



# Retrospective Mixed Model and Propensity Score Methods for Case Control Data

## Citation

Hayeck, Tristan. 2015. Retrospective Mixed Model and Propensity Score Methods for Case Control Data. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

## Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:23845417>

## Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

## Share Your Story

The Harvard community has made this article openly available.  
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

# Retrospective Mixed Model and Propensity Score Methods for Case Control Data

A dissertation presented

by

Tristan Jonathan Hayeck

to

The Biostatistics Department

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biostatistics

Harvard University

Cambridge, Massachusetts

June 2015

© 2015 Trista Jonathan Hayeck

All rights reserved.

## Retrospective Mixed Model and Propensity Score Methods for Case Control Data

## Abstract

In chapter one a Liability Threshold Mixed Linear Model (LTMLM) association statistic is introduced for ascertained case-control studies that increases power vs. existing mixed model methods for diseases with low prevalence, with a well-controlled false-positive rate. 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 estimated using a multivariate Gibbs sampler. In a Welcome Trust Case Control Consortium 2 (WTCCC2) multiple sclerosis data set LTMLM attained a 4.3% improvement ( $P=0.005$ ) in chi-square statistics (vs. existing mixed model methods) at 75 known associated SNPs. Family-biased ascertainment is considered in chapter 2, where cases and controls are ascertained non-randomly with respect to family relatedness. We introduce a family based association statistic (LT-Fam) that is robust to this problem. For type 2 diabetes cases and controls (in the Jackson Heart Study) we down-sampled to increase relatedness among cases and observed: ATT was inflated and MLM was deflated, while LT-Fam was properly calibrated. Finally, in chapter three, we propose a 2-Step Bayesian Model Averaging (2-Step BMA) method with Propensity Score (PS) adjustment that targets the primary treatment of interest characterizing the treatment effect while controlling for a high dimensional set of unknown confounders including metabolites and other epidemiological factors. This method improves on existing methods by averaging over the entire model space of both the treatment and outcome

models to control for confounding while targeting treatment effect and without need of an arbitrary number of confounders to include a priori.

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

## Introduction

Mixed model association statistics are a widely used approach to correct for population structure and cryptic relatedness in genome-wide association studies (GWAS)<sup>1-11</sup>. However, recent work shows that existing mixed model association statistics suffer a loss in power relative to standard logistic regression in ascertained case-control studies<sup>11</sup>. It is widely known that appropriate modeling of case-control ascertainment can produce substantial increases in power for case-control studies with fixed-effect covariates<sup>12-14</sup>, 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 sampler<sup>15</sup>. 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.

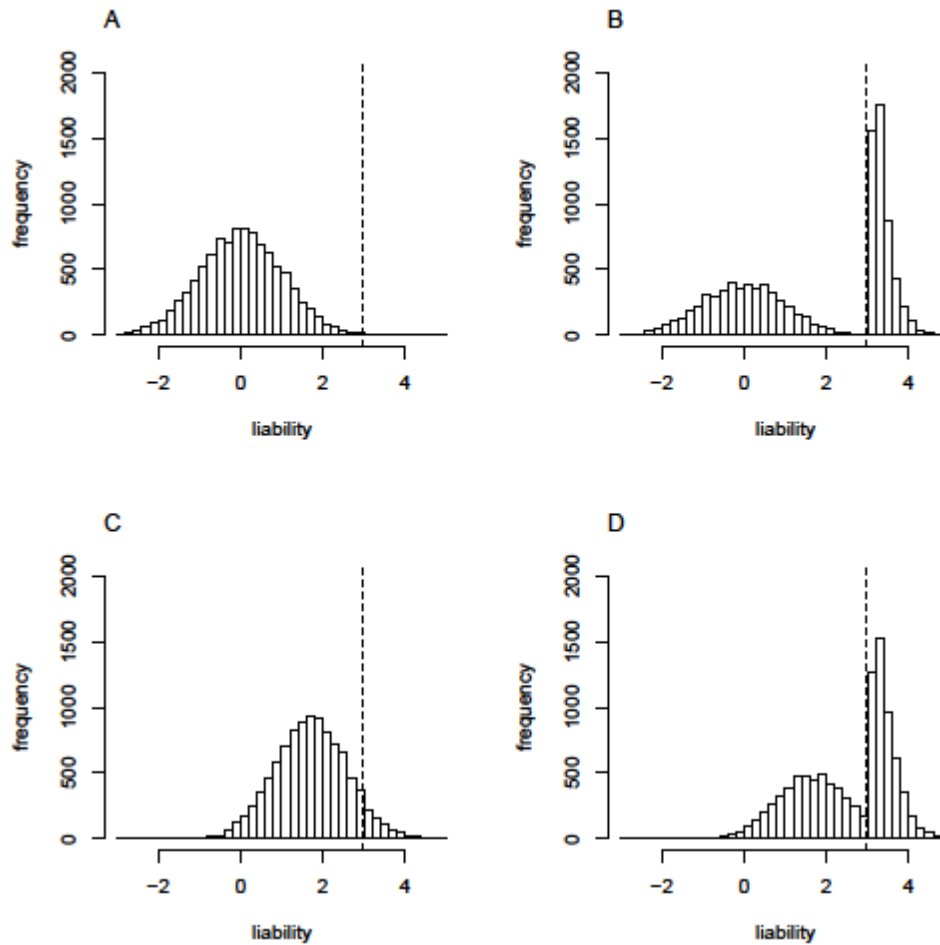
The LTMLM statistic provides an increase in power in simulations of ascertained case-control studies of diseases with low prevalence based on either simulated or real genotypes. In a WTCCC2 multiple sclerosis data set with >10,000 samples, LTMLM was correctly calibrated and attains a 4.3% improvement ( $P=0.005$ ) 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 methods<sup>11</sup> 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*).



**Figure 1.1. 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.<sup>17</sup>). 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.

The toy example in Figure 1.1 provides an illustration of how genetic relatedness to a disease case can increase an individual's PML. In Figure 1.1a and 1.1b, 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 1.1c and 1.1d, 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.<sup>4</sup> 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 ( $h_g^2$ )<sup>16</sup> 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 S1.1.

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 prevalence<sup>17</sup> (Figure S1.1).

Standard mixed model association methods generally estimate  $h^2$  from a genetic relationship matrix (GRM) and phenotypes using restricted maximum likelihood (REML)<sup>4; 11</sup>. Genotypic data is used to build a GRM (excluding the candidate SNP<sup>11</sup>):

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

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 GRM<sup>18; 19</sup>:

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

where  $\pi_i$  denotes the case-control status of individual  $i$  and  $\widehat{\boldsymbol{\theta}}_{ik}$  is the genetic relatedness of individuals  $i$  and  $k$ . This gives an estimate on the observed scale which is then transformed to the liability scale<sup>20</sup>:

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

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 sample<sup>20</sup>.

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

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

where  $\widehat{\boldsymbol{\theta}}$  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. <sup>20</sup>, 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 status<sup>20</sup>.

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 work<sup>15</sup> (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:

Initialization: for each individual  $j$ , (8)

$\varphi_i = \text{PML}_{\text{uni,case}}$  if  $\pi_i=1$  or  $\varphi_i = \text{PML}_{\text{uni,control}}$  if  $\pi_i=0$

For each MCMC iteration  $n$

For each individual  $i$

Sample  $\varphi_i$  from the constrained conditional univariate normal distribution  
 $L(\varphi_i) \sim \exp(-\varphi^T \mathbf{V}^{-1} \varphi / 2)$  and constraint  $\varphi_i \geq t$  if  $\pi_i=1$ ,  $\varphi_i < t$  if  $\pi_i=0$   
 (where  $\varphi_{\neq i}$  are fixed)

We use 100 burn-in iterations followed by 1,000 additional MCMC iterations. We estimate the  $\text{PML}_{\text{multi}}$  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  $\varphi_i$  is sampled.

#### *LTMLM Association Statistic*

The LTMLM association statistic is calculated using  $\text{PML}_{\text{multi}}$ . 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  $(\boldsymbol{\varphi}, \mathbf{x})$  and derive the joint likelihood for these combined terms. The covariance of  $\boldsymbol{\varphi}$  and  $\mathbf{x}$  between individual  $i$  and  $k$  is:

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

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  $\widehat{\boldsymbol{\theta}}$  estimated from the data.)

The variance of  $(\boldsymbol{\varphi}, \mathbf{x})$  as a function of effect size  $\beta$  is:

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

thus

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

where both of these matrices are  $2N$  by  $2N$ . (We note that the product of the matrices in equation 10 and equation 11 is  $\begin{pmatrix} \mathbf{I} + O(\beta^2) & 0 \\ 0 & \mathbf{I} + O(\beta^2) \end{pmatrix}$ , whose difference from the identity contains only  $O(\beta^2)$  terms.)

The joint likelihood of the liability and genotypes are distributed as a multivariate normal  $N(0, \mathbf{C}(\beta))$ , and thus

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

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  $\beta$ , so the terms with  $\mathbf{V}$  alone drop out when we take the derivative:

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

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

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

This leads to the score statistic:

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

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

In equations 9-15 the liability was assumed to be known, for simplicity. 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:

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

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

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

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

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

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

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 LTMMLM score statistic generalizes to non-normally distributed genotypes<sup>21-23</sup>. The overall computational cost of computing the LTMMLM statistic is  $O(MN^2)$  when  $M > N > \#iterations$  (Table S1.2). We have fixed the number of iterations at 100 burn-in iterations followed by 1,000 additional iterations.

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<sup>4; 6; 11</sup>.

### *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. All simulations included  $M$  candidate SNPs and an independent set of  $M$  SNPs used to calculate the GRM (so that candidate SNPs are not included in the GRM). For each scenario a set of 100 simulations were run. We set 10 candidate SNPs and 10 GRM SNPs to be causal in simulations with  $N=1,000$  samples, and set 50 candidate SNPs and 50 GRM SNPs to be causal in simulations with  $N=5,000$  samples, ensuring that causal SNPs have similar average chi-square statistics independent of  $M$  and  $N$ . The resulting 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 Armitage Trend Test (ATT), Logistic Regression (LogR), MLM, and LTMLM statistics (see Table 1.1). MLM statistics were computed using the GCTA-LOCO statistic described in ref. <sup>11</sup>, with the heritability parameters estimated using the GCTA software<sup>24</sup>. We evaluated performance using average chi-square statistics at causal, null, and all markers,  $\lambda_{GC}$  at all markers (median chi-square divided by 0.455)<sup>25</sup>, and proportion of causal and null markers that were significant at P-value thresholds of 0.05, 0,001,  $1 \times 10^{-6}$  and  $5 \times 10^{-8}$ .

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

**Table 1.1. List of association statistics.** We list properties of the Armitage Trend Test (ATT), standard mixed model association statistic (MLM), and proposed statistic (LTMLM).  $\pi^*$  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)

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  $[0.1, 0.9]$ .

To test the impact of the generative distribution, the underlying distribution was simulated using a logit model instead of a liability threshold model. The causal 10 causal candidate SNPs were simulated with alternating fixed effect size of  $\beta=0.4$  or  $\beta=-0.4$ . Then, case-control phenotypes were generated from a binomial distribution where the probability of being a



case was ( $case$ ) =  $\frac{1}{1+\exp(-[c+\beta x])}$ , shifted by the affine term,  $c$ , based on the desired disease prevalence.

### *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 previously<sup>11</sup>. 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, LogR, MLM and LTMLM statistics as described above.

### *WTCCC2 Genotypes and MS Phenotypes*

Finally, we analyzed WTCCC2 individuals with ascertained case-control phenotypes for MS<sup>11</sup>, a disease with a prevalence of around 0.1%. As in previous work, we assume that the disease prevalence is known based on external epidemiological literature<sup>20; 26; 27</sup>. For the WTCCC2 MS data, we used a threshold of 3.0, corresponding to a disease prevalence of 0.1%.<sup>26</sup> We computed ATT, LogR, MLM and LTMLM statistics as described above. Although the underlying MS study was appropriately matched for ancestry<sup>28</sup>, 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 S1.2). 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 MS<sup>11</sup>. 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, LogR, MLM and LTMLM. We report average chi-square statistics at causal, null, and all markers and  $\lambda_{GC}$  at all markers in Table 1.2, and the

Table 1.2

<b>N</b>	<b>M</b>	<b>Prev</b>	<b>Set</b>	<b>Statistic</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
5000	5000	50%	causal	average	16.492(0.325)	16.399(0.321)	16.88(0.332)	16.867(0.332)
			null	average	0.988(0.002)	0.988(0.002)	0.990(0.002)	0.989(0.002)
			all	average	1.143(0.004)	1.142(0.004)	1.148(0.004)	1.148(0.004)
			all	$\lambda_{GC}$	1.010(0.003)	1.010(0.003)	1.014(0.003)	1.014(0.003)
		25%	causal	average	18.637(0.388)	18.509(0.383)	19.014(0.396)	19.056(0.398)
			null	average	1.000(0.002)	1.000(0.002)	1.000(0.002)	1.001(0.002)
			all	average	1.177(0.005)	1.175(0.005)	1.18(0.005)	1.181(0.005)
			all	$\lambda_{GC}$	1.012(0.004)	1.012(0.004)	1.013(0.004)	1.013(0.004)
		10%	causal	average	25.235(0.501)	25.014(0.492)	25.386(0.506)	25.778(0.514)
			null	average	0.993(0.002)	0.992(0.002)	0.991(0.002)	0.992(0.002)
			all	average	1.235(0.006)	1.233(0.006)	1.235(0.006)	1.24(0.007)
			all	$\lambda_{GC}$	1.005(0.003)	1.005(0.003)	1.007(0.003)	1.008(0.003)
		1%	causal	average	45.376(0.878)	44.682(0.852)	42.594(0.825)	<b>46.691(0.913)</b>
			null	average	1.000(0.002)	0.999(0.002)	0.990(0.002)	1.000(0.002)
			all	average	1.444(0.011)	1.436(0.011)	1.406(0.01)	1.457(0.011)
			all	$\lambda_{GC}$	1.020(0.003)	1.020(0.003)	1.011(0.003)	1.019(0.003)
		0.1%	causal	average	68.648(1.301)	67.099(1.248)	56.303(1.082)	<b>70.81(1.364)</b>
			null	average	1.000(0.002)	1.000(0.002)	0.918(0.002)	1.000(0.002)
			all	average	1.677(0.016)	1.661(0.016)	1.472(0.013)	1.698(0.017)
			all	$\lambda_{GC}$	1.026(0.003)	1.026(0.003)	0.942(0.005)	1.025(0.003)

Table 1.2 (Continued)

<b>N</b>	<b>M</b>	<b>Prev</b>	<b>Set</b>	<b>Statistic</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
5000	50000	50%	causal	average	16.624(0.331)	16.529(0.327)	16.673(0.332)	16.69(0.333)
			null	average	1.000 (0.001)	0.999(0.001)	1.000 (0.001)	1.000 (0.001)
			all	average	1.015(0.001)	1.015(0.001)	1.016(0.001)	1.015(0.001)
			all	$\lambda_{GC}$	1.003(0.001)	1.002(0.001)	0.999(0.001)	0.999(0.001)
		25%	causal	average	18.965(0.37)	18.843(0.366)	19.03(0.372)	19.04(0.372)
			null	average	1.003(0.001)	1.003(0.001)	1.003(0.001)	1.003(0.001)
			all	average	1.021(0.001)	1.02(0.001)	1.021(0.001)	1.021(0.001)
			all	$\lambda_{GC}$	1.006(0.001)	1.006(0.001)	1.003(0.001)	1.003(0.001)
		10%	causal	average	23.71(0.444)	23.528(0.437)	23.868(0.449)	23.91(0.45)
			null	average	1.001(0.001)	1.000 (0.001)	1.001(0.001)	1.001(0.001)
			all	average	1.023(0.001)	1.023(0.001)	1.024(0.001)	1.024(0.001)
			all	$\lambda_{GC}$	1.007(0.001)	1.007(0.001)	1.005(0.001)	1.004(0.001)
		1%	causal	average	46.683(0.883)	45.969(0.859)	46.44(0.881)	47.368(0.905)
			null	average	0.999(0.001)	0.999(0.001)	0.999(0.001)	0.999(0.001)
			all	average	1.045(0.001)	1.044(0.001)	1.045(0.001)	1.045(0.001)
			all	$\lambda_{GC}$	1.004(0.001)	1.004(0.001)	1.001(0.001)	1(0.001)
		0.1%	causal	average	67.059(1.278)	65.561(1.225)	65.232(1.251)	<b>68.618(1.333)</b>
			null	average	0.999(0.001)	0.999(0.001)	0.999(0.001)	0.999(0.001)
			all	average	1.065(0.002)	1.063(0.002)	1.063(0.002)	1.067(0.002)
			all	$\lambda_{GC}$	1.004(0.001)	1.004(0.001)	1.000 (0.001)	1.000 (0.001)

**Table 1.2. Results on simulated genotypes and simulated phenotypes.** We report average  $\chi^2$  statistics across 100 simulations for each individual scenario (standard errors in parentheses).  $N$  is the number of individuals and  $M$  is the number of SNPs. 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%. In bold are the settings where LTMLM demonstrates at least a 5% improvement over MLM in  $\chi^2$  statistics at causal markers.

proportion of causal and null markers that were significant at various P-value thresholds in Table S1.3. The LTMLM statistic outperforms the ATT, LogR and MLM statistics for diseases with low prevalence. Improvements in average chi-square statistics at causal markers (which have a natural interpretation as the increase in effective sample size) were larger than improvements in power to detect an association at a given P-value threshold, likely due to the variable (normally

distributed) effect sizes in this simulation (see below). For LTMLM vs. MLM at disease prevalences of 0.1%, 26% and 5% improvements in average chi-square statistics at causal markers 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  $N/M$  (Tables S1.3 and S1.4). Simulations with population structure demonstrate similar results, but with inflation in the ATT statistic as expected (Tables S1.5 and S1.6).

The MLM statistics were calculated using an  $h^2$  parameter estimated using Restricted Maximum Likelihood Methods (REML)<sup>4</sup>, 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<sup>18;20</sup> (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 1.3 and Table S1.7), consistent with previous work (see ref.<sup>29</sup> and Supp Table 9 of ref.<sup>11</sup>). Population structure resulted in bias of both REML and HE-regression estimates of  $h^2$ , with higher bias for the REML estimates (Table S1.8).

		Liability			Observed	
N	M	Prevalence	H-E	REML	H-E	REML
5000	5000	50%	0.255(0.006)	0.253(0.006)	0.162(0.004)	0.161(0.004)
		25%	0.24(0.005)	0.236(0.004)	0.172(0.003)	0.17(0.003)
		10%	0.24(0.004)	0.23(0.004)	0.228(0.004)	0.219(0.004)
		1%	0.25(0.004)	0.21(0.002)	0.453(0.007)	0.381(0.004)
		0.1%	0.253(0.003)	0.158(0.000)	0.719(0.009)	0.449(0.001)
	50000	50%	0.272(0.009)	0.274(0.009)	0.173(0.006)	0.174(0.006)
		25%	0.246(0.010)	0.252(0.010)	0.177(0.007)	0.181(0.007)
		10%	0.225(0.004)	0.231(0.004)	0.214(0.004)	0.219(0.004)
		1%	0.246(0.004)	0.241(0.004)	0.445(0.008)	0.437(0.008)
		0.1%	0.259(0.004)	0.241(0.004)	0.734(0.012)	0.684(0.01)

**Table 1.3. Heritability parameter estimates on simulated genotypes and phenotypes.** We report results on both liability and observed scales averaged over 100 simulations. The true  $h^2$  explained by the SNPs used to build the GRM is 25% on the liability scale for all simulations. These results correspond to the same sets of simulations in table 2.

These biases do not inflate LTMLM or MLM statistics under the null (Tables S1.5 and S1.6).

We note that previous work has shown that running MLM using the correct  $h^2$  parameter does not ameliorate the loss in power for MLM<sup>11</sup>.

We also evaluated performance in settings where the liability threshold (equivalently, the disease prevalence) is mis-specified (Tables S1.9 and S1.10). The LTMLM statistic remains properly calibrated under the null, and continues to outperform the MLM statistic as the impact of mis-specifying the liability threshold is small. Mis-specifying the liability threshold leads to bias in liability-scale heritability estimates, due to the inaccurate conversion from observed scale to liability scale (Table S1.11).

Finally, we evaluated performance when phenotypes were generated using a logit model instead of a liability threshold model, using a fixed effect size for causal candidate SNPs (see Materials and Methods). At low disease prevalence, we observed improvements for LTMLM

both in average chi-square statistics at causal markers (Table S1.12) and in power to detect an association at a given P-value threshold (Table S1.13); improvements in power depend heavily on the distribution of causal effect sizes, and are larger in simulations with fixed causal effect sizes than in simulations with variable causal effect sizes.

*Simulations: WTCCC2 Genotypes and Simulated Phenotypes*

We next conducted simulations using real WTCCC2 genotypes and simulated ascertained case-control phenotypes (see Materials and Methods)<sup>11; 28</sup>. 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.

M	Prev	Set	Statistic	ATT	ATT+PCs	LogR	LogR +PCs	MLM	LTMLM
1000	50%	causal	average	16.234(0.723)	15.775(0.705)	15.566(0.674)	15.14(0.658)	17.48(0.780)	17.412(0.775)
		null	average	1.444(0.011)	1.425(0.010)	1.433(0.010)	1.415(0.010)	1.472(0.011)	1.471(0.011)
		all	average	1.592(0.014)	1.569(0.013)	1.574(0.013)	1.552(0.013)	1.632(0.015)	1.630(0.015)
		all	$\lambda_{GC}$	1.214(0.017)	1.224(0.017)	1.214(0.017)	1.223(0.017)	1.234(0.017)	1.226(0.016)
	25%	causal	average	19.277(0.771)	18.581(0.754)	18.47(0.721)	17.822(0.706)	20.493(0.820)	20.642(0.831)
		null	average	1.551(0.012)	1.507(0.011)	1.538(0.012)	1.495(0.011)	1.571(0.012)	1.577(0.013)
		all	average	1.728(0.015)	1.678(0.014)	1.707(0.015)	1.658(0.014)	1.761(0.016)	1.768(0.016)
		all	$\lambda_{GC}$	1.256(0.016)	1.245(0.017)	1.255(0.016)	1.245(0.017)	1.241(0.015)	1.243(0.015)
	10%	causal	average	22.838(0.961)	21.406(0.878)	21.618(0.881)	20.336(0.808)	23.865(1.022)	24.661(1.064)
		null	average	1.664(0.014)	1.583(0.012)	1.647(0.014)	1.568(0.012)	1.668(0.014)	1.695(0.015)
		all	average	1.876(0.018)	1.781(0.016)	1.846(0.017)	1.756(0.015)	1.890(0.019)	1.925(0.019)
		all	$\lambda_{GC}$	1.271(0.016)	1.285(0.015)	1.270(0.016)	1.284(0.015)	1.251(0.017)	1.270(0.016)
10000	50%	causal	average	16.898(0.702)	16.815(0.698)	16.229(0.659)	16.156(0.655)	17.279(0.725)	17.26(0.725)
		null	average	1.078(0.002)	1.083(0.002)	1.074(0.002)	1.078(0.002)	1.078(0.002)	1.076(0.002)
		all	average	1.094(0.002)	1.099(0.002)	1.089(0.002)	1.094(0.002)	1.095(0.002)	1.092(0.002)
		all	$\lambda_{GC}$	1.039(0.005)	1.047(0.005)	1.039(0.005)	1.046(0.005)	1.039(0.005)	1.035(0.005)
	25%	causal	average	17.573(0.726)	17.293(0.696)	16.856(0.679)	16.61(0.654)	17.976(0.749)	18.077(0.758)
		null	average	1.076(0.002)	1.078(0.002)	1.071(0.002)	1.073(0.002)	1.074(0.002)	1.073(0.002)
		all	average	1.092(0.002)	1.094(0.002)	1.087(0.002)	1.089(0.002)	1.091(0.002)	1.090(0.002)
		all	$\lambda_{GC}$	1.040(0.004)	1.045(0.004)	1.040(0.004)	1.044(0.004)	1.040(0.004)	1.039(0.004)
	10%	causal	average	24.379(1.026)	24.116(1.014)	23.127(0.944)	22.894(0.934)	24.987(1.071)	25.399(1.091)
		null	average	1.112(0.002)	1.115(0.002)	1.107(0.002)	1.110(0.002)	1.108(0.002)	1.111(0.002)
		all	average	1.136(0.003)	1.138(0.003)	1.129(0.002)	1.132(0.002)	1.131(0.003)	1.135(0.003)
		all	$\lambda_{GC}$	1.051(0.005)	1.059(0.004)	1.050(0.005)	1.058(0.004)	1.044(0.004)	1.047(0.004)

**Table 1.4. Results on real genotypes and simulated phenotypes.** We report average  $\chi^2$  statistics.  $M$  is the number of SNPs, and sample size is fixed at 500 cases and 500 controls.



