosset, S. Accurate estimation of heritability in genome wide studies using random effects models. *Bioinformatics* **27**, i317-323 (2011).

43. Kang, H.M. et al. Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709-1723 (2008).

44. Boyd, S.P. & Vandenberghe, L. Convex Optimization. (Cambridge University Press, 2004).

45. Hastie, T., Tibshirani, R. & Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. (Springer, 2009).

46. Lippert, C. et al. FaST linear mixed models for genome-wide association studies. *Nature methods* **8**, 833-835 (2011).

47. Balding, D.J. & Nichols, R.A. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica* **96**, 3-12 (1995).

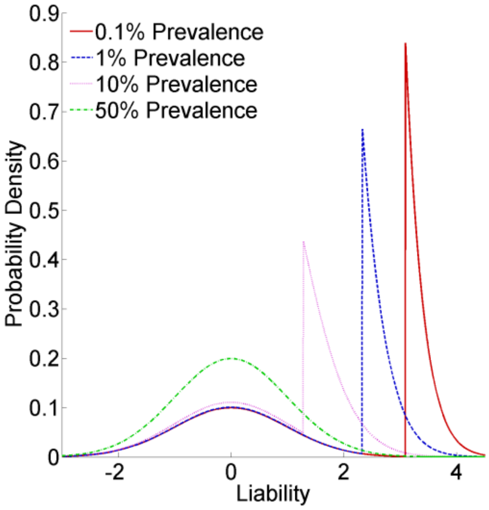
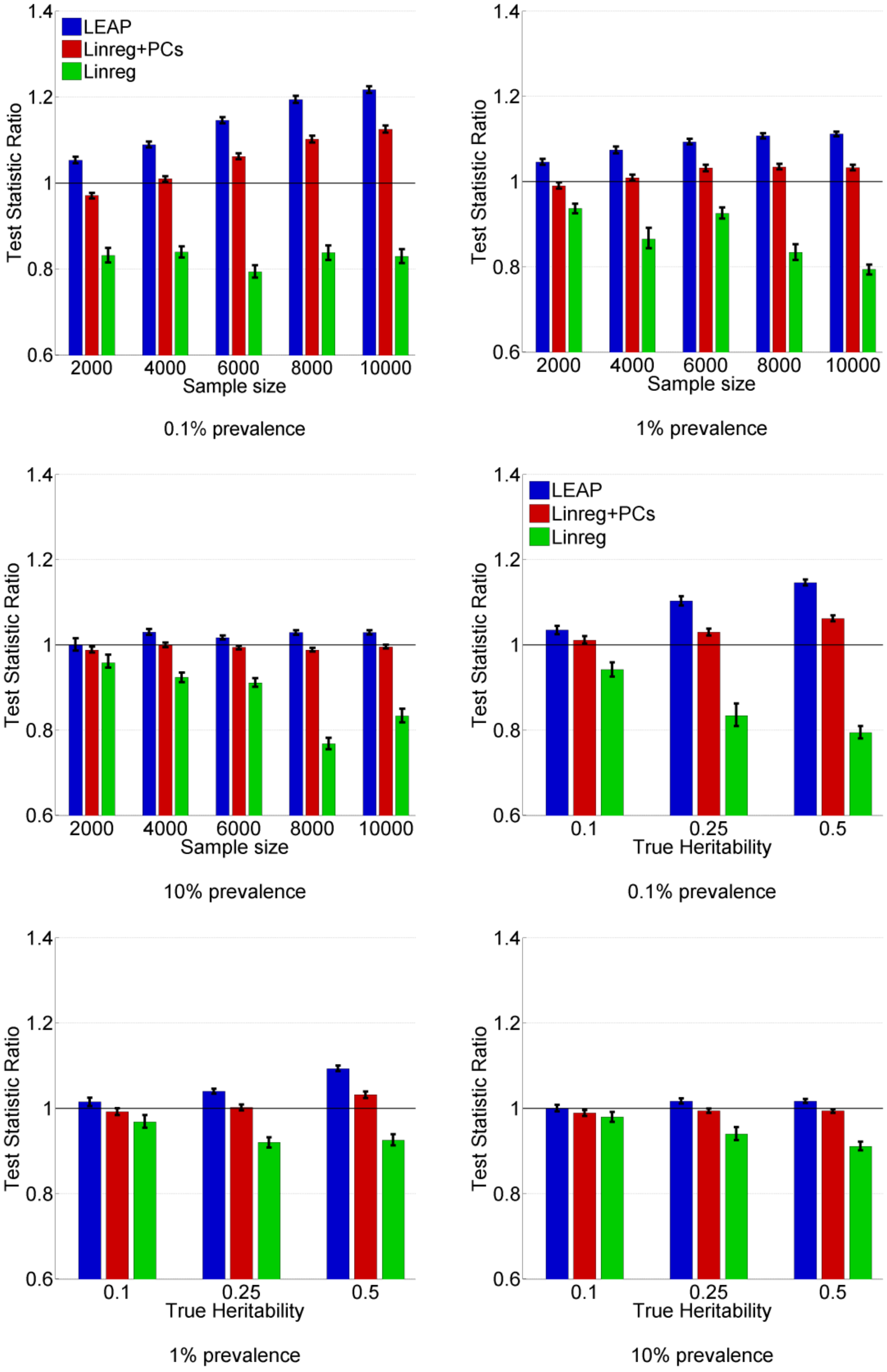


Figure 1: Liability distributions in balanced case-control data sets. Individuals with liability greater than the prevalence-specific cutoff are cases, and the remainder are controls. The liabilities of controls and of cases follow a zero-mean normal distribution, conditioned on being smaller or greater than the liability cutoff, respectively. The distribution of case liabilities becomes increasingly sharply peaked as prevalence decreases.