Results for 1,000 and 10,000 SNPs ( $M$ ) are displayed in Tables 1.4, S1.14, and S1.15, with sample size fixed at 500 cases and 500 controls. 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 1.5. The H-E regression estimates are unbiased, but the REML estimates are again downwardly biased at lower prevalence and large  $N/M$ .

M	Prevalence	Liability		Observed	
		H-E	REML	H-E	REML
1000	50%	0.259(0.013)	0.252(0.010)	0.165(0.008)	0.161(0.006)
	25%	0.241(0.010)	0.238(0.008)	0.173(0.007)	0.171(0.006)
	10%	0.245(0.011)	0.242(0.007)	0.233(0.010)	0.230(0.007)
10000	50%	0.236(0.014)	0.245(0.013)	0.150(0.009)	0.156(0.008)
	25%	0.250(0.014)	0.264(0.013)	0.180(0.010)	0.190(0.010)
	10%	0.259(0.012)	0.261(0.009)	0.246(0.011)	0.248(0.009)

**Table 1.5. Heritability parameter estimates on real genotypes and simulated phenotypes.** These results are from the same simulations used to generate Tables 4, S14, and S15. 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.

*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<sup>11;28</sup>(see Materials and Methods). We compared ATT, ATT with 5 PCs (ATT+PCs),<sup>22</sup> LogR, LogR+PCs, MLM and LTMLM. We evaluated calibration using the average  $\chi^2$  and  $\lambda_{GC}$  over all SNPs; we note that the average  $\chi^2$  and  $\lambda_{GC}$  are expected to be greater than 1 due to polygenic effects<sup>11;30</sup>. We believe that LTMLM is effective in correcting for confounding, and that a higher value of  $\lambda_{GC}$  for LTMLM vs. MLM is likely due to true polygenic signal, reflecting the higher power of LTMLM vs. MLM.

We evaluated power using the average  $\chi^2$  over the 75 published SNPs (Table 1.6) and the proportion of published SNPs that were significant at various P-value thresholds (Table S1.16). We evaluated power using the average  $\chi^2$  over the 75 published SNPs (Table 1.6) and the proportion of published SNPs that were significant at various P-value thresholds (Table S1.16).

Category	Metric	ATT	ATT+PCs	LogR	LogR+PCs	MLM	LTMLM
Published	Mean	11.661 (1.169)	9.98(0.984)	11.619 (1.161)	9.871 (0.965)	9.919 (0.974)	10.587 (1.017)
Genome Wide	Mean	1.379 (0.003)	1.152 (0.003)	1.378 (0.003)	1.142 (0.003)	1.144(0.003)	1.172 (0.003)
Genome Wide	$\lambda_{GC}$	1.343	1.125	1.343	1.115	1.115	1.141
Published	Mean/ $\lambda_{GC}$	8.695	8.880	8.663	8.860	8.905	9.292

**Table 1.6. Results on WTCCC2 MS data set.** We report going down the rows: the average  $\chi^2$  over 75 published SNPs, genome-wide including 360,557 SNPs, the  $\lambda_{GC}$  Genome Wide, and then the average across 75 published SNPs after normalizing by the  $\lambda_{GC}$ . All results are based on analysis of 10,034 individuals (see main text).

The LTMLM method performed best, with a 4.3% improvement in average chi-square statistics scaled by  $\lambda_{GC}$  vs. MLM (jackknife  $P=0.005$ ; see Materials and Methods) and an even larger improvement versus ATT and ATT+PCs, consistent with simulations (Tables 1.2 and S1.3). LTMLM also detected 56/75 known associations as nominally significant ( $P<0.05$ ) after  $\lambda_{GC}$  correction vs. 53/75 for MLM, although this difference is not statistically significant. Similar

results are obtained when calibrating association statistics via LD Score regression<sup>31</sup> (Table S1.17). A perfectly matched data set with 4,094 MS cases and 4,094 controls yielded a similar improvement for LTMLM vs. MLM (Table S1.18). 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 S1.19). 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 S1.18).

## **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. We emphasize that the improvement applies to ascertained case-control studies of diseases with low prevalence. We note that logistic and linear regression generally produce similar results, and logistic mixed model score tests that do not explicitly model case-control ascertainment are likely to produce results similar to standard linear mixed model methods.

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 S1.2). 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<sup>12; 20</sup>. 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 alternative<sup>31</sup>. Third, we did not consider ascertained case-control studies in family data sets, which also represents a future direction. The LTMLM score statistic in its current form is appropriate for association testing in population case-control samples with low levels of relatedness. In family data sets, other approaches to calibration such as LD Score regression could potentially be explored<sup>31</sup>. Fourth, the method relies on the assumption of an underlying normally distributed liability. Although this assumption is widely accepted by many geneticists<sup>32; 33</sup>, and the method also performs well under a different generative model (Tables S1.12 and S1.13), further work on whether case-control traits are accurately modeled using normally distributed liabilities is warranted. Fifth, the method does not estimate odds ratios; in this respect the method is similar to other mixed model association methods.<sup>4; 5; 8; 11</sup> However, liability-scale effect sizes can be converted to odds ratios.<sup>13</sup> Sixth, LTMLM requires running time  $O(MN^2)$  when  $M > N > \text{\#MCMC iterations}$ , analogous to standard mixed model association methods. This may be computationally intractable in very large data sets. We are developing much faster mixed model methods<sup>34</sup>, 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. Seventh, potential application of the LTMLM statistic to rare variant data sets is not considered here, and remains as a future direction. Finally, our methods could potentially be extended to multiple traits<sup>7; 27; 35</sup>.

## 2. Mixed Model Association with Family-Biased Case-Control Ascertainment

### Introduction

Mixed models have become the tool of choice for genetic association studies<sup>4; 11; 34</sup>; however, existing mixed model methods may be poorly calibrated or underpowered in settings of family sampling bias<sup>36</sup> and/or case control ascertainment. In our previous work<sup>37</sup>, we introduced a liability threshold based mixed model association statistic (LTMLM) that addresses the power loss of standard mixed model methods under case-control ascertainment. Here, we consider studies in which cases and controls are ascertained non-randomly with respect to family relatedness, such as in a discordant sibling study. We refer to this as family-biased case-control ascertainment, or simply family-biased ascertainment.

Previous work has shown that family-biased ascertainment can severely bias heritability estimates<sup>38</sup>, and we show here that it also impacts mixed model association statistics. We introduce a family based association statistic, LT-Fam, that is robust to this problem. Similar to LTMLM, LT-Fam is computed from posterior mean liabilities (PML) under a liability threshold model conditional on every individual's case-control status and the disease prevalence. However, LTMLM is susceptible to miscalibration under family-biased ascertainment, due to biased narrow-sense heritability estimation and calibration based on phenotypic covariance. The LT-Fam statistic is constructed to specifically address family-biased ascertainment, using published narrow sense heritability estimates as well as properly controlling for relatedness.

The LT-Fam statistic demonstrates proper calibration and is robust to family-biased ascertainment. We compared LT-Fam to existing Armitage Trend Test (ATT), MLM and

LTMLM statistics. MLM has been found to perform at least as well as other methods in family-based association studies<sup>36</sup>. Other noteworthy family-based association methods statistics as MQLS<sup>21</sup> and ROADTRIPS<sup>39</sup> are applicable to case-control data; however, they do not explicitly model both case-control ascertainment bias and family-biased ascertainment (see Discussion).

First, the LT-Fam statistic was compared to ATT and MLM in different settings of family biased ascertainment by simulating sib pairs under different levels of discordant and concordant sampling. Simulating all concordant siblings settings of low disease prevalence LT-Fam is properly calibrated whereas MLM appears deflated while both LTMLM and ATT appear to be inflated (average  $\chi^2$  of 1.502). Simulating all discordant siblings both ATT and MLM are deflated, again LT-Fam again appears close to properly calibrated. Then, the LT-Fam statistic was compared against other methods using the Jackson Heart Study (JHS) type two diabetes (T2D) cases and controls. Then subjects were intentionally down sampled, to induce family-biased ascertainment, increasing the relative relatedness among cases of T2D. After down sampling the LT-Fam statistic was properly calibrated whereas other statistics demonstrated inflation or deflation, consistent with simulations.

## **Materials and Methods**

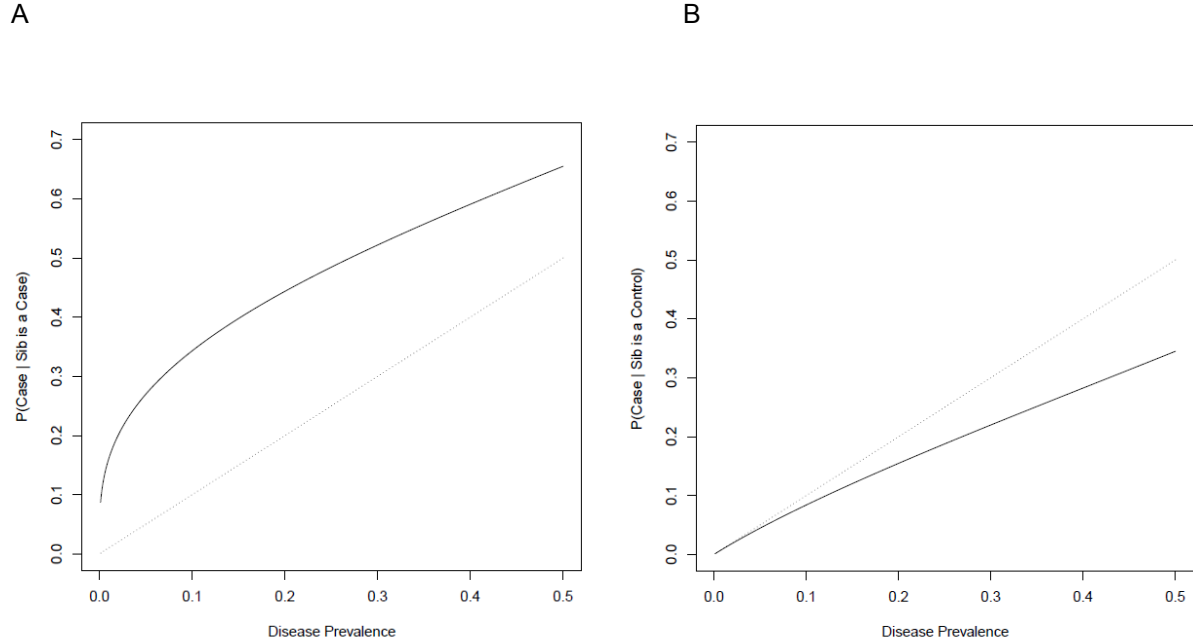
### *Overview of Method*

The LT-Fam method consists of three main steps. First, a genetic relationship matrix (GRM) is calculated and then restricted to include only related individuals by changing entries below a threshold to 0. The narrow sense heritability is then either assumed to be known (from previous publications or other resources) or calculated in settings without family-biased ascertainment

bias. Second, using a truncated multivariate Gibbs sampler the Posterior Mean Liabilities (PML) are estimated (see Posterior Mean Liabilities). These PMLs are conditional on the relatedness, case control status of all individuals, and prevalence of the disease. If we had measurements of a continuous liability then the mixed model analyses would have increased power and improved calibration. Since we only observe the case control phenotypes a method for estimating the unobserved liability is needed. One method is to make inference on the posterior mean liabilities. Finally, these components are used to calculate the  $\chi^2$  (1 d.o.f) association score statistic between the candidate SNP and the PML. We have released open-source software implementing the LT-Fam statistic (see Web Resources).

To better understand the need to directly account for family-biased ascertainment it is helpful to consider a toy example. The probability of being a case or control depends on the case-control status of related samples and the prevalence of the disease. Figure 2.1 depicts the conditional probabilities of being a case given that an individual's sibling is a case (A) and the probability of being a case given that an individual's sibling is a control (B). Assuming for simplicity individuals are sibling pairs with 0.50 genetic relatedness with 1 heritability the phenotypes/liabilities are generated using bivariate normal. Looking at how these two curves lie relative to the dotted line (the disease prevalence plotted against itself) shows the relative level discordant and concordant siblings at different disease prevalences.





**Figure 2.1** Based on analytic derivation of siblings (assuming correlation of 0.5 and underlying bivariate normal liability) above are the plots of the conditional probabilities of being a case given that individual’s sibling is a case (A) and the probability of being a case given the individual’s sibling is a control (B). The dotted line is the disease prevalence plotted against itself, demonstrating that siblings status will result in a sibling with the same case control status at a rate higher than the population prevalence. The dotted line is the disease prevalence plotted against itself, demonstrating that siblings’ status will result in a sibling with the same case control status at a rate higher than the population prevalence.

### *Narrow sense Heritability and Threshold GRM*

An important component of mixed model analysis is characterizing the heritability. In GWAS settings the fraction of heritability explained by genotyped SNPs,  $h_g^2$ , is different from the narrow sense heritability  $h^2$ . The narrow sense heritability,  $h^2$ , is the fraction of the phenotypic variance explained by all genetic variants under an additive model<sup>16; 32; 38</sup>. Estimates of narrow sense  $h^2$  are typically obtained by comparing the phenotypic correlation among monozygotic and dizygotic twins. Some mixed model statistics<sup>11; 34; 37</sup> use  $h_g^2$ ; however, since the current work focuses on related individuals our analysis focuses on estimating  $h^2$ . The  $h_g^2$  is usually less than the  $h^2$  and restricted to only those SNPs used to in the formulation  $h_g^2$  since it corresponds the heritability explained by genotypes SNPs which<sup>11</sup>.

Standard mixed model association methods generally estimate  $h_g^2$  from a genetic relationship matrix (GRM) and phenotypes using restricted maximum likelihood (REML)<sup>4; 11</sup>. Genotypic data is used to build an identity by state (IBS) GRM (excluding the candidate SNP<sup>11</sup>). Other methods may use IBD estimates that rely on known pedigrees<sup>36</sup>. In our setting this information is unknown. Here we build on existing methods that take a cut-off,  $c$ , IBS GRM to estimate narrow sense heritability, where all the all estimated covariance values below  $c$  are set to zero<sup>38</sup>:

$$\begin{aligned}\hat{\boldsymbol{\theta}}^* &= \frac{\mathbf{X}^T \mathbf{X}}{M} \\ \hat{\boldsymbol{\theta}}_{i,k} &= 0 \text{ if } I(\hat{\boldsymbol{\theta}}_{i,k}^* < c) \\ \hat{\boldsymbol{\theta}}_{i,k} &= \hat{\boldsymbol{\theta}}_{i,k}^* \text{ if } I(\hat{\boldsymbol{\theta}}_{i,k}^* \geq c)\end{aligned}\tag{1}$$

where  $\mathbf{X}$  is a matrix of SNPs normalized to mean 0 and variance 1 and  $M$  is the number of SNPs. In some settings Leave Once Chromosome Out (LOCO) analysis is recommended. However, here we're approximating the pedigree by using a thresholded GRM, where the issue of proximal contamination makes a bigger difference to IBS<sup>6; 11</sup>. For this reason, the candidate SNPs will be included in the calculation of the GRM for the JHS sample. Using the IBS GRM in settings of unrelated individuals would provide estimates of  $h_g^2$ . The primary expectation is to use published narrow sense estimates; the H-E regression estimates are included for comparison purposes only as it is expected to yield biased estimates of  $h^2$  under family-biased ascertainment. The H-E regression estimate is obtained by regressing the product of the case-control phenotypes on the off diagonal terms of the cut-off GRM followed by a transformation to liability scale<sup>18; 19</sup>.

### *Posterior Mean Liabilities*

Since the continuous trait is not observed in case control settings, the latent liability is approximated using the 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.<sup>20</sup>, these correspond to the expected value of the liability conditional on the case control status:

$$PML_{uni,case} = E[\varphi|\pi_i = 1] = z/K$$

$$PML_{uni,control} = E[\varphi|\pi_i = 0] = -z/(1 - K) \quad (2)$$

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 status<sup>20</sup>.

$$PML_{multi,i} = E[\varphi|\boldsymbol{\pi}, h^2, \mathbf{V}] \quad (3)$$

The  $PML_{multi}$  for each individual is conditional on that individual's case-control status, every other individual's case-control status, and on their  $\mathbf{V}$ . One way to estimate the PML is by using a Gibbs sampler. We'll sample from a truncated multivariate normal distribution in order to get estimates of the PMLs. 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. A large sample of each of these random variables will be generated then averaged across to get the posterior means. We use 100 burn-in iterations followed by 1,000 additional MCMC iterations. We estimate the  $PML_{multi}$  by averaging over MCMC iterations. Details of the algorithm Gibbs sampler algorithm are described in the LTMLM manuscript<sup>37</sup>.

### *Liability Threshold Model and LT-Fam Association Statistic*

The goal of this work is to test for association between a candidate SNP and a phenotype

while controlling for family biased ascertainment. We first consider a quantitative trait:

$$\boldsymbol{\varphi} = \boldsymbol{\beta}\mathbf{x} + \mathbf{u} + \mathbf{e} \quad . \quad (4)$$

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  $\mathbf{x}$  with mean 0 and variance 1, with effect size  $\boldsymbol{\beta}$ . The quantitative trait value depends on the fixed effect of the candidate SNP ( $\boldsymbol{\beta}\mathbf{x}$ ), the genetic random effect excluding the candidate SNP ( $\mathbf{u}$ ), and the environmental component ( $\mathbf{e}$ ). We extend to case-control traits via the liability threshold model <sup>17</sup>, 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 prevalence and a control if the individual has liability below  $t$ .

The liability is modeled as a multivariate normal distribution that uses the Lee transformation to liability scale <sup>20</sup>:

$$L(\boldsymbol{\varphi}) = (2\pi)^{\frac{-n}{2}} |\mathbf{V}|^{-1/2} \exp\left(\frac{-1}{2} (\boldsymbol{\varphi})^T \mathbf{V}^{-1} (\boldsymbol{\varphi})\right) \quad (5)$$

We note that we observe the case-control phenotypes of the individuals and not the continuous liabilities. The random component of the mixed model comes from the phenotypic covariance  $\mathbf{V}$ . The variance between the individuals is modeled as the phenotypic covariance

$$\mathbf{V} = h^2 \widehat{\boldsymbol{\Theta}} + (I - h^2) \mathbf{I} , \quad (6)$$

where  $\widehat{\boldsymbol{\Theta}}$  is the  $N$  by  $N$  cut-off GRM,  $\mathbf{V}$  is the phenotypic covariance,  $h^2$  is the heritability parameter, and  $\mathbf{I}$  is the identity matrix. In order to estimate the phenotypic covariance the heritability parameter and Genetic Relatedness Matrix are needed (described in *Narrow sense Heritability and Threshold GRM*).

The LT-Fam association statistic is a modification of the LTMLM association statistic focused on controlling family-biased ascertainment. In this case control setting, we'll assume there is some latent continuous trait, or liability. The method uses a retrospective association score statistic assuming a liability threshold model but, now directly accounting for related individuals<sup>37</sup>. 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  $(\boldsymbol{\varphi}, \boldsymbol{x})$  and derive the joint likelihood for these combined terms. The covariance of  $\boldsymbol{\varphi}$  and  $\boldsymbol{x}$  between individual  $i$  and  $k$  is:

$$\text{Cov}(\boldsymbol{\varphi}_i, \boldsymbol{x}_k) = E[\boldsymbol{\varphi}_i, \boldsymbol{x}_k] - E[\boldsymbol{\varphi}_i]E[\boldsymbol{x}_k] = E[\boldsymbol{\varphi}_i, \boldsymbol{x}_k] = E[\beta \boldsymbol{\theta}_{i,k}, \boldsymbol{x}_k] = \beta \boldsymbol{\theta}_{i,k}, \quad (7)$$

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  $(\boldsymbol{\varphi}, \boldsymbol{x})$  as a function of effect size  $\beta$  is:

$$\boldsymbol{C}(\beta) = \begin{pmatrix} \boldsymbol{V} & \beta \boldsymbol{\theta} \\ \beta \boldsymbol{\theta}^T & \boldsymbol{\theta} \end{pmatrix}, \quad (8)$$

thus

$$\boldsymbol{C}(\beta)^{-1} = \begin{pmatrix} \boldsymbol{V}^{-1} & -\beta \boldsymbol{V}^{-1} \\ -\beta (\boldsymbol{V}^{-1})^T & \boldsymbol{\theta}^{-1} \end{pmatrix} + O(\beta^2), \quad (9)$$

where both of these matrices are  $2N$  by  $2N$ . (We note that the product of the matrices in equation

10 and equation 11 is  $\begin{pmatrix} \mathbf{I} + O(\beta^2) & 0 \\ 0 & \mathbf{I} + O(\beta^2) \end{pmatrix}$ , whose difference from the identity contains

only  $O(\beta^2)$  terms.)

The joint likelihood of the liability and genotypes are distributed as a multivariate normal  $N(\mathbf{0}, \boldsymbol{C}(\beta))$ , and thus

$$L(\mathbf{x}, \boldsymbol{\varphi} | \boldsymbol{\beta}) = (2\pi)^{\frac{-n}{2}} |\mathbf{C}(\boldsymbol{\beta})|^{-1/2} \exp\left(\frac{-1}{2} (\boldsymbol{\varphi}, \mathbf{x})^T \mathbf{C}(\boldsymbol{\beta})^{-1} (\boldsymbol{\varphi}, \mathbf{x})\right). \quad (10)$$

Taking the derivative of the log likelihood results in the score equation. This leads to the score statistic (more detailed derivation in <sup>37</sup>):

$$\text{Score statistic} = \frac{(\mathbf{x}^T \mathbf{V}^{-1} \boldsymbol{\varphi})^2}{(\mathbf{V}^{-1} \boldsymbol{\varphi})^T \boldsymbol{\Theta} (\mathbf{V}^{-1} \boldsymbol{\varphi})}, \quad (11)$$

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

In equations 9-15 the liability was assumed to be known, for simplicity. The liability is approximated (more detailed derivation in <sup>37</sup>).

$$LT - \text{Fam score statistic} = \frac{(\mathbf{x}^T \mathbf{V}^{-1} \mathbf{PML}_{multi})^2}{(\mathbf{V}^{-1} \mathbf{PML}_{multi})^T \boldsymbol{\Theta} (\mathbf{V}^{-1} \mathbf{PML}_{multi})} \quad (12)$$

The key distinctions between the LT-Fam and LTMLM statistics come from the derivation of the GRM, where LTMLM does not use a cut-off, and the heritability parameter where LTMLM  $h_g^2$  as opposed to the narrow sense estimate  $h^2$ . For consistency, the LT-Fam statistic is constructed from  $\mathbf{PML}_{multi}$  and  $\mathbf{V}$  using the cut-off GRM as well. Additionally, the LT-Fam statistic assumes related samples whereas LTMLM does not so, the denominator for the LT-Fam includes the GRM, as opposed to LTMLM which contains the identity matrix. The overall computational cost of computing the LTMLM statistic is  $O(MN^2)$  when  $M > N > \#iterations$ . We have fixed the number of iterations at 100 burn-in iterations followed by 1,000 additional iterations.

#### *Simulated Genotypes and Simulated Phenotypes*

We performed simulations using simulated genotypes and simulated phenotypes, all with  $N/2$  sibling pairs. Under each simulation scenario approximately 50% cases and 50% controls were ascertained and 100 separate simulations were run. For settings where  $N = 5,000$  a random set of 100 SNPs were causal and for  $N = 1,000$  a random set of 20 SNPs were set to be causal. All simulations included  $M$  candidate SNPs (either 50,000 or 10,000) and an independent set of  $M$

GRM SNPs, which were used for estimating the random component of the mixed model statistics. Half of the causal SNPs were candidate and the other half were GRM SNPs.

Siblings were simulated by generating genotypes of parents for each sib pair, 25 blocks of SNPs from each parent haplotype were randomly passed along to the children to simulate mating. Three different family-biased ascertainment schemes were considered: unbiased sampling, all concordant siblings, and all discordant siblings.

At a disease prevalence of  $f$ , sib pairs are generated then only retaining a subset based on the case control status of each pair. Under the unbiased scheme all case-case siblings are retained, case-control siblings are retained with probability  $f*(1-f)$ , and control-control siblings are retained with probability  $[f*(1-f)]^2$ . The expected proportion of cases is then probability of getting a case-case times the retention rate plus the probability of getting a case control siblings times the retention,  $2*f*f*1 + 2*f*(1-f)*f*(1-f)$ . an approximately 50/50 ascertainment scheme with varying levels of discordant and concordant sibling pairs based on the prevalence. For the concordant scheme,  $N/4$  sibling pairs are case-case and  $N/4$  sibling pairs are control-control. For the discordant scheme, all  $N/2$  sibling pairs are discordant (Table 2.1). The corresponding  $h^2$  and  $h_g^2$ , where  $h_g^2$  estimates are used for the standard MLM and LTMLM versus  $h^2$  which used for LT-Fam in settings of unknown narrow sense heritability, estimates were calculated knowing that (in the concordant and discordant schemes) they would be severely biased, for both the cut-off GRM, assuming  $c=0.05$ , and unrestricted GRM (Table S2.1).

In the discordant and concordant simulations the  $h^2$  is set to 0.50 and the LT-Fam statistic assumes this parameter to be known; for the unbiased setting the HE-regression estimate is used. In some scenarios the narrow sense heritability might be expected to be mis-specified, so to test

the sensitivity simulations where the narrow sense heritability is incorrectly specified were performed as well.

We compared Armitage Trend Test (ATT), MLM, LTMMLM, and LT-Fam statistics. We evaluated performance using average  $\chi^2$  statistics at causal, null, and all markers,  $\lambda_{GC}$  at all markers (median  $\chi^2$  divided by 0.455)<sup>25</sup>, and proportion of causal and null markers that were significant at various P-value thresholds.

### *JHS Genotypes and T2D Phenotypes*

We analyzed JHS individuals with case-control phenotypes for type 2 diabetes (T2D), a disease with a prevalence around 8%. The data set contained 339 cases and 1778 controls genotyped at 736,614 SNPs after QC<sup>40</sup>. We compared ATT, MLM, and LT-Fam statistics. LTMMLM is expected to be improperly calibrated, so it is not included. In order to test the impact of family-biased ascertainment three down sampling schemes were used: controls that were related to cases were removed (at a level above the cut-off), cases that did not have a case relative were removed (at a level below the cut-off), or both. A total of 94 cases and 1318 controls remained after down sampling both cases and controls.

The heritability estimates were calculated and then converted to the liability scale for: HE-regression and REML<sup>16; 18; 19</sup>, using either the full or cut-off GRM. The value of narrow-sense heritability used for the down sampled data was set to 0.257, the HE-regression cut-off estimate from the full sample. We also ran LT-Fam with mis-specified narrow-sense heritability values ranging from 0.25 and 0.75.

## **Results**

### *Simulations: Simulated Genotypes and Simulated Phenotypes*



We compared the performance of ATT, MLM, LTMLM and LT-Fam in simulated studies of siblings. These statistics were considered because Eu-ahsunthornwattana finds that MLM performs at least as well as the other methods in family-based association studies<sup>36</sup>. We considered unbiased (without family-biased ascertainment), concordant sibling and discordant sibling studies (see Materials and Methods). Results are displayed in Table 2.1 where all of the statistics appear to be close to properly calibrated in the unbiased studies, having mean near 1.000 for null SNP sets. In settings of both concordant siblings for lower prevalence diseases LT-Fam again is properly calibrated (average  $\chi^2$  of 0.999) and MLM appears deflated (0.674) whereas both LTMLM and ATT appear to be inflated (1.499). As the disease prevalence decrease to 1% and lower, LT-Fam attains 3% higher power than MLM and 8% higher power than ATT after properly calibrating using the respective  $\lambda_{GC}$ . MLM and ATT are deflated in settings of discordant sibling sampling, LTMLM gets unstable estimates, and LT-Fam is properly calibrated.

Table 2.1

M	N	Sibling Set	Prevalence	Set	ATT	MLM	LTMLM	LT-Fam
50000	1000	Unbiased	50%	Causal	16.091(0.668)	15.423(0.641)	14.26(0.591)	14.502(0.601)
				Null	1.105(0.001)	1.044(0.001)	0.965(0.001)	1.004(0.001)
				All	1.108(0.001)	1.047(0.001)	0.968(0.001)	1.007(0.001)
				All $\lambda_{GC}$	1.108(0.002)	1.047(0.002)	0.967(0.005)	1.007(0.002)
			10%	Causal	21.763(0.892)	20.484(0.839)	21.67(0.897)	20.889(0.865)
				Null	1.062(0.001)	0.986(0.001)	1.026(0.001)	1.004(0.001)
				All	1.066(0.001)	0.99(0.001)	1.03(0.001)	1.008(0.001)
				All $\lambda_{GC}$	1.062(0.004)	0.987(0.004)	1.026(0.002)	1.003(0.002)
			1%	Causal	36.139(1.419)	31.663(1.237)	39.148(1.576)	35.62(1.419)
				Null	1.054(0.001)	0.921(0.001)	1.085(0.001)	1.004(0.001)
				All	1.061(0.001)	0.927(0.001)	1.093(0.001)	1.011(0.001)
				All $\lambda_{GC}$	1.055(0.01)	0.921(0.007)	1.086(0.002)	1.005(0.001)
		Concord	50%	Causal	26.58(1.098)	12.13(0.512)	27.751(1.246)	17.785(0.746)
				Null	1.502(0.001)	0.68(0)	1.489(0.001)	1.001(0.001)
				All	1.507(0.001)	0.683(0)	1.494(0.001)	1.004(0.001)
				All $\lambda_{GC}$	1.501(0.002)	0.68(0.001)	1.489(0.001)	0.987(0.005)
			10%	Causal	36.048(1.49)	16.675(0.7)	27.04(1.193)	24.767(1.057)
				Null	1.499(0.001)	0.676(0)	1.446(0.001)	0.999(0.001)
				All	1.506(0.001)	0.679(0)	1.451(0.001)	1.003(0.001)
				All $\lambda_{GC}$	1.499(0.002)	0.676(0.002)	1.446(0.004)	0.986(0.005)
			1%	Causal	67.318(2.393)	31.717(1.155)	63.901(2.418)	47.75(1.773)
				Null	1.499(0.001)	0.674(0)	1.475(0.001)	0.999(0.001)
				All	1.513(0.001)	0.68(0.001)	1.487(0.001)	1.008(0.001)
				All $\lambda_{GC}$	1.502(0.002)	0.674(0.001)	1.484(0.003)	0.985(0.005)
Discord	50%	Causal	6.354(0.27)	6.354(0.27)	N/A	12.272(0.514)		
		Null	0.504(0)	0.504(0)	N/A	1.003(0.001)		
		All	0.505(0)	0.505(0)	N/A	1.005(0.001)		
		All $\lambda_{GC}$	0.516(0.001)	0.516(0.001)	N/A	1.059(0.001)		
	10%	Causal	9.096(0.376)	9.096(0.376)	N/A	17.871(0.736)		
		Null	0.504(0)	0.504(0)	N/A	1.005(0.001)		
		All	0.506(0)	0.506(0)	N/A	1.009(0.001)		
		All $\lambda_{GC}$	0.516(0.001)	0.516(0.001)	N/A	1.061(0.002)		
	1%	Causal	12.594(0.543)	12.594(0.543)	7.057(0.315)	25.24(1.101)		
		Null	0.503(0)	0.503(0)	0.893(0.001)	1.004(0.001)		
		All	0.506(0)	0.506(0)	0.895(0.001)	1.009(0.001)		
		All $\lambda_{GC}$	0.515(0.001)	0.515(0.001)	0.893(0.008)	1.057(0.003)		

Table 2.2 (Continued)

M	N	Sibling Set	Prevalence	Set	ATT	MLM	LTMLM	LT-Fam
50000	5000	Unbiased	50%	Causal	17.147(0.342)	16.517(0.329)	15.36(0.306)	15.579(0.31)
				Null	1.101(0.001)	1.043(0.001)	0.97(0.001)	1.002(0.001)
				All	1.118(0.001)	1.059(0.001)	0.985(0.001)	1.017(0.001)
				All $\lambda_{GC}$	1.102(0.001)	1.045(0.001)	0.972(0.002)	1.002(0.001)
			10%	Causal	22.422(0.447)	21.154(0.423)	22.003(0.441)	21.063(0.421)
				Null	1.069(0.001)	0.992(0.001)	1.03(0.001)	1.002(0.001)
				All	1.091(0.001)	1.012(0.001)	1.051(0.001)	1.022(0.001)
				All $\lambda_{GC}$	1.07(0.002)	0.992(0.002)	1.03(0.001)	1.002(0.001)
			1%	Causal	32.44(0.672)	28.431(0.585)	34.503(0.712)	31.108(0.64)
				Null	1.041(0.001)	0.912(0.001)	1.092(0.001)	1.003(0.001)
				All	1.073(0.001)	0.94(0.001)	1.125(0.001)	1.033(0.001)
				All $\lambda_{GC}$	1.042(0.006)	0.912(0.005)	1.094(0.002)	1.004(0.001)
		Concord	50%	Causal	27.243(0.537)	17.567(0.347)	24.986(0.515)	18.19(0.359)
				Null	1.504(0.001)	0.983(0.001)	1.421(0.001)	1.002(0.001)
				All	1.53(0.001)	1(0.001)	1.445(0.001)	1.019(0.001)
				All $\lambda_{GC}$	1.506(0.002)	0.985(0.002)	1.427(0.008)	1(0.002)
			10%	Causal	39.44(0.777)	25.372(0.502)	N/A	26.419(0.524)
				Null	1.507(0.001)	0.976(0.001)	N/A	1.004(0.001)
				All	1.545(0.001)	1(0.001)	N/A	1.029(0.001)
				All $\lambda_{GC}$	1.509(0.002)	0.979(0.001)	N/A	1.005(0.002)
			1%	Causal	73.249(1.418)	46.183(0.898)	57.341(1.149)	49.503(0.973)
				Null	1.511(0.001)	0.949(0.001)	1.317(0.001)	1.006(0.001)
				All	1.583(0.002)	0.995(0.001)	1.373(0.002)	1.054(0.001)
				All $\lambda_{GC}$	1.513(0.002)	0.951(0.002)	1.316(0.011)	1.004(0.002)
Discord	50%	Causal	6.215(0.126)	6.215(0.126)	N/A	12.352(0.25)		
		Null	0.503(0)	0.503(0)	N/A	1.004(0.001)		
		All	0.509(0)	0.509(0)	N/A	1.016(0.001)		
		All $\lambda_{GC}$	0.505(0)	0.505(0)	N/A	1.011(0)		
	10%	Causal	8.061(0.165)	8.062(0.165)	N/A	16.063(0.329)		
		Null	0.502(0)	0.502(0)	N/A	1.003(0.001)		
		All	0.51(0)	0.51(0)	N/A	1.018(0.001)		
		All $\lambda_{GC}$	0.505(0)	0.505(0)	N/A	1.011(0)		
	1%	Causal	12.841(0.262)	12.841(0.262)	N/A	25.621(0.522)		
		Null	0.501(0)	0.501(0)	N/A	1.001(0.001)		
		All	0.514(0)	0.514(0)	N/A	1.025(0.001)		
		All $\lambda_{GC}$	0.504(0)	0.504(0)	N/A	1.011(0)		

**Table 2.1:** Sibling study of simulated genotypes and phenotypes, where the unbiased analysis has individuals sampled such that acceptance probabilities for case-case siblings is 100%, case-control siblings is  $f^*(1-f)$ , and control-control siblings is  $[f^*(1-f)]^2$ . This will produce an approximately 50/50 ascertainment scheme with varying levels of discordant and concordant sibling pairs based on the prevalence. The concordant sibling setting has 50/50 case-case and control-control sibling pairs, whereas the discordant are all discordant sibling pairs. There are M

Table 2.2 (Continued)

SNPs, N Individuals, and 100 simulations for each scenario of disease prevalence, M, N, and sibling ascertainment scheme. Settings where results for LTMLM read N/A come from convergence issues do to largely mis-calculated heritability estimates which, is expected when looking at highly related individuals.

LT-Fam results are based on knowledge of the correct  $h^2$  (and did not use REML or H-E regression estimates), whereas other methods are not designed to use this knowledge. We determined that  $h^2$  estimates from both REML and H-E regression were in fact biased (Table S2.1), which explains the mis-calibration of MLM and LTMLM statistics in Table 2.1. Impact of mis-specification in the LT-Fam statistic

The impact of mis-specification of the narrow sense heritability  $h^2$  was tested, where the true value was 0.50 and the specified parameter was set to 0.25, 0.40, 0.60, and 0.75 at a disease prevalence of 1% where improper calibration was seen for other statistics. LT-Fam demonstrated a slight loss in power relative to LT-Fam with correct  $h^2$ , but was still properly calibrated (Table S2.2).

### *JHS Genotypes and T2D Phenotypes*

We analyzed 2,117 JHS individuals (339 cases and 1,778 controls) with T2D phenotypes typed on genome-wide arrays. We analyzed both the full data set, and down-sampled data sets with family-biased ascertainment in which we removed controls that were related to cases and/or removed cases that did not have a case relative, based on a down-sampling relatedness cut-off of 0.05 (see Materials and Methods). The first row of Table 2.2 depicts the setting when all individuals are used, then the subsequent rows down sample to artificially induce highly levels of family-biased ascertainment. In the first row we observe properly formed for ATT, MLM and LT-Fam, all relatively close to an average  $\chi^2$  of . However, after down sampling the statistics become increasingly biased as family biased ascertainment also increases, except for the LT-Fam statistic. In this setting of 1440 samples, where only controls that aren't related to cases and

cases related to another case remain, the average  $\chi^2$  for LT-Fam was 0.980, whereas ATT was inflated 1.337, and MLM was deflated 0.820 (Table 2.2). Down-sampling runs with a down-sampling relatedness cutoff of 0.025 produced qualitatively similar results (Table S2.3).

Controls	Controls	Cases	Cases	Total	ATT	MLM	LT-Fam
1778	All	339	All	2145	1.044(0.002)	1.000(0.002)	1.019(0.002)
1318	Unrelated	339	All	1684	1.135(0.002)	0.950(0.002)	0.995(0.002)
1778	All	94	Related	1900	1.236(0.002)	0.973(0.002)	0.994(0.002)
1318	Unrelated	94	Related	1440	1.337(0.002)	0.820(0.001)	0.980(0.002)

**Table 2. 2** The average  $\chi^2$  statistics genome wide are reported at a GRM cut-off of 0.05 for LT-Fam run on real genotypes and phenotypes with different family ascertainment bias. To mimic family-biased ascertainment: controls that were related to cases were removed, cases that did not have a case relative were removed, or both. Unrelated means the controls that were related to any case were removed leaving only the controls that were not related to any case (based on the cut-off level). The Related term refers to cases that did not have any case relatives being removed, leaving only the cases related to another case (based on the cut-off level).

We determined that  $h^2$  estimates from both REML and H-E regression were biased in the down-sampling runs (Table S2.4), which explains the mis-calibration of MLM and LTMLM statistics in Table 2.2. In the most extreme down-sampling of both cases and controls, the REML estimate of narrow-sense heritability was 4.017 (or 1.726 at a GRM cut-off of 0.025), which is outside the plausible 0-1 range. This is consistent with our simulations, in which biased heritability estimates also produced mis-calibrated MLM and LTMLM statistics in settings with family-biased ascertainment.

## Discussion

Through both simulated sibling studies and using real JHS T2D samples we have demonstrated mis-calibration of existing  $\chi^2$  statistics and introduced our properly calibrated LT-

Fam statistic. The mis-calibration of standard  $\chi^2$  statistics comes from the family-biased ascertainment of extreme phenotypes. For this reason, narrow sense estimation of  $h^2$  from a reference population in conjunction with the LT-Fam statistic results in a properly calibrated test.

The primary methods developed for similar settings, family data in retrospective case control settings, were developed by Thorton and McPeck. Thorton and McPeck developed several statistics and corresponding pieces of software to analyze case control diseases in family settings, specifically ROADTRIPS and MQLS (whereas MASTOR is for quantitative traits)<sup>21; 36; 39</sup>. Both statistics also adjust for known relatedness while using a retrospective model but, have been found to demonstrate lower power than MLM in certain settings<sup>36</sup>. One clear advantage of ROADTRIPS and MQLS is they take advantage of all phenotype information, even for individuals that have not been genotyped. The improvement from our LT-Fam statistics is it explicitly models the latent liability while control for family-biased ascertainment while, ROADTRIPS and MQLS does not.

Several limitations of our LT-Fam method restrict it's use but, also allow for future directions of advancement. The LT-Fam method does require the  $h^2$  the parameter from the literature; however, we did demonstrate that the LT-Fam statistic is robust to misspecification (Table S2.2). Futur directions include improvements in speed of the algorithm. The current algorithm runs in  $O(MN^2)$  whereas existing methods run much faster (<sup>34</sup> which is much faster for unrelated samples). Incorporation of fixed effect covariates and large effect SNPs have been shown to increase power in mixed model settings, we have not been considered here but, this is a future avenue to improve the statistic. As with the LTMLM statistic, the method relies on the assumption of underlying normally distributed liability but, has been shown to be a reasonable

assumption<sup>32; 33</sup>. Much work is being done to increase power in settings of multiple phenotypes, the LT-Fam statistic could be extended for such settings of multiple case control phenotypes.

# 3. Two Step Bayesian Model Averaging using Propensity Score Adjustment for Treatment Effect Estimation

## 3.1). *Introduction*

In clinical settings often the scientific question of interest is isolating the effect of a primary exposure, say, elevated expression of a specific metabolite, on Body Mass Index (BMI) or pancreatic ductal adenocarcinoma (PDAC). Other factors such as subject demographics and other epidemiological measures that are associated with the primary exposure and the outcome may act as confounders, clouding or completely inverting the true effect of the primary exposure.

The goal here is to understand the causal effect of elevated levels of BCAAs on either PDAC or BMI. By characterizing the effect of the primary exposure on the outcome, we understand how limiting or targeting the exposure may impact outcomes. This can lead to efforts to directly modify the primary exposure, policy or preventative medicine schemes, or indirectly modifying the primary exposure through developing drugs or other targeted interventions. Note the distinction between this goal and that of predicting future cases PDAC or obesity.

Absent the benefits of randomized exposure, statistical methods for confounding adjustment are required to estimate quantities that can be interpreted, at least approximately, as causal effects of a particular exposure. Many techniques will leverage Propensity Score (PS) methods, where the probability of assignment to treatment is estimated and then the outcome are compared between treated versus untreated individuals with similar estimates of the PS. PS characterize discrepancies between individuals probabilities of being treated caused by confounders.



Despite the proven value of propensity scores for confounding adjustment in observational studies, implementation can be difficult in high dimensional settings. It could even be argued that this difficulty is limiting the use of PS methods in lots of scientific contexts (e.g., genetic or metabolomic applications). Often PS models will use the “kitchen sink” approach, where all of the covariates are used to calculate the PS estimates. In high dimensional settings the “kitchen sink” approach becomes untenable, using all covariates in the model or to inform the propensity score results in unstable estimates. Thus, researchers are increasingly confronted with the need to decide which of a high-dimensional set of covariates that are genuine confounders to include in the PS.

Note that the wide array of methods for penalized regression, such as Lasso or more involved penalties like adaptive elastic net, penalize the inclusion of covariates in a regression model. These methods do not control for confounding, but instead focus on association with covariates and the outcome, as opposed to targeting the effect of primary exposure of interest. Such methods are tuned for prediction purposes, but are limited for effect estimation because they do not target the inclusion of genuine confounders. Model selection techniques, such as stepwise selection, have been used to select variables to include in the PS, but, as with all model selection techniques, limit the inference to specific set of confounders selected and may arbitrarily miss the impact of some true confounders. Schneeweiss developed a reasoned algorithm specific to PS estimation that first dichotomizes covariates to estimate their relative prevalence in treated and untreated subjects and relative risk with exposure. Then based on the prevalences and RR, each covariate is ranked and a pre specified  $k$  number of them are used in the outcome model making the method sensitive to the selection of  $k$ <sup>41</sup>. In addition to being sensitive to the choice of how many covariates to include and it is not designed to detect complex relationships among

confounders. Techniques that select a single subset of potential confounders then condition all inference on the selected set and do not account for the uncertainty inherent to the choice of confounders to include for adjustment. This level of model uncertainty is increasingly important in high-dimensional settings.

The need to prioritize genuine confounders while properly accounting for confounding uncertainty has motivated a recent vein of new methods<sup>42; 43</sup>. These methods are rooted in the core ideas of Bayesian Model Averaging<sup>44</sup> in that, rather than choose the confounders to include in the propensity score a priori, they average over a large space of PS specifications according to posterior belief that each measured covariate is or is not important for adjustment. In metabolomics a high dimensional set of unknown confounders make for a novel setting for such propensity score methods that may improve effect estimation, where new methods are needed that leverage PS in high dimensional scientific contexts.

Continuing the line of research in Zigler and Dominici<sup>45</sup>, the purpose of this paper is to provide a flexible 2-Step Bayesian Model Averaging method and demonstrate its effectiveness for treatment effect estimation in settings of high dimensional confounding<sup>43; 45</sup>. This method improves on existing methods by averaging over the model space controlling for confounding while targeting treatment effect. It does so without need of an arbitrary number of confounders to include a priori and making assumptions about the underlying model. BMA weights each model according to the posterior model weight, which is calculated in a manner to prioritize the variables most important for causal effect estimation<sup>44-46</sup>. We'll compare our method to several of these methods that use model selection, adjustment for propensity scores, and existing methods we're expanding upon to help highlight when the 2-Step BMA is most appropriately applied. The proposed method will be compared under different simulation scenarios and in a

metabolomics example using data from the Nurses Health Study (NHS) and Health Professionals Follow Up-Study (HPFS). Comparison methods will include a gold standard method assuming knowledge of the true data-generating mechanism (Gold), the kitchen sink approach (Kitch), step-wise model selection of the PS model (SW), High Dimensional Propensity Score Adjustment (HDPS), Generalized Boosted Regression for the PS model (GBR), and the Approximately Bayesian Model Averaging MC<sup>3</sup> proposed in Zigler and Dominici (AB-MC3)<sup>41</sup>;<sup>45; 47</sup>. These methods are described and compared in tables 3.1 and 3.2.