0.1% prevalence

1% prevalence

10% prevalence

0.1% prevalence

1% prevalence

10% prevalence

Figure 2: The mean ratio of normalized test statistics for causal SNPs between each evaluated method and an LMM, and its 95% confidence interval, under different sample sizes (top row) and heritability levels (bottom row). Larger mean ratios

indicate higher power. Values above the horizontal line indicate that a method has test statistics that are on average greater than that of an LMM.

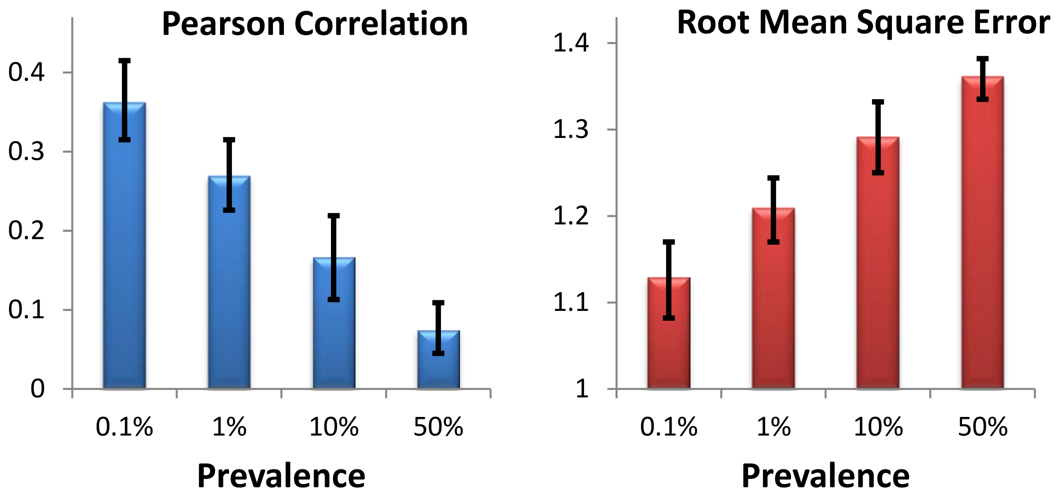


Figure 3: Similarity between the estimated and true liabilities of controls, for data sets with 6,000 individuals, and their 95% confidence intervals (computed via 10-fold cross validation for each data set, averaged over 10 data sets). The similarity measures shown are the Pearson correlation and the root mean square error, after normalizing the liabilities to have zero mean and unit variance.

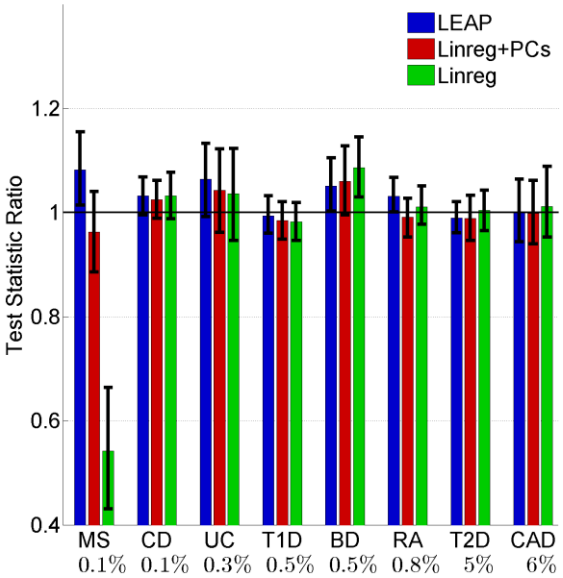


Figure 4: Analysis of real data sets. The values shown are the mean of the ratios of normalized test statistics of tag SNPs between each evaluated method and an LMM, and its 95% confidence interval. A higher mean ratio indicates higher power. Values above the horizontal line indicate that a method has test statistics that are on average greater than that of an LMM. The diseases shown are multiple sclerosis (MS), Crohn's disease (CD), ulcerative colitis (UC), type 1 diabetes (T1D), bipolar disorder (BD), rheumatoid arthritis (RA), type 2 diabetes (T2D), and coronary artery disease (CAD). The prevalence of each disease is shown below its name. Results for hypertension are omitted because its analysis contains only three tag SNPs with P value smaller than 0.01, leading to an unreliable estimate of the statistics ratio.