Method	Abbreviation	Brief Description
<b>Gold Standard</b>	Gold	A PS model is fit using only the terms known to be associated with the outcome. This can only be used in settings where the true covariates associated with the outcome are known, such as the simulation scenarios.
<b>2-Step Bayesian Model Averaging</b>	2-Step BMA	Performs BMA averaging on both the exposure model and outcome model separately to cut the feedback. The PS are calculated by first performing parameter selection just based on the exposure and then a function of the PS is use to perform BMA on the outcome model.
<b>Approximately Bayesian MC<sup>3</sup> Approach</b>	AB-MC3	Approximately Bayesian approach that separates the exposure and outcome model set while performing MCMC integration across the model space.
<b>High Dimensional Propensity Score Adjustment</b>	HDPS	First the covariates are dichotomized then the criteria for assignment is the Apparent Relative Risk (ARR). The ARR is a function of the imbalance in prevalence of those exposed versus unexposed in conjunction with the independent association between potential confounder and outcome. After ranking based on the ARR then a pre-specified number of covariates are included as confounders in the PS.
<b>Toolkit for Weighting and Analysis of Nonequivalent Groups Kitchen Sink</b>	T-KS T-ES Or generally GBR	Generalized boosted regression modeling that uses a multivariate nonparametric approach using a regression tree following a recursive algorithm to estimate a function characterizing treatment assignment. Two different stopping criteria, KS and ES.
<b>Step-Wise</b>	SW	Performs stepwise selection using BIC criteria on the propensity score model

Table 3.1: Brief description comparing the different methods being evaluated. For consistency of comparison, each method was implemented to use PS broken up by quintiles.

Method	Kitch	Step-Wise	2-Step BMA	AB-MC3	HDPS	GBR
<b>Accounts for terms associated with X &amp; Y</b>	PS	PS	PS and outcome	PS	PS	PS
<b>Accounts for terms associated with Y</b>	PS	No	outcome	No	PS	PS
<b>Accounts for terms associated with X</b>	PS	PS	PS	PS	PS	PS
<b>Augmented PS with covariates</b>	Inefficient	Yes	Yes	Yes	Yes	No
<b>Treats Covariates as Binary</b>	No	No	No	No	Yes	No
<b>Pre-specified number <math>\alpha</math> terms</b>	Yes	No	No	No	Yes	No
<b>Model Uncertainty</b>	No	No	Yes	Yes	No	No

Table 3.2: Comparison of the different features of the methods being evaluated.

The scientific context we explore here involves metabolic diseases. Metabolites are known to impact diseases such as pancreatic cancer; however, the 2-Step BMA method is particularly beneficial in settings such as metabolomics because it is hard to reason scientifically about individual metabolites. Here we wish to estimate the treatment effect of elevated levels of branched-chain amino acids (BCAAs) on pancreatic cancer outcomes while properly controlling for this large set of potential confounders. As comparison and validation of our method, we will conduct analyses analogous to those in Mayers et al<sup>48</sup> examining the association between BCAA and PDAC<sup>48</sup> and those in Liu et al<sup>49</sup> examining the marginal association between BCAAs and BMI (Liu pending).

## **2) Estimating causal effects with PS**

Here we develop a method for estimating the effect of a binary treatment,  $X$ , binary outcome of interest,  $Y$ , while adjusting for a set of potential confounders  $C$ , where  $C$  is possibly high dimensional.

### **2.1) Notation, estimand**

We'll denote our binary treatment with  $X=[0,1]$  and the outcome of interest with  $Y=[0,1]$  with a vector of  $p$  measured covariates as  $C$ . The propensity score ( $PS$ ) will be defined as the conditional probability of being treated,  $X=1$ , given the confounders  $C$  where throughout the manuscript we'll assume strong ignorability ("no unmeasured confounding")<sup>50</sup>. A confounder is a covariate that is associated with both the exposure and outcome. Failure to adjust for confounding will distort estimates of treatment effects. The main goal of our analysis is for improved estimation of the average causal effect of the treatment on the outcome while accounting for confounders with  $E[Y|C, X=1] - E[Y|C, X=0]$ .

<b>Term</b>	<b>Description</b>
$X$	Primary exposure of interest
$Y$	Outcome
$C$	Set of confounders
$\alpha_x$	Vector of indicators of terms to include the PS model
$\gamma$	Coefficients in the PS model
$g_x (*)$	Link function for PS model
$\alpha_y$	Vector of indicators of terms to include the outcome model
$\beta$	Coefficients in the outcome model
$g_x (*)$	Link function for PS model
$PS$	The propensity score
$h(*)$	Generalized function of the PS

Table 3.3: Description of terms used in formulation.

## 2.2) Propensity scores and Confounding Uncertainty

The PS is probability of treatment given the set of true confounders,  $C$  (Table 3.3).

$$PS = E[X|C] \quad (3.1)$$

If the true confounders in our model were known, the generalized linear model for exposure  $X$  is  $g_x(E[X_i|C_i]) = \sum_{k=0}^p \gamma_k C_{i,k}$  where the  $\gamma$  are our exposure coefficients. GLMs of this form will be used estimate (1). We extend the GLM formulation for the PS model to accommodate settings where the true set of confounders to be included in the model is unknown.

Towards this end, we introduce the vector of indicators for terms in our exposure model,  $\alpha_x$ .

$$g_x(E[X_i|C_i]) = \sum_{k=0}^p \alpha_x \gamma_k C_{i,k} \quad (3.2)$$

The  $\alpha_x$  parameter is a vector of indicators for which confounders are in the model. From the GLM the PS scores can be estimated, plugging back in, probability of treatment given confounders is:

$$PS(\gamma, \alpha, C_i) = g_x^{-1}(\sum_{k=0}^p \alpha_x \gamma_k C_{i,k}) \quad (3.3)$$

The confounders and PS are included to potentially gain efficiency and model residual imbalances, similar to existing matching techniques<sup>51</sup>. And a generalized linear model for our outcome Y:

$$g_y(E[Y_i|X_i, C_i]) = \beta_0 + \beta_X X_i + h(PS(\gamma, \alpha_x, C_i); \beta) + \sum_{k=0}^p \alpha_y \beta_k C_{i,k} \quad (3.4)$$

So, the outcome model will be conditional on the treatment, a function of the PS, and the covariates associated with Y. Similar to the  $\alpha_x$  term, the  $\alpha_y$  is key the second step of the 2-Step BMA algorithm, allowing for flexibility to search of the outcome model space conditional on the PS (see supplement for details).

It is common practice to discretize the PS into quintiles or deciles, we'll assume an  $h(.)$  of quintiles (equation 4) throughout all methods used in the analysis for consistent comparison. This will allow for a dimension reduction of confounders while be generalized enough to allow for non-linear effects of the matched sets. Similarly, we have a set of indicators for terms in the outcome model  $\alpha_y$  and a new set of covariates related to the outcome,  $\beta$ .

### 3) BMA for PS

The  $\alpha_x$  and  $\alpha_y$  parameters should be regarded as unknown parameters, since the true model is unknown there is some large set of possible models,  $M$ , being considered. They are treated separately for the purposes of prioritizing variables based on being associated with either treatment or outcome. The treatment effect will be estimated by averaging over the different set of models weighted by the probability of each model.

$$P(ACE | X, C) \approx \sum_{\alpha_x, \alpha_y \in M} P(ACE | \alpha_x, \alpha_y, X, C) P(\alpha_x, \alpha_y | X, C) \quad (3.5)$$

The goal is to find the proper set of  $\alpha_x$  and  $\alpha_y$  that contain the true set of confounders and terms truly associated with Y. A Markov Chain Monte Carlo (MCMC) algorithm will be

detailed below which samples from the joint posterior space to get estimates of the posterior ACE. By building on existing BMA this technique specifically targets the treatment effect, without restricting to a single model. In doing so BMA approaches account for model uncertainty while controlling for confounding, which is robust to spurious correlation between true confounders and other related covariates <sup>43; 44; 46</sup>.

### 3.2) Two-Step BMA

The method we propose searches the possible model space by first fitting the PS based on the exposure model. Then iteratively using a function of the PS estimates to fit the outcome model conditional on the PS. The model space of the exposure model and outcome models are averaged over to get the treatment effect estimates. This model assumes:

- Strong ignorability, where PS require specification of all necessary confounders.
- $\alpha_x$  and  $\alpha_y$  independent and flat priors
- The Bayes factor approximated by the BIC assumption (details in the supplement and see <sup>43; 44; 52</sup>)
- Generalized  $g_x(\cdot)$  and  $g_y(\cdot)$
- Equal prior probability of each model,  $M^t$



Assuming a retrospective likelihood, where both the exposure and outcome are modeled jointly.

$$L(Y, X|C, \gamma, \alpha, \beta, \delta)$$

$$= \prod_{i=1}^n \left[ g_x^{-1} \left( \sum_{k=0}^p \alpha_x \gamma_k C_{i,k} \right) \right]^{X_i} \left[ 1 - g_x^{-1} \left( \sum_{k=0}^p \alpha_x \gamma_k C_{i,k} \right) \right]^{1-X_i} * \\ \left[ g_y^{-1} \left( \beta_0 + \beta_X X_i + h(PS(\gamma, \alpha_x, C_i); \beta) + \sum_{k=0}^p \alpha_y \beta_k C_{i,k} \right) \right]^{Y_i} * \\ \left[ 1 - g_y^{-1} \left( \beta_0 + \beta_X X_i + h(PS(\gamma, \alpha_x, C_i); \beta) + \sum_{k=0}^p \alpha_y \beta_k C_{i,k} \right) \right]^{1-Y_i} \quad (3.6)$$

Or more simply without the model parameters, the retrospective model maybe thought of as a two parameter binomial:

$$L(Y, X|C, \gamma, \alpha, \beta, \delta) = \prod_{i=1}^n [g_x^{-1}(*)]^{X_i} [1 - g_x^{-1}(*)]^{1-X_i} * [g_y^{-1}(*)]^{Y_i} [1 - g_y^{-1}(*)]^{1-Y_i}$$

The posterior samples of  $(\alpha_x, \gamma, \alpha_y, \beta)$  are obtained by iteratively sampling from  $P(\alpha_x | X, C)$ ,  $P(\gamma | \alpha_x, X, C)$ ,  $P(\alpha_y | D, f(\alpha_x, \gamma))$ , and  $P(\beta | \alpha_y, D, f(\alpha_x, \gamma))$  (Details of this derivation can be found in the supplement). Where our observed data in a retrospective model consists of  $D=(X, C, Y)$ . The first stage of the MCMC algorithm is to get  $\alpha_x$  and  $\gamma$  as a way to come up with a posterior distribution for the PS that is then used as a fixed prior distribution for  $p(\beta, \alpha_y, PS|Data)$ , which is used to estimate the ACE (See Appendix 3 for more information).

### 3.3) Rationale/justification for 2-stage approach

2-Step BMA allows both the C-X and the C-Y associations to be considered when prioritizing confounders, but that this is complicated by issues of "feedback". This also provides a more straightforward integration of the PS model, so this method could use any generalized linear model. Breaking up the steps, leads to the initial drawback a PS only informed by the

association with  $X$ . This method, in part, builds on existing methods by then averaging over the set of outcome models ( $\alpha_y$ ), conditional on the posterior predictive distribution of the PS, to then account for those terms not picked up in the first step that are directly associated with  $Y$ <sup>45</sup>. We will investigate the causal effect of the dietary intervention on onset of diabetes, adjusting for all measured metabolites with the NHS and HPFS data set.

#### 4) Simulation Study

##### ***Simulation Study Comparing Different Methods***

To compare the different methods (Table 3.1 and 3.2) a series of different simulations studies were developed with varying effects, number of confounders, and correlation structure among the confounders. For all settings data are generated as follows, for  $1, 2, \dots, n$  individuals, at  $n=500$  we simulate  $p=200$ . The covariates  $C_i = (C_{i1}, C_{i2}, \dots, C_{ip})$  are drawn from a  $MVN(0, \Sigma)$  where  $\Sigma$  is either the identity or a series of bivariate normals with correlation of 0.5. The treatment  $X_i$ , is simulated from a Bernoulli distribution with probability  $P(X_i=1) = \Phi^{-1}(\cdot)$  (Equation 2). Similarly, the exposure  $Y_i$  is simulated Bernoulli distribution with probability from  $P(Y_i=1) = \text{logit}(\cdot)$  (Equation 4). Where the treatment effect is simulated with  $\beta_x = [0, 0.2, 0.4]$ .

Several scenarios were compared, there are 2 covariates are associated with  $X$  only, 2 covariates associated with  $Y$  only, and 2 covariates associated with both  $X$  and  $Y$  at  $\pm 0.3$ . Another setting will include that same set of covariates then additionally a set of low coefficients, where there are 28 covariates are associated with  $X$  only, 28 covariates associated with  $Y$  only, and 28 covariates associated with both  $X$  and  $Y$  at coefficients  $\pm 0.05$ . This results in a total of 30 covariates associated with  $X$  only, 30 terms associated with  $Y$  only, and 30 covariates associated with both  $X$  and  $Y$ .

Finally using 10 terms associated with X only, 10 terms associated with Y, and 10 confounders each with coefficients of +/-0.5. Each confounder has alternating directionality of coefficient and correlated using an autocorrelation covariance matrix with parameter of 0.4. So, considering the first two confounders, they'll have  $Cov(C_1, C_2) = 0.4$  with  $\beta_{C_1}=0.5$  and  $\beta_{C_2}=-0.5$ .

We simulated 1000 replicated data sets under each of the different simulation scenarios. The Starting values for the MCMC searches are initially at the MLE (in settings of small n and large p there maybe issues and it is recommended to star the search at zero). There was a burn in of 2000 and 8000 for the outer MCMC simulations searching through the space of the  $\alpha$ 's. The method is then currently on the magnitude the hundreds so 10,000 iterations, or two orders of magnitude larger will allow for proper search of the model space.

The proposed method will be compared under different simulation scenarios and in the NHS and HPFS data. Under each scenario the 2-Step BMA method will be compared to: Gold, Kitch, SW, HDPS, GBR, and AB-MC3.

## ***Simulation Results***

### *Comparing the Methods Under Different Simulation Scenarios*

The base scenario of simulations at  $\beta_X=0.0$ ,  $n=500$  and  $p=200$  with 2 covariates associated with X, 2 associated with Y, and 2 associated with X and Y are depicted in figures 3.1 and 1.2. First is a comparison of the average inclusion of each of the different  $\alpha$ 's over 1,000 (figure 3.1). Where both methods appear to be picking up the true sets of confounders. Using the same simulation scenarios as in figure 3.1 where are the biases of the different methods with the box plots labeled with the average bias and the MSE of each method under the abbreviation for

the method. The 2-Step, HDPS, and AB-MC3 appear to all be demonstrating similar level of unbiased estimation of the treatment effect compared relative to the gold standard.

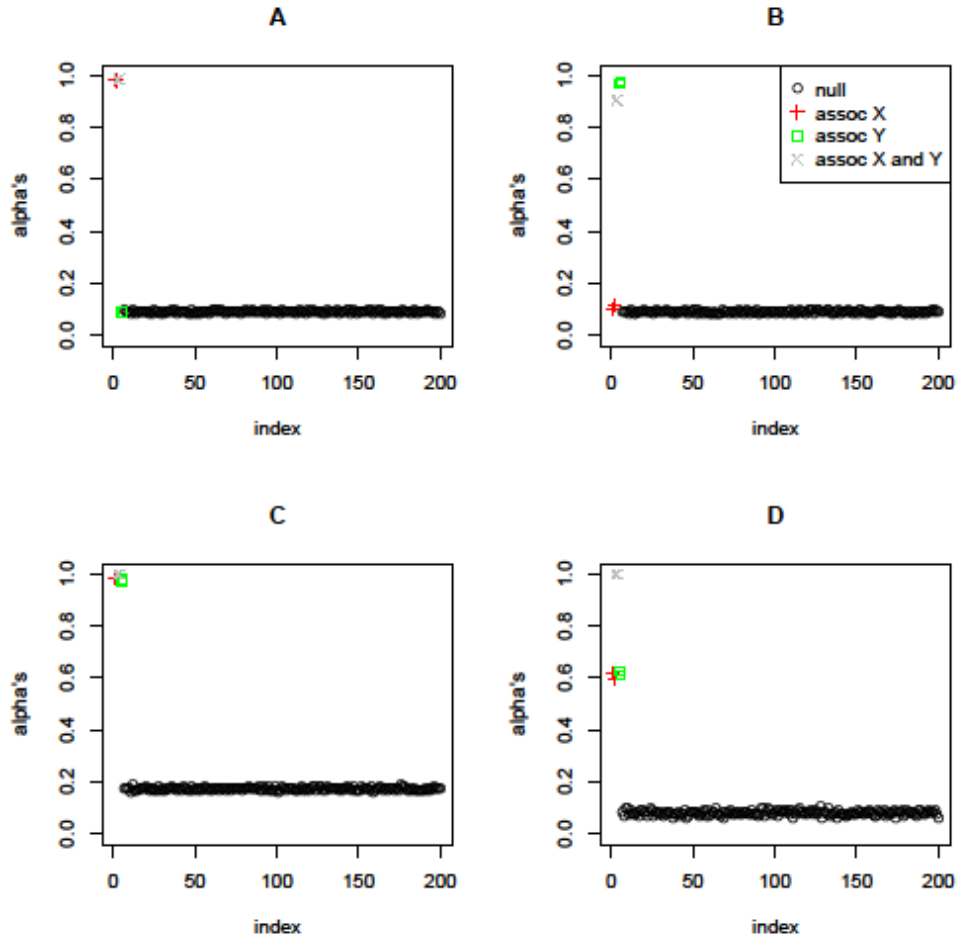


Figure 3.1: The average inclusion of each of the different  $\alpha$ 's over 1,000 simulations in the base setting of at  $\beta_X=0.0$ ,  $n=500$  and  $p=200$  with 2 covariates associated with X, 2 associated with Y, and 2 associated with X and Y. Panel 1A corresponds to the  $\alpha_x$ 's of the 2-Step, 1B are the  $\alpha_y$ , 1C are the union of  $\alpha_x$  and  $\alpha_y$ , and 1D corresponds to the HDPS method.

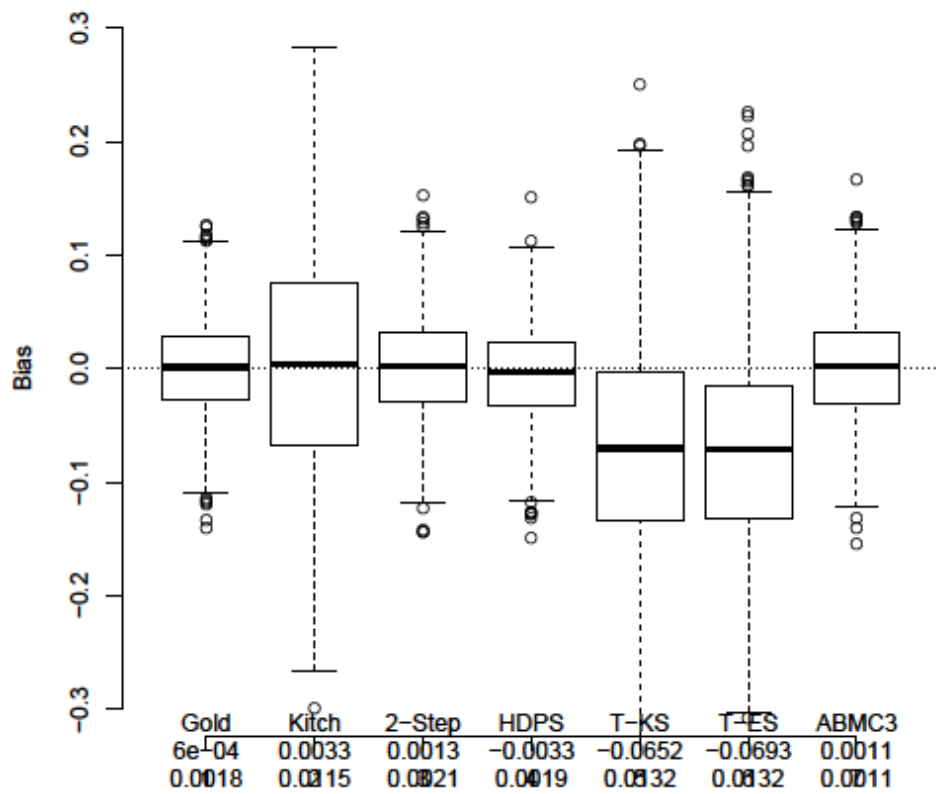


Figure 3.2: Using the same simulation scenarios as in figure 3.1 where are the biases of the different methods with the box plots labeled with the average bias and the MSE of each method under the abbreviation for the method.

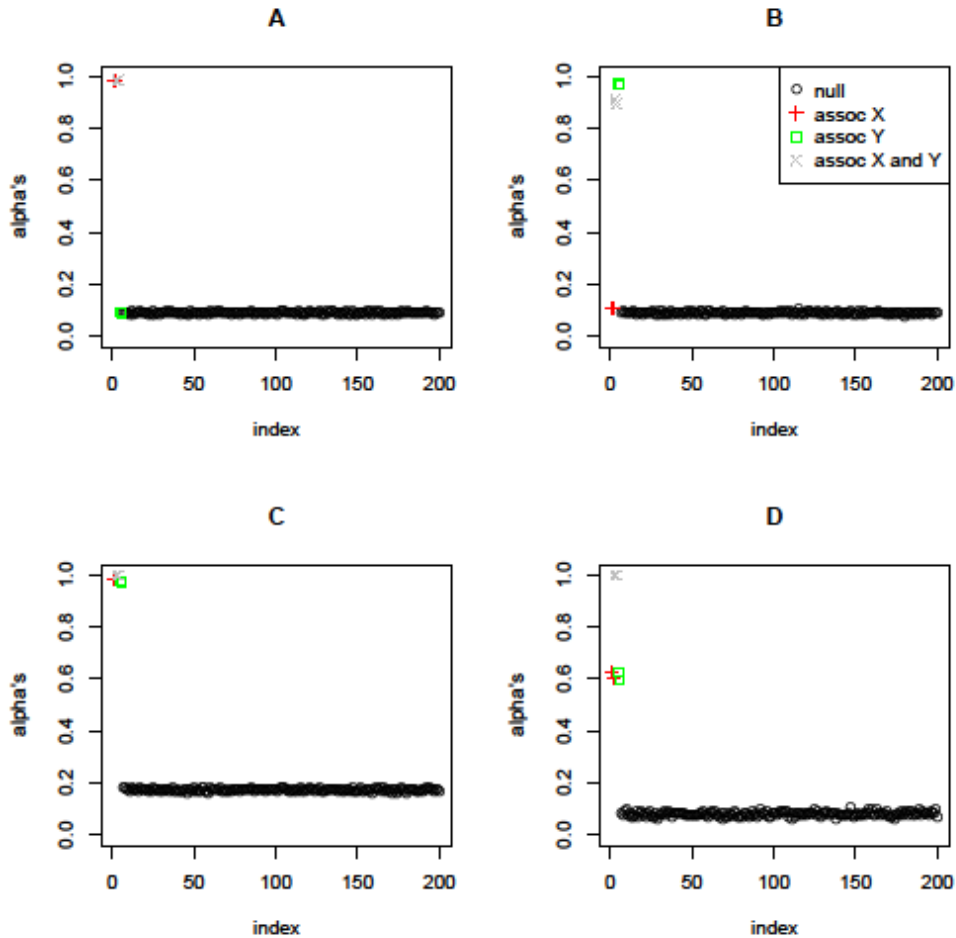


Figure 3.3: Now the treatment effect is simulated to be  $\beta_x=0.2$ . The average inclusion of each of the different  $\alpha$ 's over 1,000 simulations with  $n=500$ , and  $p=200$  with 2 covariates associated with X, 2 associated with Y, and 2 associated with X and Y. Panel 1A corresponds to the  $\alpha_x$ 's of the 2-Step, 1B are the  $\alpha_y$ , 1C are the union of  $\alpha_x$  and  $\alpha_y$ , and 1D corresponds to the HDPS method.

Next the settings of non-zero treatment effect were tested, first at  $\beta_x=0.2$  (Figures 3.3 and 4) and next at  $\beta_x=0.4$  (figures 3.5 and 3.6). The 2-Step and HDPS both appear to be picking up the expected models similar to figure 3.1 (figures 3.3 and 3.5) . Again the 2-Step and AB-MC3 methods appear to give unbiased estimates of the treatment effect, but the HDPS appears to be bias downward, potentially including covariates that are not truly association with Y biasing the observed ACE of the exposure on the outcome. This bias appears to increase as the treatment effect increases, resulting in more conservative estimates (figures 3.4 and 3.6).

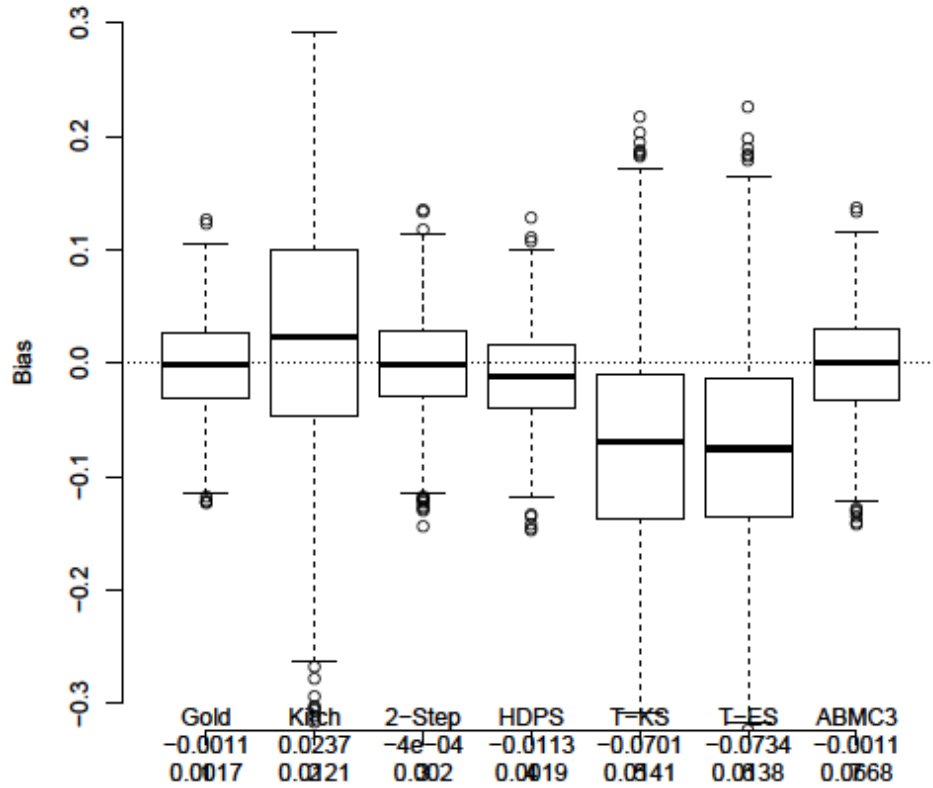


Figure 3.4: Using the same simulation scenarios as in figure 3.3, at  $\beta_X=0.2$ , where are the biases of the different methods with the box plots labeled with the average bias and the MSE of each method under the abbreviation for the method.

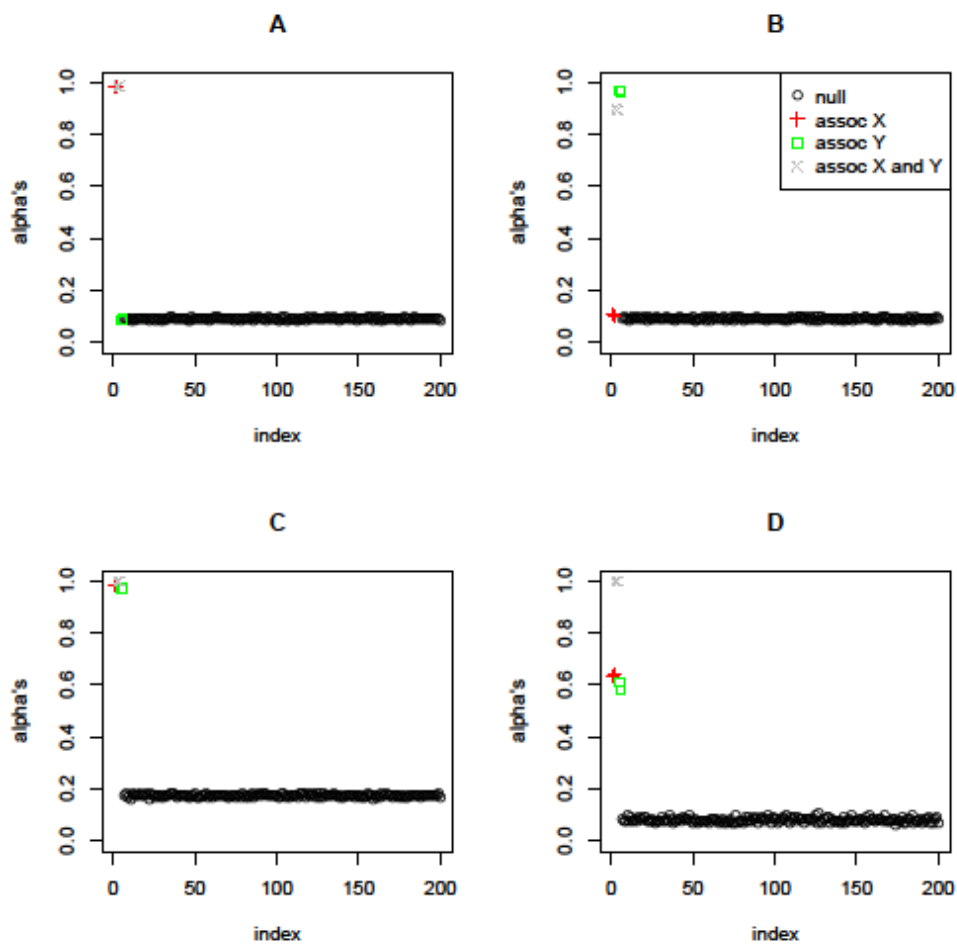


Figure 3.5: Now the treatment effect is simulated to be  $\beta_X=0.4$  The average inclusion of each of the different  $\alpha$ 's over 1,000 simulations with  $n=500$ , and  $p=200$  with 2 covariates associated with X, 2 associated with Y, and 2 associated with X and Y. Panel 1A corresponds to the  $\alpha_x$ 's of the 2-Step, 1B are the  $\alpha_y$ , 1C are the union of  $\alpha_x$  and  $\alpha_y$ , and 1D corresponds to the HDPS method.



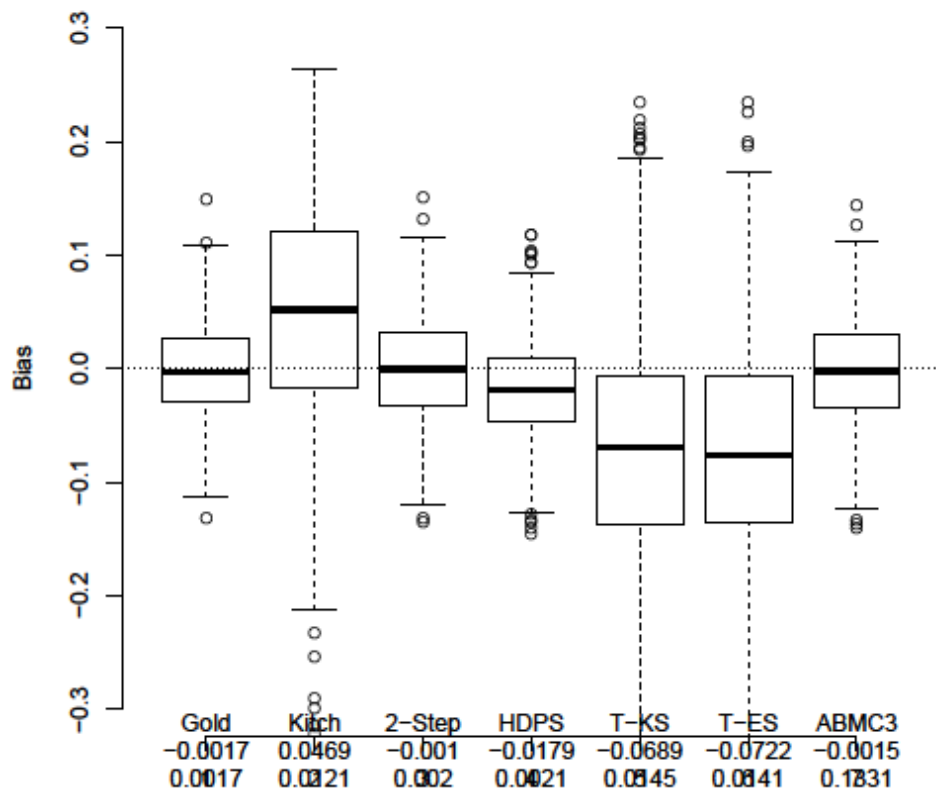


Figure 3.6: Using the same simulation scenarios as in figure 3.5, at  $\beta_X=0.4$ , where are the biases of the different methods with the box plots labeled with the average bias and the MSE of each method under the abbreviation for the method.

Then the scenario with 30 covariates associated with X, 30 associated with Y, and 30 associated with X and Y at very effect levels and each of these covariates is correlated with a null covariate (figures 3.7 and 3.8). Now it appears the HDPS method picks up the spuriously associated covariates (7D). There appears to be bias in all of the methods in this setting, however the 2-step method appears to be outperforming the AB-MC3 method.

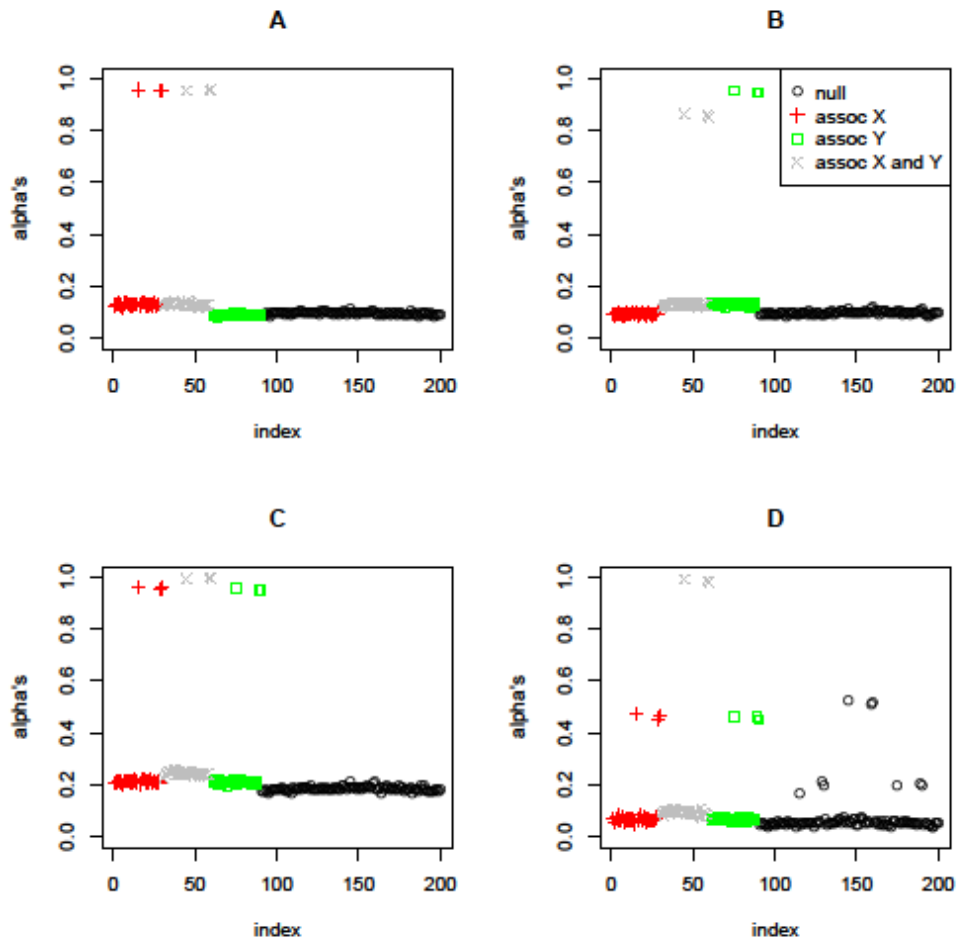


Figure 3.7: Now the treatment effect is simulated to be  $\beta_X=0.0$  but there are now with 30 covariates associated with X, 30 associated with Y, and 30 associated with X and Y at very effect levels. The average inclusion of each of the different  $\alpha$ 's over 1,000 simulations with  $n=500$ , and  $p=200$ . Panel 1A corresponds to the  $\alpha_x$ 's of the 2-Step, 1B are the  $\alpha_y$ , 1C are the union of  $\alpha_x$  and  $\alpha_y$ , and 1D corresponds to the HDPS method.

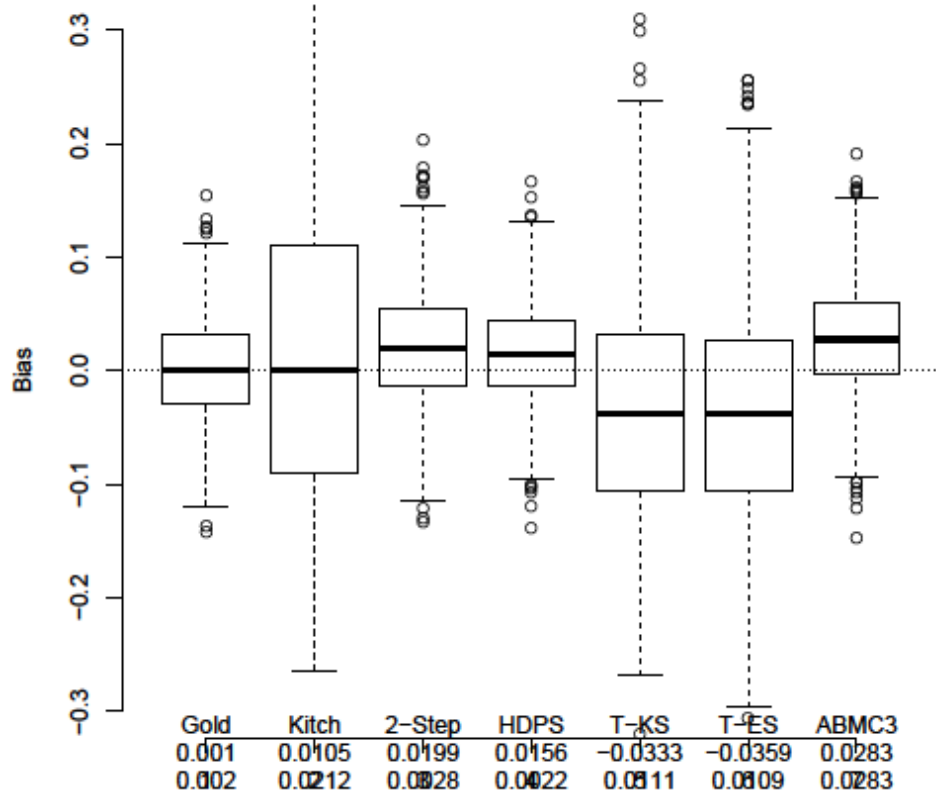


Figure 3.8: Using the same simulation scenarios as in figure 3.7, at varying effect levels and correlated covariates, where are the biases of the different methods with the box plots labeled with the average bias and the MSE of each method under the abbreviation for the method.

Figures 3.9 and 3.10 describes the are the posterior inclusion probabilities,  $\alpha$ , in the setting with 10 terms associated with X only, 10 terms associated with Y, and 10 confounders each with coefficients of +/-0.5 In figure a) are the results for the 2-Step BMA  $\alpha_x$  b) are the results for 2-Step BMA  $\alpha_y$  c) are the union of  $\alpha_x$  and  $\alpha_y$  d) HDPS using 10% of the covariates.

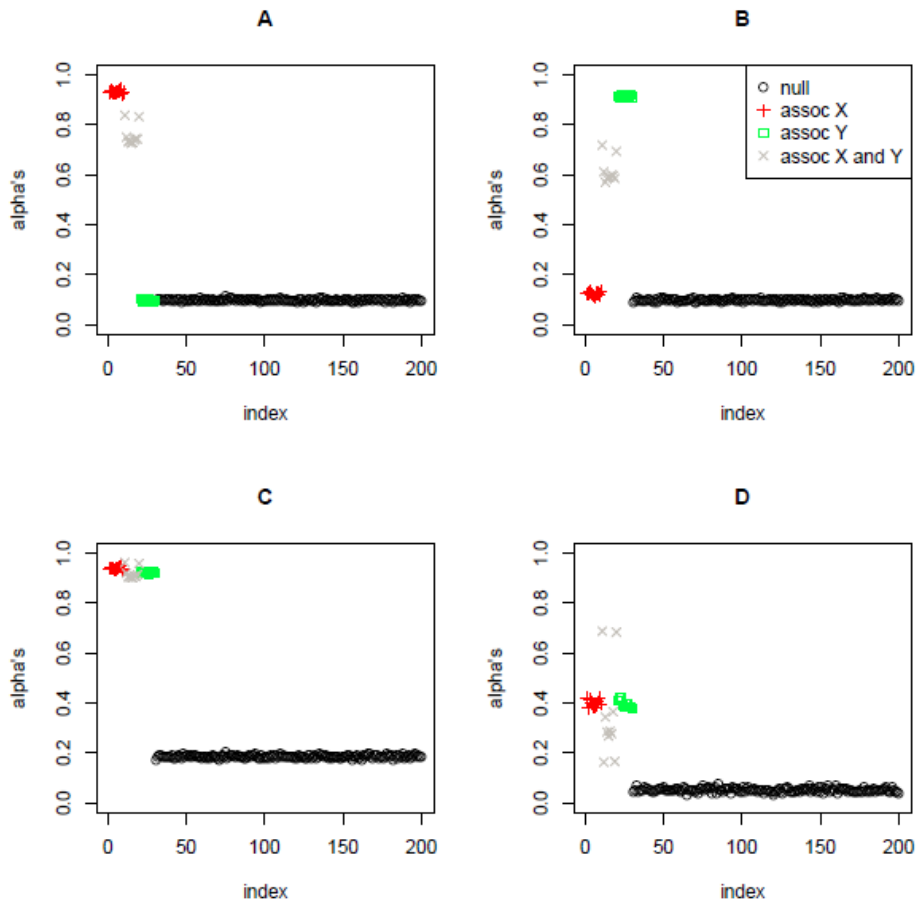


Figure 3.9: Now the treatment effect is simulated to be  $\beta_X=0.0$  but there are now with 10 covariates associated with X, 10 associated with Y, and 10 associated with X and Y with coefficients of 0.25. All of the confounders are now correlated using an autocorrelation scheme with a parameter of 0.4, having alternating directions of coefficient effect directionality. The average inclusion of each of the different  $\alpha$ 's over 1,000 simulations with  $n=500$ , and  $p=200$ . Panel 1A corresponds to the  $\alpha_x$ 's of the 2-Step, 1B are the  $\alpha_y$ , 1C are the union of  $\alpha_x$  and  $\alpha_y$ , and 1D corresponds to the HDPS method.

It appears HDPS is having a harder time picking up the correlated confounders (typically less than 50% of the time for the majority of the confounders) while, the 2-Step BMA appears to be considerably more effective at picking them up (over 90% of the time). The 2-Step BMA method demonstrates the least bias and MSE relative to all the other methods (Figure 3.10). The 2-Step BMA method appears to be the most robust to complicated confounder interactions.

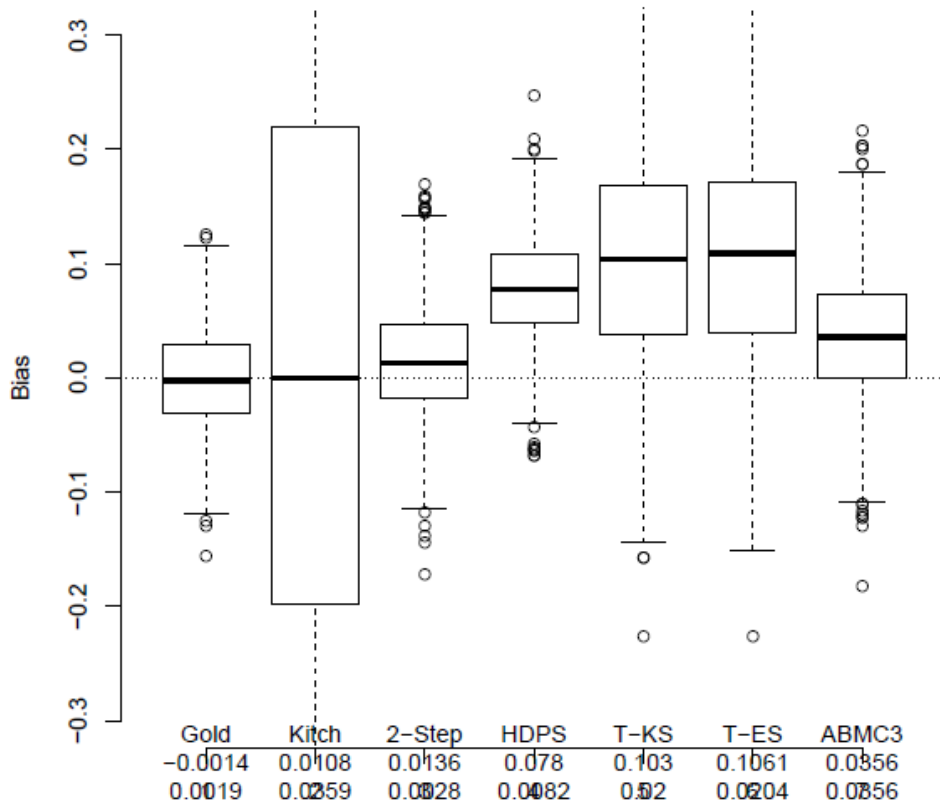


Figure 3.10: Using the same simulation scenarios as in figure 3.9, with autocorrelated confounders with alternating directionality of effect. The biases of the different methods with the box plots labeled with the average bias and the MSE of each method under the abbreviation for the method.

### *Overview of data Analysis*

Metabolomics are known to be associated with cancer<sup>48</sup> but, key metabolites are hard to isolate because there are many that are potentially associated; different techniques, such as principle component analysis (PCA), have been used to better understand and reduce the dimension of the problem<sup>53</sup>. Some known risk factors of PDAC have been identified as well as some of the complex biological pathways involved in the disease development; however, it is known to be associated with systemic metabolism<sup>54-56</sup> (Batch 2013 Metabolism, Lackey 2013 Am J Physiol Endocrinol Metab, Newgard 2009 Cell Metabolism, Mayers et. al. 2014 Nat Medicine). Further, it is known that BMI is associated with progression of PDAC. Elevated plasma levels of branched-chain amino acids (BCAAs) are associated with increased risk of pancreatic cancer and suspected to be associated with BMI as well. Two analyses will be performed to estimate the effect of elevated levels of BCAAs: one estimating the effect on BMI and another estimating the effect on PDAC. It is common in metabolic studies to group metabolites when studying their effects because they represent a complicated and dynamic biochemical state of the sample<sup>53</sup>.

We analyzed two prospective cohort studies, Health Professional Follow-Up Study (HPFS) and Nurses Health Study. Initial quality control was performed on the 592 subjects from NHS and HPFS consisting of 197 cases of PDAC. Removing individuals based on missing data left 453 subjects with 137 cases of PDAC. When BMI was considered the primary outcome of interest the subset of 316 control samples were used.

Quality control was done on the 133 targeted Metabolites, of those measured Metabolites 79 were used in both analyses. Of the over 50 metabolites that were excluded, they were

removed from the analysis for either poor stability with processing delay, or missing data at a level of greater than 10%. For PDAC there were a total of 99 covariates including metabolic profiles and epidemiological profiles and 98 for BMI (BMI was used a covariate for PDAC). The metabolite measures were transformed to the natural log scale. The primary exposure examined was the sum of the log-transformed plasma levels of three branched-chain amino acids (BCAAs): valine, leucine, and isoleucine, which have previously been shown to be linked to BMI and certain metabolic diseases<sup>48</sup>. The sum of the log of these metabolite levels were taken then dichotomized at the top quintile of expression, with those above this threshold being considered to have elevated expression.

#### *Data Analysis: Primary Outcome of BMI*

In Liu's<sup>49</sup> analysis confounding was adjusted for in a linear mixed model using the baseline covariates: baseline obesity status as a dummy variable (obese was defined as baseline BMI > 30 kg/m<sup>2</sup>), smoking status (never, past and current smokers) as dummy variables, physical activity (MET-hours) as a continuous variable, alcohol intake (g/day) as a continuous variable, total calorie intake (kcal/day) as a continuous variable, age (years) at blood draw and the follow-up years since baseline (take value 0 at baseline). In contrast, our analysis looks at baseline BMI was dichotomized with a cut-off of 30.

As a comparison to Liu's analysis both a marginal analysis of each BCAA was performed as well as a joint analysis after selecting out the top confounders found using the 2-Step BMA technique. The goal of this secondary analysis was to isolate if any discrepancy between Liu's results and our results had to potentially do with confounders as opposed to difference in defining the primary exposure.

Although the 2-Step BMA technique (as well as the HDPS adjustment) demonstrates an ATE of elevated levels BCAAs on BMI around zero (Table 3.4), it appears confounders that were previously uncontrolled for are now being incorporated. This should result in more accurate description of the treatment effect.

Method	Outcome	ATE	OR
<b>2-Step BMA</b>	PDAC	0.131 (0.075, 0.186)	1.807 (1.411, 2.272)
	BMI	0.030 (-0.025, 0.107)	1.364 (0.763, 2.444)
<b>HDPS</b>	PDAC	0.086 (-.039, 0.233 )	1.4815 (0.830, 2.765 )
	BMI	0.002 ( -0.079, 0.1739)	1.026 (0.377, 3.868)
<b>Conditional Logistic+ BMI, physical activity and reported diabetes at blood collection</b>	PDAC		1.89 (1.17,3.06)
<b>Conditional Logistic+ BMI, physical activity, reported diabetes, HbA1c, plasma insulin, proinsulin and C-peptide</b>	PDAC		2.19 (1.44,3.34)

Table 3.4: Comparison of ATE and OR with confidence/credible intervals in parenthesis for the 2-Step BMA and HDPS methods. The last two columns are results from Mayers et. al. 2014 Nat Med found in table 3.1. The primary exposure for those rows are top versus bottom quintile of log summed BCAAs.

*Data Analysis: Understanding the impact of Confounders with Primary Outcome of BMI*

The goal of this analysis was not to isolate a subset of confounders, better understanding their impact on elevated levels of BCAAs is important though. By using BMA techniques to account for the model uncertainty in this setting, our results indicate high levels of confounding (posterior probabilities of  $\alpha_x$  and  $\alpha_y$  listed in Table 3.5) maybe driving what was previously thought to be a clear association between elevated levels of BCAAs and BMI. The 2-Step BMA method found acetylglycine to be associated with BMI with inclusion probabilities are:  $p(\alpha_x)=7.8\%$  and  $p(\alpha_y)= 42.1\%$  (Liu’s found this to be a novel metabolite).



## Outcome

Selected Covariates	PDAC			BMI		
	HDPS	2S-BMA $P(\alpha_x)$	2S-BMA $P(\alpha_y)$	HDPS	2S-BMA $P(\alpha_x)$	2S-BMA $P(\alpha_y)$
phenylalanine	0	1.00	0.01	0	0.98	0.12
aminoadipate	0	1.00	0.02	1	0.97	0.09
cohort	0	0.73	0.01	0	0.64	0.11
smoking	0	0.77	0.02	0	0.39	0.38
anthranilic acid	0	0.63	0.01	0	0.64	0.05
glutamate	0	0.88	0.02	1	0.30	0.62
methionine	0	0.66	0.01	0	0.32	0.13
quinolinate	1	0.44	0.03	0	0.68	0.40
ADMA SDMA	1	0.05	0.03	0	0.08	0.06
alanine	1	0.06	0.01	0	0.46	0.07
malate	1	0.06	0.02	0	0.09	0.31
ornithine	1	0.07	0.01	0	0.16	0.06
proline	1	0.41	0.03	1	0.31	0.37
pyroglutamic acid	1	0.20	0.03	0	0.20	0.29
tryptophan	1	0.20	0.03	0	0.20	0.29
DMAPI	1	0.05	0.03	1	0.06	0.06
HAA	1	0.34	0.02	0	0.16	0.06
acetylglycine	0	0.06	0.02	1	0.08	0.42
glycine	0	0.08	0.02	1	0.10	0.32
methionine sulfoxide	0	0.06	0.01	1	0.08	0.05
tyrosine	0	0.12	0.02	1	0.15	0.08
xanthurenate	0	0.08	0.02	1	0.08	0.09

Table 3.5: Comparison of the posterior inclusion probabilities of the HDPS method and 2-Step BMA. The set of covariates either selected by HDPS and/or at a level of  $P(\alpha) > 0.50$  using the 2-Step BMA technique.

Interestingly, we're getting higher posterior probabilities for glutamate: with posterior inclusion probabilities of glutamate  $p(\alpha_x) = 30.1\%$  and  $p(\alpha_y) = 61.8\%$ ; this appears to be strong evidence for a potential confounder. Previously, this was not picked up by Liu. In terms of development of our methodology, we're less concerned with the posterior inclusion probabilities of these potential confounders but, can be thought of as a form of validation. Elevated levels of glutamate have previously been found to be associated with obesity and is thought to be associated with the regulation of appetite<sup>57</sup>. Glutamate is known to be associated with sophisticated functionality of neurotransmission modulation. Glutamatergic synapses have been found to be associated with a variety of neurobiological diseases as well as non-neurobiological

diseases such as osteoporosis <sup>58</sup>. Picking up these potential confounders bolsters the argument for BMA techniques that account for model uncertainty in high dimension settings, such as this one.

This is not a one to one comparison of the analysis done in Liu however, they looked at marginal effects of each metabolite separately. A follow up analysis was done, running a logistic regression marginally with each BCAA we found them to be significantly associated with continuous BMI (with p-values for valine, leucine, and isoleucine of: 0.000456, 0.0023, and, 0.00245 respectively), which is consistent with the findings of Liu. However, if we look at each BCAA individual but include the top 4 confounders (smoking, glutamate, quinolinate, and proline) picked up by the 2-Step BMA method as potential confounders, none of them are significant any more (Table 3.6). Interesting, glutamate also does not appear to be significantly associated. Smoking and proline aren't picked up by HDPS however, are significant in all joint analysis. Further, this may indicate elevated levels of BCAAs are less important in the biologic pathway resulting in BMI than previously expected.

	<b>BCAA Included in Joint Model</b>		
	<b>Valine</b>	<b>Leucine</b>	<b>Isoleucine</b>
<b>Valine</b>	0.089918	NA	NA
<b>Leucine</b>	NA	0.207617	NA
<b>Isoleucine</b>	NA	NA	0.223370
<b>Smoking</b>	0.032060	0.031895	0.029182
<b>Glutamate</b>	0.792327	0.741130	0.701842
<b>Quinolinate</b>	4.62e-05	2.6e-05	2.5e-05
<b>Proline</b>	0.013607	0.014261	0.019594

Table 3.6: The reported p-values for the different joint models, where the column indicates which BCAA is included in the analysis.

*Data Analysis: Primary Outcome of (PDAC)*

Initial quality control was performed on the 592 subjects from NHS and HPFS consisting of 197 cases of PDAC with a total of 99 covariates including metabolic profiles and epidemiological profiles. After removing individuals based on missing data, leaving 453 subjects with 137 cases of PDAC.

The HDPS method found the OR of elevated levels of BCAAs to be 1.4815 (0.830, 2.765), whereas 2-Step BMA method found 1.807 (1.411, 2.272) closer to the Mayers reported 1.89 (1.17–3.06) when they controlled for: BMI, physical activity, reported diabetes, HbA1c, plasma insulin, proinsulin and C-peptide. This provides validation for our method, in a setting of head to head comparison of our methods without outside scientific knowledge we observed similar results to a setting with expert analysis with expected confounders included in a conditional logistic analysis run by Mayers. It should be noted Mayers looked at top versus bottom quintile of log summed BCAAS as opposed to top versus lower 4 quintiles as in our analysis. The  $\alpha_x$  and  $\alpha_y$  posterior inclusion probabilities at or above levels of 50% and 90% and parameter selection of HDPS demonstrated little overlap (Table 3.4). This may indicate model selection techniques are not properly controlling for confounding or model uncertainty that while BMA techniques do. Both the kitchen sink and TWANG methods demonstrate unrealistically high treatment effects, indicating instability in statistical estimation of the ATE.

*Commentary across both PDAC and BMI*

We've introduced the 2-Step BMA method that targets primary exposure of interest while controlling for a high dimensional set of unknown confounders. Through simulation study, it was demonstrated decreased bias and MSE in certain settings versus state of the art techniques and that the method is less susceptible to spurious correlation among covariates. With the NHS and HPFS for the 2-Step BMA method demonstrated similar treatment effects to published results as opposed to the other methods.

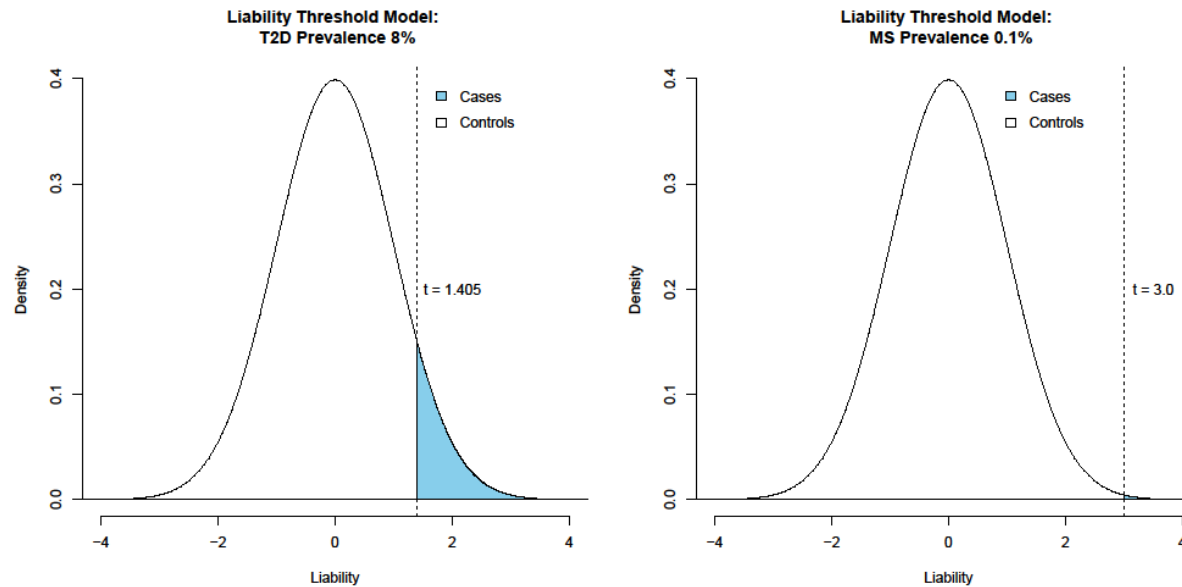
Not controlling for a strong confounder, such as glutamate when BMI was the outcome, has the potential to inflate the ATE of elevated levels of BCAAs on obesity that were otherwise observed. Our method controlled for glutamate and other potential confounders. The 2-Step BMA method provides effect estimates of BCAAs for PDAC that are on par with previous findings of Mayer, whereas HDPS demonstrates confidence intervals that include zero (Table 3.4).

Neither the HDPS nor the 2-Step BMA methods found a clearly significant effect of BCAAs when the outcome was BMI; however, Liu looked at the marginal effect of each BCAA which is a different set up than our approach. For our analysis we looked at the combined impact of all three BCAAs, log summed and dichotomized to above and below the top quintile, which may be the reason for the discrepancy. Additionally, our analysis focused on baseline BMI dichotomized above and below 30, possibly explaining some of the discrepancy with Liu. Further, the joint logistic regression follow up analysis (with the top 4 confounders found using the 2-Step BMA) did not demonstrate significant association between continuous BMI and valine, leucine, or isoleucine (Table 3.6).

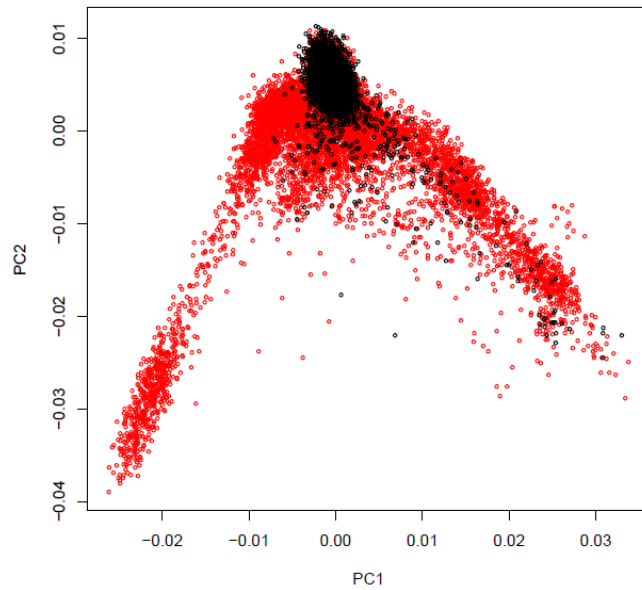
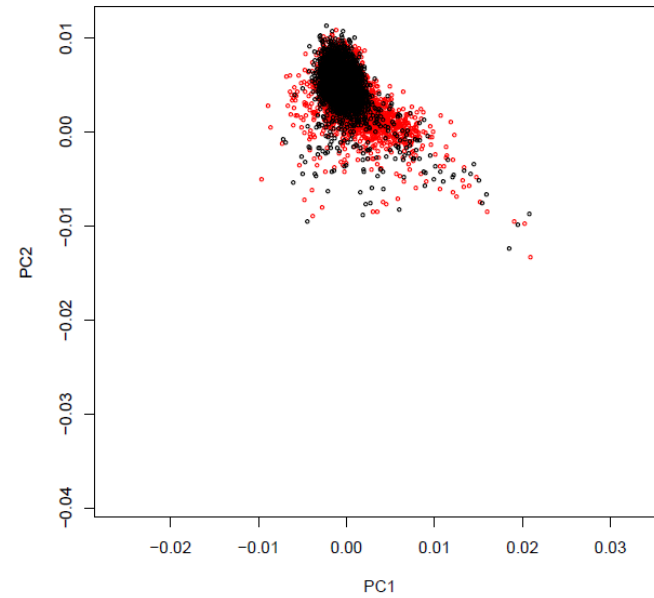
In both sets of analysis the metabolites used were the same and two metabolites in particular were found to be strongly associated with elevated levels of BCAAs using the 2-Step

BMA method: phenylalanine and amino adipate (4), demonstrating consistency in the method. This may point to the need for further investigation of the sets of metabolites treated as the primary exposure in future studies. HDPS only picked this up when looking at BMI as the outcome.

## Appendix 1 Supplemental Figures for Chapter 1



**Figure S.1.1. 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 S.1.1, 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  $PML_{\text{indiv}}$  for a control is 0.00 and 3.33 for a case. As the disease prevalence goes down the difference in the  $PML_{\text{indiv}}$  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 S.1.2. 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.

<b>Term</b>	<b>Description</b>
$\varphi$	Quantitative liability, the unobserved trait
$\beta$	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
$\pi$	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
$\Theta$	True underlying Genetic Relationship Matrix (GRM)
$V$	Phenotypic covariance matrix
$I$	Identity matrix
$h^2$	Heritability parameter

**Table S.1.1: Description of notation used and a brief description of the terms.**

<b>Computation</b>	<b>ATT</b>	<b>MLM</b>	<b>LTMLM</b>
<b>GRM and <math>V^{-1}</math></b>	NA	$O(MN^2)$	$O(MN^2)$
<b>PML</b>	NA	NA	$O(MN^2)$
<b>Assoc. Statistic</b>	$O(MN)$	$O(MN)$ or $O(MN^2)$	$O(MN)$
<b>Overall</b>	$O(MN)$	$O(MN^2)$	$O(MN^2)$

**Table S.1.2. Computational cost.**  $M$  is the number of SNPs and  $N$  is the number of individuals. We assume that  $M > N > \#MCMC$  iterations. The details of the computational costs of MLM are provided in Table 1 of ref<sup>1</sup>.



Table S.1.3

N	M	Prev	$\alpha$	set	ATT Liability	MLM Liability	ATT	LogR	MLM	LTMLM
<b>1000</b>	1000	50%	0.05	causal	0.732(0.015)	0.741(0.015)	0.663(0.016)	0.663(0.016)	0.671(0.016)	0.670(0.016)
			0.001	causal	0.548(0.017)	0.562(0.017)	0.444(0.017)	0.443(0.017)	0.447(0.017)	0.444(0.017)
			$1 \times 10^{-6}$	causal	0.331(0.016)	0.353(0.016)	0.227(0.014)	0.224(0.014)	0.236(0.014)	0.234(0.014)
			$5 \times 10^{-8}$	causal	0.275(0.015)	0.287(0.015)	0.179(0.013)	0.175(0.013)	0.177(0.013)	0.178(0.013)
			0.05	null	0.050(0.001)	0.049(0.001)	0.051(0.001)	0.051(0.001)	0.051(0.001)	0.051(0.001)
		0.001	null	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $2 \times 10^{-4}$ )	0.001( $2 \times 10^{-4}$ )	0.001( $2 \times 10^{-4}$ )	0.001( $2 \times 10^{-4}$ )	
		$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	
		$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	
		25%	0.05	causal	0.704(0.015)	0.712(0.015)	0.648(0.016)	0.648(0.016)	0.652(0.016)	0.653(0.016)
			0.001	causal	0.544(0.017)	0.557(0.017)	0.456(0.017)	0.455(0.017)	0.460(0.017)	0.466(0.017)
	$1 \times 10^{-6}$		causal	0.379(0.016)	0.383(0.016)	0.271(0.015)	0.268(0.015)	0.267(0.015)	0.272(0.015)	
	$5 \times 10^{-8}$		causal	0.318(0.016)	0.334(0.016)	0.219(0.014)	0.215(0.014)	0.226(0.014)	0.224(0.014)	
	0.05		null	0.051(0.001)	0.051(0.001)	0.050(0.001)	0.049(0.001)	0.050(0.001)	0.049(0.001)	
	0.001	null	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )		
	$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
	$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
	10%	50%	0.05	causal	0.761(0.014)	0.752(0.014)	0.708(0.015)	0.708(0.015)	0.708(0.015)	0.711(0.015)
			0.001	causal	0.589(0.016)	0.601(0.016)	0.51(0.017)	0.509(0.017)	0.514(0.017)	0.516(0.017)
			$1 \times 10^{-6}$	causal	0.410(0.016)	0.417(0.016)	0.325(0.016)	0.323(0.016)	0.325(0.016)	0.330(0.016)
			$5 \times 10^{-8}$	causal	0.362(0.016)	0.377(0.016)	0.266(0.015)	0.259(0.015)	0.268(0.015)	0.277(0.015)
0.05			null	0.050(0.001)	0.051(0.001)	0.049(0.001)	0.049(0.001)	0.050(0.001)	0.050(0.001)	
0.001		null	0.001( $2 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )		
$1 \times 10^{-6}$		null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
$5 \times 10^{-8}$		null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		

Table S.1.3 (Continued)

N	M	Prev	$\alpha$	set	ATT Liability	MLM Liability	ATT	LogR	MLM	LTMLM
		1%	0.05	causal	0.803(0.013)	0.796(0.013)	0.767(0.014)	0.767(0.014)	0.753(0.014)	0.769(0.014)
			0.001	causal	0.666(0.016)	0.665(0.016)	0.632(0.016)	0.632(0.016)	0.619(0.016)	0.638(0.016)
			$1 \times 10^{-6}$	causal	0.548(0.016)	0.543(0.016)	0.496(0.017)	0.491(0.017)	0.477(0.017)	0.501(0.017)
			$5 \times 10^{-8}$	causal	0.504(0.016)	0.500(0.016)	0.449(0.016)	0.447(0.016)	0.433(0.016)	0.451(0.016)
			0.05	null	0.050(0.001)	0.048(0.001)	0.050(0.001)	0.050(0.001)	0.049(0.001)	0.050(0.001)
			0.001	null	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
		0.1%	0.05	causal	0.820(0.013)	0.811(0.013)	0.815(0.013)	0.815(0.013)	0.801(0.013)	0.818(0.013)
			0.001	causal	0.727(0.015)	0.689(0.015)	0.697(0.015)	0.697(0.015)	0.656(0.016)	0.695(0.015)
			$1 \times 10^{-6}$	causal	0.585(0.016)	0.559(0.016)	0.551(0.016)	0.547(0.016)	0.510(0.017)	0.559(0.016)
			$5 \times 10^{-8}$	causal	0.548(0.016)	0.511(0.016)	0.512(0.016)	0.510(0.016)	0.475(0.017)	<b>0.525(0.016)</b>
			0.05	null	0.050(0.001)	0.046(0.001)	0.050(0.001)	0.049(0.001)	0.048(0.001)	0.050(0.001)
			0.001	null	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)

Table S.1.3 (Continued)

<b>N</b>	<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>	
<b>5000</b>	1000	50%	0.05	causal	0.700(0.016)	0.700(0.016)	0.642(0.017)	0.639(0.017)	0.637(0.017)	0.636(0.017)	
			0.001	causal	0.526(0.017)	0.529(0.017)	0.422(0.017)	0.420(0.017)	0.424(0.017)	0.424(0.017)	
			$1 \times 10^{-6}$	causal	0.337(0.016)	0.341(0.016)	0.234(0.015)	0.231(0.015)	0.240(0.015)	0.241(0.015)	
			$5 \times 10^{-8}$	causal	0.290(0.015)	0.288(0.015)	0.19(0.014)	0.185(0.013)	0.191(0.014)	0.190(0.013)	
			0.05	null	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	
			0.001	null	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	
	25%	1000	50%	0.05	causal	0.727(0.015)	0.723(0.015)	0.649(0.016)	0.649(0.016)	0.647(0.016)	0.646(0.016)
				0.001	causal	0.534(0.017)	0.537(0.017)	0.446(0.017)	0.444(0.017)	0.443(0.017)	0.445(0.017)
				$1 \times 10^{-6}$	causal	0.369(0.016)	0.376(0.016)	0.271(0.015)	0.269(0.015)	0.270(0.015)	0.270(0.015)
				$5 \times 10^{-8}$	causal	0.324(0.016)	0.324(0.016)	0.228(0.014)	0.227(0.014)	0.228(0.014)	0.228(0.014)
				0.05	null	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )
				0.001	null	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )
				$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
				$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
10%	1000	50%	0.05	causal	0.741(0.015)	0.747(0.015)	0.691(0.016)	0.690(0.016)	0.694(0.016)	0.694(0.016)	
			0.001	causal	0.579(0.016)	0.586(0.016)	0.518(0.017)	0.516(0.017)	0.520(0.017)	0.516(0.017)	
			$1 \times 10^{-6}$	causal	0.405(0.016)	0.403(0.016)	0.305(0.016)	0.303(0.016)	0.305(0.016)	0.307(0.016)	
			$5 \times 10^{-8}$	causal	0.349(0.016)	0.355(0.016)	0.248(0.015)	0.245(0.015)	0.249(0.015)	0.249(0.015)	
			0.05	null	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.051( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.051( $4 \times 10^{-4}$ )	
			0.001	null	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $7 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $7 \times 10^{-5}$ )	
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	

Table S.1.3 (Continued)

N	M	Prev	$\alpha$	set	ATT Liability	MLM Liability	ATT	LogR	MLM	LTMLM
		1%	0.05	causal	0.827(0.012)	0.834(0.012)	0.806(0.013)	0.806(0.013)	0.801(0.013)	0.807(0.013)
			0.001	causal	0.700(0.015)	0.700(0.015)	0.668(0.016)	0.664(0.016)	0.66(0.016)	0.663(0.016)
			$1 \times 10^{-6}$	causal	0.555(0.016)	0.551(0.016)	0.499(0.017)	0.496(0.017)	0.494(0.017)	0.503(0.017)
			$5 \times 10^{-8}$	causal	0.514(0.016)	0.509(0.016)	0.443(0.016)	0.441(0.016)	0.44(0.016)	0.446(0.016)
			0.05	null	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )
			0.001	null	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
		0.1%	0.05	causal	0.845(0.012)	0.854(0.012)	0.852(0.012)	0.852(0.012)	0.848(0.012)	0.852(0.012)
			0.001	causal	0.771(0.014)	0.770(0.014)	0.775(0.014)	0.775(0.014)	0.770(0.014)	0.770(0.014)
			$1 \times 10^{-6}$	causal	0.689(0.015)	0.667(0.015)	0.657(0.016)	0.657(0.016)	0.631(0.016)	<b>0.688(0.015)</b>
			$5 \times 10^{-8}$	causal	0.666(0.016)	0.654(0.016)	0.620(0.016)	0.593(0.016)	0.555(0.016)	0.590(0.016)
			0.05	null	0.049( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.051( $4 \times 10^{-4}$ )	0.051( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.051( $4 \times 10^{-4}$ )
			0.001	null	0.001( $7 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)

Table S.1.3 (Continued)

<b>N</b>	<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>LogR</b>	<b>N</b>	<b>M</b>	
<b>5000</b>	5000	50%	0.05	causal	0.696(0.007)	0.703(0.007)	0.630(0.007)	0.629(0.007)	0.633(0.007)	0.632(0.007)	
			0.001	causal	0.519(0.008)	0.530(0.008)	0.415(0.008)	0.415(0.008)	0.424(0.008)	0.424(0.008)	
			$1 \times 10^{-6}$	causal	0.332(0.007)	0.344(0.007)	0.229(0.006)	0.229(0.006)	0.236(0.007)	0.235(0.007)	
			$5 \times 10^{-8}$	causal	0.285(0.007)	0.295(0.007)	0.186(0.006)	0.184(0.006)	0.191(0.006)	0.190(0.006)	
		0.05	null	0.0491( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )		
		0.001	null	0.001 ( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )		
		$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
		$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
	25%	5000	50%	0.05	causal	0.694(0.007)	0.700(0.007)	0.628(0.007)	0.628(0.007)	0.632(0.007)	0.632(0.007)
				0.001	causal	0.512(0.008)	0.523(0.008)	0.427(0.008)	0.427(0.008)	0.434(0.008)	0.434(0.008)
				$1 \times 10^{-6}$	causal	0.344(0.007)	0.355(0.007)	0.260(0.007)	0.260(0.007)	0.265(0.007)	0.265(0.007)
				$5 \times 10^{-8}$	causal	0.297(0.007)	0.307(0.007)	0.211(0.006)	0.211(0.006)	0.215(0.006)	0.215(0.006)
		0.05	null	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )		
		0.001	null	0.001 ( $7 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001 ( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )		
		$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
		$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
10%	5000	50%	0.05	causal	0.751(0.006)	0.758(0.006)	0.704(0.007)	0.703(0.007)	0.706(0.007)	0.709(0.007)	
			0.001	causal	0.592(0.007)	0.597(0.007)	0.516(0.008)	0.516(0.008)	0.517(0.008)	0.520(0.008)	
			$1 \times 10^{-6}$	causal	0.428(0.007)	0.434(0.007)	0.335(0.007)	0.334(0.007)	0.335(0.007)	0.339(0.007)	
			$5 \times 10^{-8}$	causal	0.373(0.007)	0.384(0.007)	0.278(0.007)	0.277(0.007)	0.280(0.007)	0.285(0.007)	
	0.05	null	0.049( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.049( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )			
	0.001	null	0.001 ( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001 ( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )			
	$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)			

Table S.1.3 (Continued)

<b>N</b>	<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
		1%	0.05	causal	0.805(0.006)	0.803(0.006)	0.775(0.006)	0.775(0.006)	0.773(0.006)	0.780(0.006)
			0.001	causal	0.668(0.007)	0.664(0.007)	0.626(0.007)	0.626(0.007)	0.618(0.007)	0.631(0.007)
			$1 \times 10^{-6}$	causal	0.526(0.007)	0.516(0.007)	0.479(0.007)	0.479(0.007)	0.468(0.007)	0.485(0.007)
			$5 \times 10^{-8}$	causal	0.482(0.007)	0.469(0.007)	0.429(0.007)	0.428(0.007)	0.415(0.007)	0.433(0.007)
			0.05	null	0.050( $4 \times 10^{-4}$ )	0.045( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )
			0.001	null	0.001( $6 \times 10^{-5}$ )	0.001( $5 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
		0.1%	0.05	causal	0.833(0.005)	0.812(0.006)	0.818(0.006)	0.818(0.006)	0.799(0.006)	0.821(0.006)
			0.001	causal	0.726(0.007)	0.697(0.007)	0.702(0.007)	0.702(0.007)	0.670(0.007)	0.705(0.007)
			$1 \times 10^{-6}$	causal	0.599(0.007)	0.554(0.007)	0.562(0.007)	0.560(0.007)	0.516(0.007)	0.564(0.007)
			$5 \times 10^{-8}$	causal	0.554(0.007)	0.510(0.007)	0.516(0.007)	0.516(0.007)	0.470(0.007)	0.519(0.007)
			0.05	null	0.050( $4 \times 10^{-4}$ )	0.0324( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )	0.040( $4 \times 10^{-4}$ )	0.050( $4 \times 10^{-4}$ )
			0.001	null	0.001 ( $6 \times 10^{-5}$ )	$3 \times 10^{-4}$ ( $4 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )	0.001( $5 \times 10^{-5}$ )	0.001( $6 \times 10^{-5}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)

Table S.1.3 (Continued)

<b>N</b>	<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
<b>5000</b>	50000	50%	0.05	causal	0.702(0.007)	0.703(0.007)	0.631(0.007)	0.631(0.007)	0.631(0.007)	0.631(0.007)
			0.001	causal	0.518(0.008)	0.520(0.008)	0.418(0.008)	0.418(0.008)	0.419(0.008)	0.420(0.008)
			$1 \times 10^{-6}$	causal	0.331(0.007)	0.333(0.007)	0.231(0.006)	0.230(0.006)	0.232(0.006)	0.231(0.006)
			$5 \times 10^{-8}$	causal	0.278(0.007)	0.280(0.007)	0.183(0.006)	0.182(0.006)	0.183(0.006)	0.184(0.006)
			0.05	null	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )
			0.001	null	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
		25%	0.05	causal	0.706(0.007)	0.710(0.007)	0.652(0.007)	0.652(0.007)	0.652(0.007)	0.652(0.007)
			0.001	causal	0.539(0.008)	0.540(0.008)	0.450(0.008)	0.449(0.008)	0.450(0.008)	0.450(0.008)
			$1 \times 10^{-6}$	causal	0.362(0.007)	0.363(0.007)	0.259(0.007)	0.259(0.007)	0.262(0.007)	0.261(0.007)
			$5 \times 10^{-8}$	causal	0.308(0.007)	0.312(0.007)	0.207(0.006)	0.206(0.006)	0.209(0.006)	0.209(0.006)
			0.05	null	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )
			0.001	null	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
10%	0.05	causal	0.751(0.006)	0.755(0.006)	0.726(0.007)	0.726(0.007)	0.725(0.007)	0.720(0.007)		
	0.001	causal	0.629(0.007)	0.631(0.007)	0.548(0.008)	0.545(0.008)	0.555(0.007)	0.551(0.007)		
	$1 \times 10^{-6}$	causal	0.466(0.007)	0.474(0.007)	0.355(0.007)	0.355(0.007)	0.355(0.007)	0.355(0.007)		
	$5 \times 10^{-8}$	causal	0.381(0.007)	0.385(0.007)	0.272(0.007)	0.270(0.007)	0.274(0.007)	0.273(0.007)		
	0.05	null	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )		
	0.001	null	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )		
	$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
	$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		

Table S.1.3 (Continued)

<b>N</b>	<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
		1%	0.05	causal	0.793(0.006)	0.795(0.006)	0.770(0.006)	0.770(0.006)	0.769(0.006)	0.770(0.006)
			0.001	causal	0.671(0.007)	0.671(0.007)	0.634(0.007)	0.634(0.007)	0.634(0.007)	0.637(0.007)
			$1 \times 10^{-6}$	causal	0.526(0.007)	0.526(0.007)	0.477(0.007)	0.476(0.007)	0.476(0.007)	0.480(0.007)
			$5 \times 10^{-8}$	causal	0.484(0.007)	0.485(0.007)	0.431(0.007)	0.431(0.007)	0.429(0.007)	0.432(0.007)
			0.05	null	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )
			0.001	null	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
		0.1%	0.05	causal	0.835(0.005)	0.833(0.005)	0.821(0.006)	0.820(0.006)	0.817(0.006)	0.821(0.006)
			0.001	causal	0.719(0.007)	0.717(0.007)	0.697(0.007)	0.697(0.007)	0.688(0.007)	0.697(0.007)
			$1 \times 10^{-6}$	causal	0.588(0.007)	0.587(0.007)	0.561(0.007)	0.560(0.007)	0.557(0.007)	0.560(0.007)
			$5 \times 10^{-8}$	causal	0.547(0.007)	0.545(0.007)	0.517(0.007)	0.517(0.007)	0.508(0.007)	0.518(0.007)
			0.05	null	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )	0.050( $1 \times 10^{-4}$ )
			0.001	null	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )	0.001( $2 \times 10^{-5}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)

**Table S.1.3. Percentage of SNPs achieving alpha levels for simulated genotypes and simulated phenotypes.** We report the true positive and false positives at different  $\alpha$  levels. 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. In bold are the settings where LTMLM demonstrates at least a 5% power improvement over MLM.



Table S.1.4

<b>N</b>	<b>M</b>	<b>Prev</b>	<b>Set</b>	<b>Statistic</b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
<b>1000</b>	1000	50%	causal	average	25.716(1.091)	26.947(1.132)	16.847(0.715)	16.367(0.675)	17.254(0.735)	17.229(0.736)
			null	average	1.005(0.005)	1.003(0.005)	1.008(0.005)	1.006(0.005)	1.008(0.005)	1.007(0.005)
			all	average	1.252(0.014)	1.262(0.015)	1.166(0.010)	1.160(0.009)	1.171(0.010)	1.169(0.01)
			all	$\lambda_{GC}$	1.034(0.008)	1.033(0.008)	1.031(0.009)	1.031(0.009)	1.033(0.008)	1.032(0.008)
	25%	causal	average	28.716(1.229)	30.022(1.275)	19.112(0.824)	18.475(0.773)	19.553(0.849)	19.571(0.849)	
		null	average	1.003(0.005)	1.004(0.005)	1.000(0.004)	0.998(0.004)	1.002(0.004)	1.001(0.004)	
		all	average	1.280(0.016)	1.294(0.016)	1.181(0.011)	1.173(0.010)	1.187(0.011)	1.187(0.011)	
		all	$\lambda_{GC}$	1.019(0.007)	1.021(0.007)	1.014(0.007)	1.013(0.007)	1.026(0.007)	1.02(0.007)	
	10%	causal	average	34.541(1.501)	35.559(1.539)	23.777(1.021)	22.821(0.943)	24.084(1.053)	24.738(1.098)	
		null	average	1.001(0.005)	1.004(0.005)	1.001(0.004)	0.999(0.004)	1.004(0.004)	1.003(0.004)	
		all	average	1.336(0.019)	1.349(0.019)	1.229(0.013)	1.217(0.012)	1.235(0.014)	1.241(0.014)	
		all	$\lambda_{GC}$	1.021(0.008)	1.028(0.008)	1.025(0.008)	1.025(0.008)	1.030(0.008)	1.029(0.008)	
	1%	causal	average	55.732(2.12)	54.012(2.06)	42.316(1.565)	39.858(1.421)	40.611(1.539)	<b>45.336(1.756)</b>	
		null	average	1.002(0.004)	0.988(0.004)	0.999(0.004)	0.997(0.004)	0.992(0.004)	0.999(0.004)	
		all	average	1.550(0.028)	1.518(0.027)	1.412(0.021)	1.386(0.019)	1.388(0.020)	1.443(0.023)	
		all	$\lambda_{GC}$	1.027(0.008)	1.012(0.007)	1.02(0.008)	1.019(0.008)	1.009(0.008)	1.018(0.008)	
	0.1%	causal	average	73.159(2.847)	63.052(2.549)	58.946(2.157)	54.625(1.924)	51.507(2.045)	<b>65.991(2.656)</b>	
		null	average	0.997(0.004)	0.974(0.004)	0.996(0.004)	0.994(0.004)	0.978(0.004)	0.996(0.004)	
		all	average	1.718(0.037)	1.595(0.032)	1.575(0.029)	1.53(0.026)	1.483(0.026)	1.646(0.034)	
		all	$\lambda_{GC}$	1.035(0.009)	1.008(0.007)	1.023(0.007)	1.023(0.007)	1.000(0.007)	1.014(0.008)	

Table S.1.4 (Continued)

N	M	Prev	Set	Statistic	ATT Liability	MLM Liability	ATT	LogR	MLM	LTMLM
<b>5000</b>	1000	50%	causal	average	25.224(1.076)	25.503(1.084)	16.488(0.703)	16.029(0.666)	16.652(0.713)	16.625(0.714)
			null	average	1.002(0.002)	1.002(0.002)	1.000(0.002)	0.998(0.002)	1.000(0.002)	0.999(0.002)
			all	average	1.050(0.003)	1.051(0.003)	1.031(0.003)	1.028(0.003)	1.032(0.003)	1.031(0.003)
			all	$\lambda_{GC}$	1.006(0.003)	1.007(0.003)	1.005(0.003)	1.005(0.003)	1.006(0.003)	1.005(0.004)
	25%	causal	average	28.012(1.156)	28.279(1.163)	18.545(0.767)	17.985(0.726)	18.707(0.778)	18.757(0.783)	
		null	average	1.004(0.002)	1.004(0.002)	1.002(0.002)	1.000(0.002)	1.002(0.002)	1.002(0.002)	
		all	average	1.058(0.004)	1.059(0.004)	1.037(0.003)	1.034(0.003)	1.038(0.003)	1.037(0.003)	
		all	$\lambda_{GC}$	1.010(0.003)	1.011(0.003)	1.003(0.003)	1.003(0.003)	1.004(0.003)	1.004(0.003)	
	10%	causal	average	32.953(1.379)	33.282(1.395)	22.883(0.948)	22.028(0.886)	23.065(0.965)	23.497(0.996)	
		null	average	1.005(0.002)	1.005(0.002)	1.004(0.002)	1.002(0.002)	1.004(0.002)	1.003(0.002)	
		all	average	1.069(0.004)	1.070(0.004)	1.047(0.003)	1.044(0.003)	1.048(0.003)	1.048(0.003)	
		all	$\lambda_{GC}$	1.009(0.003)	1.009(0.004)	0.999(0.003)	0.999(0.003)	1.003(0.003)	1.001(0.003)	
	1%	causal	average	57.101(2.112)	56.546(2.100)	43.459(1.570)	40.939(1.426)	43.546(1.611)	<b>46.302(1.760)</b>	
		null	average	1.002(0.002)	1.002(0.002)	1.001(0.002)	0.999(0.002)	1.001(0.002)	1.001(0.002)	
		all	average	1.115(0.006)	1.113(0.006)	1.086(0.005)	1.079(0.004)	1.086(0.005)	1.091(0.005)	
		all	$\lambda_{GC}$	1.007(0.004)	1.009(0.003)	1.003(0.003)	1.002(0.003)	1.005(0.003)	1.003(0.004)	
	0.1%	causal	average	74.089(2.623)	68.943(2.51)	61.561(2.057)	57.182(1.828)	58.752(2.08)	<b>67.284(2.450)</b>	
		null	average	0.997(0.002)	0.993(0.002)	0.996(0.002)	0.995(0.002)	0.991(0.002)	0.996(0.002)	
		all	average	1.143(0.007)	1.129(0.007)	1.117(0.006)	1.107(0.005)	1.106(0.006)	1.128(0.007)	
		all	$\lambda_{GC}$	1.004(0.001)	0.994(0.002)	0.989(0.001)	0.988(0.001)	1.005(0.002)	0.996(0.002)	

Table S.1.4 (Continued)

N	M	Prev	Set	Statistic	ATT Liability	MLM Liability	ATT	LogR	MLM	LTMLM
<b>5000</b>	5000	50%	causal	average	25.372(0.496)	26.69(0.52)	16.492(0.325)	16.399(0.321)	16.88(0.332)	16.867(0.332)
			null	average	0.993(0.002)	0.994(0.002)	0.988(0.002)	0.988(0.002)	0.99(0.002)	0.989(0.002)
			all	average	1.237(0.006)	1.251(0.007)	1.143(0.004)	1.142(0.004)	1.148(0.004)	1.148(0.004)
			all	$\lambda_{GC}$	1.010(0.004)	1.01(0.003)	1.010(0.003)	1.010(0.003)	1.014(0.003)	1.014(0.003)
	25%	causal	average	28.119(0.588)	29.56(0.617)	18.637(0.388)	18.509(0.383)	19.014(0.396)	19.056(0.398)	
		null	average	1.001(0.002)	1.001(0.002)	1.000(0.002)	1.000(0.002)	1.000(0.002)	1.001(0.002)	
		all	average	1.272(0.007)	1.287(0.008)	1.177(0.005)	1.175(0.005)	1.180(0.005)	1.181(0.005)	
		all	$\lambda_{GC}$	1.019(0.004)	1.017(0.004)	1.012(0.004)	1.012(0.004)	1.013(0.004)	1.013(0.004)	
	10%	causal	average	36.38(0.716)	37.626(0.742)	25.235(0.501)	25.014(0.492)	25.386(0.506)	25.778(0.514)	
		null	average	0.996(0.002)	0.994(0.002)	0.993(0.002)	0.992(0.002)	0.991(0.002)	0.992(0.002)	
		all	average	1.350(0.009)	1.361(0.009)	1.235(0.006)	1.233(0.006)	1.235(0.006)	1.240(0.007)	
		all	$\lambda_{GC}$	1.016(0.003)	1.017(0.004)	1.005(0.003)	1.005(0.003)	1.007(0.003)	1.008(0.003)	
	1%	causal	average	58.172(1.131)	55.473(1.077)	45.376(0.878)	44.682(0.852)	42.594(0.825)	<b>46.691(0.913)</b>	
		null	average	1.000(0.002)	0.954(0.002)	1.000(0.002)	0.999(0.002)	0.990(0.002)	1(0.002)	
		all	average	1.571(0.014)	1.499(0.013)	1.444(0.011)	1.436(0.011)	1.406(0.010)	1.457(0.011)	
		all	$\lambda_{GC}$	1.020(0.003)	0.970(0.005)	1.020(0.003)	1.020(0.003)	1.011(0.003)	1.019(0.003)	
	0.1%	causal	average	81.745(1.565)	66.716(1.29)	68.648(1.301)	67.099(1.248)	56.303(1.082)	<b>70.81(1.364)</b>	
		null	average	1.001(0.002)	0.840(0.002)	1.000(0.002)	1.000(0.002)	0.918(0.002)	1.000(0.002)	
		all	average	1.809(0.019)	1.499(0.016)	1.677(0.016)	1.661(0.016)	1.472(0.013)	1.698(0.017)	
		all	$\lambda_{GC}$	1.022(0.003)	0.861(0.006)	1.026(0.003)	1.026(0.003)	0.942(0.005)	1.025(0.003)	

Table S.1.4 (Continued)

N	M	Prev	Set	Statistic	ATT Liability	MLM Liability	ATT	LogR	MLM	LTMLM
5000	50000	50%	causal	average	25.501(0.504)	25.642(0.507)	16.624(0.331)	16.529(0.327)	16.673(0.332)	16.69(0.333)
			null	average	0.999(0.001)	0.999(0.001)	1.000 (0.001)	0.999(0.001)	1.000 (0.001)	1.000 (0.001)
			all	average	1.023(0.001)	1.024(0.001)	1.015(0.001)	1.015(0.001)	1.016(0.001)	1.015(0.001)
			all	$\lambda_{GC}$	0.999(0.001)	1.000 (0.001)	1.003(0.001)	1.002(0.001)	0.999(0.001)	0.999(0.001)
	25%	causal	average	28.475(0.555)	28.638(0.557)	18.965(0.37)	18.843(0.366)	19.03(0.372)	19.04(0.372)	
		null	average	1.003(0.001)	1.003(0.001)	1.003(0.001)	1.003(0.001)	1.003(0.001)	1.003(0.001)	
		all	average	1.030(0.001)	1.031(0.001)	1.021(0.001)	1.020(0.001)	1.021(0.001)	1.021(0.001)	
		all	$\lambda_{GC}$	1.004(0.001)	1.004(0.001)	1.006(0.001)	1.006(0.001)	1.003(0.001)	1.003(0.001)	
	10%	causal	average	34.358(0.639)	34.704(0.647)	23.71(0.444)	23.528(0.437)	23.868(0.449)	23.910(0.450)	
		null	average	0.999(0.001)	1.000 (0.001)	1.001(0.001)	1.000 (0.001)	1.001(0.001)	1.001(0.001)	
		all	average	1.033(0.001)	1.034(0.001)	1.023(0.001)	1.023(0.001)	1.024(0.001)	1.024(0.001)	
		all	$\lambda_{GC}$	1.006(0.001)	1.008(0.002)	1.007(0.001)	1.007(0.001)	1.005(0.001)	1.004(0.001)	
	1%	causal	average	60.112(1.139)	59.863(1.135)	46.683(0.883)	45.969(0.859)	46.44(0.881)	47.368(0.905)	
		null	average	1.000 (0.001)	1.000 (0.001)	0.999(0.001)	0.999(0.001)	0.999(0.001)	0.999(0.001)	
		all	average	1.059(0.002)	1.058(0.002)	1.045(0.001)	1.044(0.001)	1.045(0.001)	1.045(0.001)	
		all	$\lambda_{GC}$	1.002(0.001)	1.002(0.001)	1.004(0.001)	1.004(0.001)	1.001(0.001)	1.000 (0.001)	
	0.1%	causal	average	79.864(1.54)	77.754(1.502)	67.059(1.278)	65.561(1.225)	65.232(1.251)	<b>68.618(1.333)</b>	
		null	average	1.000 (0.001)	0.999(0.001)	0.999(0.001)	0.999(0.001)	0.999(0.001)	0.999(0.001)	
		all	average	1.078(0.002)	1.076(0.002)	1.065(0.002)	1.063(0.002)	1.063(0.002)	1.067(0.002)	
		all	$\lambda_{GC}$	1.002(0.001)	1.001(0.001)	1.004(0.001)	1.004(0.001)	1.000(0.001)	1.000(0.001)	

**Table S.1.4. Complete results on simulated genotypes and simulated phenotypes.** Results are analogous to Table 2, but are reported for other values of  $M$  and  $N$  and consist of the same simulations as S3. 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. In bold are the settings where LTMLM demonstrates at least a 5% power improvement over MLM.

<b>M</b>	<b>Prev</b>	<b>Set</b>	<b>Statistic</b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>ATT+PCs</b>	<b>LogR</b>	<b>LogR +PCs</b>	<b>MLM</b>	<b>LTMLM</b>
<b>1000</b>	10%	<b>causal</b>	average	36.835 (1.515)	16.352 (0.683)	24.736 (1.002)	23.406(0.951)	23.622(0.93)	22.399(0.883)	13.222(0.553)	12.172(0.536)
		<b>null</b>	average	2.238 (0.032)	0.789 (0.014)	1.814 (0.022)	1.632(0.020)	1.780(0.020)	1.601(0.019)	0.849(0.012)	0.753(0.011)
		<b>all</b>	average	2.584 (0.036)	0.944 (0.016)	2.043 (0.025)	1.849(0.023)	1.999(0.023)	1.809(0.022)	0.973(0.013)	0.867(0.013)
		<b>all</b>	$\lambda_{GC}$	1.376 (0.042)	0.373 (0.006)	1.240 (0.028)	1.064(0.010)	1.239(0.028)	1.062(0.010)	0.518(0.007)	0.458(0.011)
<b>10000</b>	10%	<b>causal</b>	average	33.100 (1.375)	29.318 (1.217)	22.606 (0.943)	21.99(0.919)	21.672(0.876)	21.104(0.854)	20.412(0.867)	19.817(0.874)
		<b>null</b>	average	1.342 (0.004)	0.965 (0.003)	1.232 (0.003)	1.084(0.003)	1.226(0.003)	1.078(0.002)	0.976(0.002)	0.948(0.002)
		<b>all</b>	average	1.374 (0.004)	0.993 (0.003)	1.253 (0.003)	1.104(0.003)	1.246(0.003)	1.098(0.003)	0.995(0.003)	0.967(0.003)
		<b>all</b>	$\lambda_{GC}$	1.258 (0.036)	0.882 (0.003)	1.174 (0.025)	1.023(0.003)	1.174(0.025)	1.023(0.003)	0.920(0.003)	0.893(0.013)
<b>20000</b>	10%	<b>causal</b>	average	34.93 (1.481)	32.625 (1.39)	24.098 (1.037)	23.592(1.014)	23.031(0.963)	22.568(0.943)	22.892(1.003)	22.811(1.006)
		<b>null</b>	average	1.281 (0.002)	0.982 (0.002)	1.187 (0.002)	1.042(0.001)	1.182(0.002)	1.038(0.001)	0.986(0.001)	0.981(0.001)
		<b>all</b>	average	1.298 (0.002)	0.998 (0.002)	1.198 (0.002)	1.053(0.002)	1.193(0.002)	1.048(0.001)	0.997(0.002)	0.992(0.002)
		<b>all</b>	$\lambda_{GC}$	1.247 (0.029)	0.939 (0.002)	1.168 (0.020)	1.013(0.002)	1.167(0.020)	1.013(0.002)	0.958(0.002)	0.953(0.006)

**Table S.1.5. Results on simulated genotypes and simulated phenotypes with population structure.** We report average  $\chi^2$  statistics for simulations with population structure averaged across 100 simulations for each parameter setting (see main text).

Table S.1.6

<b>N</b>	<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>ATT+PCs</b>	<b>LogR</b>	<b>LogR +PCs</b>	<b>MLM</b>	<b>LTMLM</b>			
<b>1000</b>	1000	10%	0.05	causal	0.758 (0.014)	0.661 (0.016)	0.719 (0.015)	0.715 (0.015)	0.719 (0.015)	0.716 (0.015)	0.611 (0.017)	0.596 (0.017)			
				causal	0.623 (0.016)	0.421 (0.017)	0.541 (0.017)	0.535 (0.017)	0.540 (0.017)	0.530 (0.017)	0.36 (0.017)	0.334 (0.016)			
			1x10 <sup>-6</sup>	causal	0.435 (0.016)	0.220 (0.014)	0.323 (0.016)	0.316 (0.016)	0.315 (0.016)	0.307 (0.016)	0.183 (0.013)	0.153 (0.013)			
			5x10 <sup>-8</sup>	causal	0.373 (0.016)	0.176 (0.013)	0.274 (0.015)	0.261 (0.015)	0.265 (0.015)	0.256 (0.015)	0.143 (0.012)	0.116 (0.011)			
			0.05	null	0.097 (0.001)	0.020 (0.001)	0.086 (0.001)	0.069 (0.001)	0.085 (0.001)	0.069 (0.001)	0.022 (0.001)	0.020 (0.001)			
			0.001	null	0.022 (0.001)	0.012 (0.001)	0.018 (0.001)	0.016 (0.001)	0.018 (0.001)	0.016 (0.001)	0.011 (0.001)	0.010 (0.001)			
			1x10 <sup>-6</sup>	null	0.012 (5x10 <sup>-4</sup> )	0.006 (4x10 <sup>-4</sup> )	0.010 (4x10 <sup>-4</sup> )	0.009 (4x10 <sup>-4</sup> )	0.009 (4x10 <sup>-4</sup> )	0.009 (4x10 <sup>-4</sup> )	0.005 (4x10 <sup>-4</sup> )	0.004 (3x10 <sup>-4</sup> )			
			5x10 <sup>-8</sup>	null	0.011 (5x10 <sup>-4</sup> )	0.005 (4x10 <sup>-4</sup> )	0.008 (4x10 <sup>-4</sup> )	0.008 (4x10 <sup>-4</sup> )	0.008 (4x10 <sup>-4</sup> )	0.007 (4x10 <sup>-4</sup> )	0.004 (3x10 <sup>-4</sup> )	0.003 (3x10 <sup>-4</sup> )			
			<b>1000</b>	10000		0.05	causal	0.745 (0.015)	0.729 (0.015)	0.707 (0.015)	0.702 (0.016)	0.707 (0.015)	0.702 (0.016)	0.682 (0.016)	0.674 (0.016)
							causal	0.587 (0.016)	0.557 (0.017)	0.495 (0.017)	0.492 (0.017)	0.494 (0.017)	0.492 (0.017)	0.475 (0.017)	0.451 (0.017)
1x10 <sup>-6</sup>	causal	0.411 (0.016)				0.380 (0.016)	0.322 (0.016)	0.314 (0.016)	0.316 (0.016)	0.311 (0.016)	0.295 (0.016)	0.280 (0.015)			
5x10 <sup>-8</sup>	causal	0.347 (0.016)				0.315 (0.016)	0.270 (0.015)	0.269 (0.015)	0.266 (0.015)	0.259 (0.015)	0.245 (0.015)	0.225 (0.014)			
0.05	null	0.08 (4x10 <sup>-4</sup> )				0.038 (3x10 <sup>-4</sup> )	0.071 (4x10 <sup>-4</sup> )	0.054 (3x10 <sup>-4</sup> )	0.071 (4x10 <sup>-4</sup> )	0.053 (3x10 <sup>-4</sup> )	0.042 (3x10 <sup>-4</sup> )	0.04 (3x10 <sup>-4</sup> )			
0.001	null	0.006 (1x10 <sup>-4</sup> )				0.002 (7x10 <sup>-5</sup> )	0.004 (9x10 <sup>-5</sup> )	0.003 (7x10 <sup>-5</sup> )	0.004 (9x10 <sup>-5</sup> )	0.003 (7x10 <sup>-5</sup> )	0.002 (6x10 <sup>-5</sup> )	0.002 (7x10 <sup>-5</sup> )			
1x10 <sup>-6</sup>	null	0.001 (5x10 <sup>-5</sup> )				0.001 (5x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )			
5x10 <sup>-8</sup>	null	0.001 (5x10 <sup>-5</sup> )				0.001 (5x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )	0.001 (4x10 <sup>-5</sup> )			

Table S.1.6 (Continued)

<b>1000</b>	<b>20000</b>	<b>0.05</b>	<b>causal</b>	<b>0.762</b> <b>(0.014)</b>	<b>0.753</b> <b>(0.014)</b>	<b>0.690</b> <b>(0.016)</b>	<b>0.687</b> <b>(0.016)</b>	<b>0.690</b> <b>(0.016)</b>	<b>0.686</b> <b>(0.016)</b>	<b>0.686</b> <b>(0.016)</b>	<b>0.682</b> <b>(0.016)</b>
		0.001	causal	0.588 (0.016)	0.578 (0.017)	0.512 (0.017)	0.506 (0.017)	0.511 (0.017)	0.505 (0.017)	0.490 (0.017)	0.488 (0.017)
		$1 \times 10^{-6}$	causal	0.412 (0.016)	0.400 (0.016)	0.315 (0.016)	0.309 (0.016)	0.311 (0.016)	0.306 (0.016)	0.297 (0.015)	0.300 (0.016)
		$5 \times 10^{-8}$	causal	0.356 (0.016)	0.340 (0.016)	0.264 (0.015)	0.262 (0.015)	0.260 (0.015)	0.258 (0.015)	0.254 (0.015)	0.254 (0.015)
		0.05	null	0.078 ( $3 \times 10^{-4}$ )	0.043 ( $2 \times 10^{-4}$ )	0.069 ( $2 \times 10^{-5}$ )	0.052 ( $2 \times 10^{-4}$ )	0.068 ( $2 \times 10^{-4}$ )	0.051 ( $2 \times 10^{-4}$ )	0.046 ( $2 \times 10^{-4}$ )	0.045 ( $2 \times 10^{-4}$ )
		0.001	null	0.005 ( $7 \times 10^{-5}$ )	0.002 ( $4 \times 10^{-5}$ )	0.003 ( $6 \times 10^{-5}$ )	0.002 ( $4 \times 10^{-5}$ )	0.003 ( $5 \times 10^{-5}$ )	0.00 ( $2 \times 4 \times 10^{-5}$ )	0.002 ( $4 \times 10^{-5}$ )	0.002 ( $4 \times 10^{-5}$ )
		$1 \times 10^{-6}$	null	0.001 ( $3 \times 10^{-5}$ )	0.001 ( $3 \times 10^{-5}$ )	0.001 ( $2 \times 10^{-5}$ )	0 (NA)	0 (NA)	0 (NA)	0 (NA)	0 (NA)
		$5 \times 10^{-8}$	null	0.001 ( $2 \times 10^{-5}$ )	0.001 ( $2 \times 10^{-5}$ )	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)

**Table S.1.6. Percentage of SNPs achieving alpha levels for simulated genotypes and simulated phenotypes with population structure.** We report the true positive and false positives at different  $\alpha$  levels for the same simulations as in Table S.1.5 (see main text).

<b>M</b>	Prevalence	<b>Liability</b>		<b>Observed</b>	
		H-E	REML	H-E	REML
<b>1000</b>	50%	0.259(0.013)	0.252(0.01)	0.165(0.008)	0.161(0.006)
	25%	0.241(0.010)	0.238(0.008)	0.173(0.007)	0.171(0.006)
	10%	0.245(0.011)	0.242(0.007)	0.233(0.010)	0.230(0.007)
<b>10000</b>	50%	0.236(0.014)	0.245(0.013)	0.150(0.009)	0.156(0.008)
	25%	0.250(0.014)	0.264(0.013)	0.180(0.010)	0.190(0.010)
	10%	0.259(0.012)	0.261(0.009)	0.246(0.011)	0.248(0.009)

**Table S.1.7. Heritability parameter estimates on simulated genotypes and phenotypes.** Results are analogous to Table 3, under different settings of  $M$  and  $N$ .

<b>M</b>	Prevalence	<b>Liability</b>		<b>Observed</b>	
		H-E	REML	H-E	REML
<b>1000</b>	10%	0.470(0.013)	0.415(0.005)	0.446(0.012)	0.394(0.005)
<b>10000</b>	10%	0.407(0.031)	0.526(0.011)	0.438(0.044)	0.500(0.010)
<b>20000</b>	10%	0.383(0.030)	0.531(0.013)	0.408(0.040)	0.505(0.012)

**Table S.1.8. Heritability parameter estimates on simulated genotypes and phenotypes with population structure.** These results are from the same simulations used to generate Table S.1.5 and S6. 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.



Table S.1.9

<b>N</b>	<b>M</b>	<b>True Prev</b>	<b>Specified Prev</b>	<b>Set</b>	<b>Statistic</b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>	
<b>1000</b>	1000	1%	1%	causal	average	55.732(2.12)	54.012(2.06)	42.316(1.565)	39.858(1.421)	40.611(1.539)	45.336(1.756)	
				null	average	1.002(0.004)	0.988(0.004)	0.999(0.004)	0.997(0.004)	0.992(0.004)	0.999(0.004)	
				all	average	1.550(0.028)	1.518(0.027)	1.412(0.021)	1.386(0.019)	1.388(0.020)	1.443(0.023)	
				all	$\lambda$ GC	1.027(0.008)	1.012(0.007)	1.020(0.008)	1.019(0.008)	1.009(0.008)	1.018(0.008)	
			3.4%	causal	average	As Above	As Above	As Above	As Above	As Above	As Above	44.441(1.805)
			null	average	As Above	As Above	As Above	As Above	As Above	As Above	1.009(0.005)	
			all	average	As Above	As Above	As Above	As Above	As Above	As Above	1.444(0.023)	
			all	$\lambda$ GC	As Above	As Above	As Above	As Above	As Above	As Above	1.035(0.008)	
			0.2%	causal	average	As Above	As Above	As Above	As Above	As Above	As Above	45.532(1.851)
			null	average	As Above	As Above	As Above	As Above	As Above	As Above	1.002(0.004)	
			all	average	As Above	As Above	As Above	As Above	As Above	As Above	1.448(0.024)	
			all	$\lambda$ GC	As Above	As Above	As Above	As Above	As Above	As Above	1.023(0.008)	

Table S.1.9 (Continued)

	0.1%	0.1%	causal	average	73.159(2.847)	63.052(2.549)	58.946(2.157)	54.625(1.924)	51.507(2.045)	65.991(2.656)
			null	average	0.997(0.004)	0.974(0.004)	0.996(0.004)	0.994(0.004)	0.978(0.004)	0.996(0.004)
			all	average	1.718(0.037)	1.595(0.032)	1.575(0.029)	1.530(0.026)	1.483(0.026)	1.646(0.034)
			all	$\lambda$ GC	1.035(0.009)	1.008(0.007)	1.023(0.007)	1.023(0.007)	1.000(0.007)	1.014(0.008)
	0.5%		causal	average	As Above	As Above	As Above	As Above	As Above	65.281(2.617)
			null	average	As Above	As Above	As Above	As Above	As Above	0.998(0.004)
			all	average	As Above	As Above	As Above	As Above	As Above	1.641(0.033)
			all	$\lambda$ GC	As Above	As Above	As Above	As Above	As Above	1.022(0.007)
	0.02%		causal	average	As Above	As Above	As Above	As Above	As Above	68.616(2.729)
			null	average	As Above	As Above	As Above	As Above	As Above	1.003(0.005)
			all	average	As Above	As Above	As Above	As Above	As Above	1.679(0.035)
			all	$\lambda$ GC	As Above	As Above	As Above	As Above	As Above	1.022(0.008)

**Table S.1.9. Simulated genotypes and phenotypes with mis-specification of the liability threshold.** LTMLM was run at prevalence of 1% and 0.1% under mis-specification of the threshold,  $t = \text{true} \pm 0.5$ .

Table S.1.10

<b>N</b>	<b>M</b>	<b>True Prev</b>	<b>Specified Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT Liability</b>	<b>LogR Liability</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>	
<b>5000</b>	5000	1%	1%	0.05	causal	0.803(0.013)	0.796(0.013)	0.767(0.014)	0.767(0.014)	0.753(0.014)	0.769(0.014)	
				0.001	causal	0.666(0.016)	0.665(0.016)	0.632(0.016)	0.632(0.016)	0.619(0.016)	0.638(0.016)	
				1x10-6	causal	0.548(0.016)	0.543(0.016)	0.496(0.017)	0.491(0.017)	0.477(0.017)	0.501(0.017)	
				5x10-8	causal	0.504(0.016)	0.500(0.016)	0.449(0.016)	0.447(0.016)	0.433(0.016)	0.451(0.016)	
				0.05	null	0.050(0.001)	0.048(0.001)	0.050(0.001)	0.050(0.001)	0.049(0.001)	0.050(0.001)	
				0.001	null	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )	
			1x10-6	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
			5x10-8	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		
			3.4%	0.05	causal	As Above	As Above	As Above	As Above	As Above	As Above	0.783(0.014)
			0.001	causal	As Above	As Above	As Above	As Above	As Above	As Above	0.647(0.016)	
			1x10-6	causal	As Above	As Above	As Above	As Above	As Above	As Above	0.469(0.017)	
			5x10-8	causal	As Above	As Above	As Above	As Above	As Above	As Above	0.428(0.016)	
		0.05	null	As Above	As Above	As Above	As Above	As Above	As Above	0.051(0.001)		
		0.001	null	As Above	As Above	As Above	As Above	As Above	As Above	0.001(1x10 <sup>-4</sup> )		
		1x10-6	null	As Above	As Above	As Above	As Above	As Above	As Above	0(NA)		
		5x10-8	null	As Above	As Above	As Above	As Above	As Above	As Above	0(NA)		
		0.2%	0.05	causal	As Above	As Above	As Above	As Above	As Above	As Above	0.774(0.014)	
		0.001	causal	As Above	As Above	As Above	As Above	As Above	As Above	0.638(0.016)		
		1x10-6	causal	As Above	As Above	As Above	As Above	As Above	As Above	0.491(0.017)		
		5x10-8	causal	As Above	As Above	As Above	As Above	As Above	As Above	0.439(0.016)		
		0.05	null	As Above	As Above	As Above	As Above	As Above	As Above	0.050(0.001)		
		0.001	null	As Above	As Above	As Above	As Above	As Above	As Above	0.001(1x10 <sup>-4</sup> )		
		1x10-6	null	As Above	As Above	As Above	As Above	As Above	As Above	0(NA)		
		5x10-8	null	As Above	As Above	As Above	As Above	As Above	As Above	0(NA)		

Table S.1.10 (Continued)

<b>N</b>	<b>M</b>	<b>True Prev</b>	<b>Specified Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT Liability</b>	<b>LogR Liability</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
		0.1%	0.1%	0.05	causal	0.820(0.013)	0.811(0.013)	0.815(0.013)	0.815(0.013)	0.801(0.013)	0.818(0.013)
				0.001	causal	0.727(0.015)	0.689(0.015)	0.697(0.015)	0.697(0.015)	0.656(0.016)	0.695(0.015)
				1x10-6	causal	0.585(0.016)	0.559(0.016)	0.551(0.016)	0.547(0.016)	0.510(0.017)	0.559(0.016)
				5x10-8	causal	0.548(0.016)	0.511(0.016)	0.512(0.016)	0.510(0.016)	0.475(0.017)	0.525(0.016)
				0.05	null	0.050(0.001)	0.046(0.001)	0.050(0.001)	0.049(0.001)	0.048(0.001)	0.050(0.001)
				0.001	null	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )	0.001(1x10 <sup>-4</sup> )
				1x10-6	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
				5x10-8	null	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
			0.5%	0.05	causal	As Above	As Above	As Above	As Above	As Above	0.823(0.013)
				0.001	causal	As Above	As Above	As Above	As Above	As Above	0.717(0.015)
				1x10-6	causal	As Above	As Above	As Above	As Above	As Above	0.569(0.016)
				5x10-8	causal	As Above	As Above	As Above	As Above	As Above	0.526(0.016)
				0.05	null	As Above	As Above	As Above	As Above	As Above	0.049(0.001)
				0.001	null	As Above	As Above	As Above	As Above	As Above	0.001(2x10 <sup>-4</sup> )
				1x10-6	null	As Above	As Above	As Above	As Above	As Above	0(NA)
				5x10-8	null	As Above	As Above	As Above	As Above	As Above	0(NA)
			0.02%	0.05	causal	As Above	As Above	As Above	As Above	As Above	0.824(0.012)
				0.001	causal	As Above	As Above	As Above	As Above	As Above	0.716(0.015)
				1x10-6	causal	As Above	As Above	As Above	As Above	As Above	0.573(0.016)
				5x10-8	causal	As Above	As Above	As Above	As Above	As Above	0.523(0.016)
				0.05	null	As Above	As Above	As Above	As Above	As Above	0.050(0.001)
				0.001	null	As Above	As Above	As Above	As Above	As Above	0.001(1x10 <sup>-4</sup> )
				1x10-6	null	As Above	As Above	As Above	As Above	As Above	0(NA)
				5x10-8	null	As Above	As Above	As Above	As Above	As Above	0(NA)

Table S.1.10 (Continued)

**Table S.1.10. Percentage of SNPs achieving alpha levels for simulated genotypes and simulated phenotypes with mis-specification of the liability threshold.** This corresponds to the same set of simulations as Table S.1.9.

<b>M</b>	<b>Prevalence</b>		<b>Liability</b>		<b>Observed</b>	
	True	Specified	H-E	REML	H-E	REML
<b>1000</b>	1%	1%	0.246(0.007)	0.206(0.005)	0.445(0.013)	0.373(0.008)
		3.4%	0.350(0.012)	0.290(0.007)	0.462(0.016)	0.383(0.010)
		0.2%	0.182(0.006)	0.152(0.004)	0.445(0.014)	0.373(0.009)
	0.1%	0.1%	0.247(0.008)	0.175(0.004)	0.701(0.023)	0.498(0.011)
		0.5%	0.324(0.009)	0.234(0.004)	0.689(0.019)	0.499(0.009)
		0.02%	0.184(0.006)	0.132(0.003)	0.677(0.022)	0.486(0.011)

**Table S.1.11. Heritability parameter estimates on simulated genotypes and phenotypes with misspecification of the liability threshold.** This corresponds to the same set of simulations as Table S.1.9 and S10.

<b>N</b>	<b>M</b>	<b>Prev</b>	<b>Set</b>	<b>Statistic</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
<b>1000</b>	1000	50%	causal	average	18.085(0.248)	17.859(0.242)	18.582(0.252)	18.576(0.254)
			null	average	0.997(0.004)	0.996(0.004)	0.998(0.004)	0.997(0.004)
			all	average	1.168(0.007)	1.164(0.007)	1.174(0.008)	1.173(0.008)
			all	$\lambda_{GC}$	1.023(0.007)	1.022(0.007)	1.026(0.008)	1.02(0.007)
	25%	causal	average	18.531(0.246)	18.295(0.24)	18.987(0.251)	18.954(0.253)	
		null	average	1.005(0.005)	1.003(0.004)	1.005(0.005)	1.004(0.005)	
		all	average	1.180(0.008)	1.176(0.007)	1.185(0.008)	1.183(0.008)	
		all	$\lambda_{GC}$	1.037(0.007)	1.036(0.007)	1.036(0.007)	1.035(0.007)	
	10%	causal	average	20.926(0.268)	20.628(0.26)	21.259(0.272)	21.347(0.272)	
		null	average	1.005(0.004)	1.003(0.004)	1.005(0.004)	1.004(0.004)	
		all	average	1.204(0.008)	1.200(0.008)	1.207(0.008)	1.207(0.008)	
		all	$\lambda_{GC}$	1.031(0.007)	1.031(0.007)	1.036(0.008)	1.037(0.008)	
	1%	causal	average	28.421(0.326)	27.886(0.314)	27.697(0.323)	28.974(0.338)	
		null	average	1.009(0.005)	1.007(0.005)	1.008(0.005)	1.008(0.005)	
		all	average	1.283(0.010)	1.276(0.010)	1.275(0.010)	1.288(0.010)	
		all	$\lambda_{GC}$	1.027(0.008)	1.027(0.008)	1.025(0.008)	1.028(0.008)	
	0.1%	causal	average	35.927(0.358)	35.100(0.342)	33.172(0.347)	36.654(0.373)	
		null	average	0.999(0.004)	0.997(0.004)	0.992(0.004)	0.998(0.004)	
		all	average	1.348(0.012)	1.338(0.012)	1.314(0.012)	1.355(0.013)	
		all	$\lambda_{GC}$	1.031(0.007)	1.031(0.007)	1.012(0.007)	1.022(0.007)	

**Table S.1.12. Simulated genotypes and phenotypes generated from a logit distribution.** We report average  $\chi^2$  statistics for simulations with phenotypes generated from a logit distribution averaged across 100 simulations for each parameter setting (see main text).

Table S.1.13

<b>N</b>	<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
<b>1000</b>	1000	50%	0.05	causal	0.988(0.003)	0.988(0.003)	0.988(0.003)	0.988(0.003)
			0.001	causal	0.827(0.012)	0.821(0.012)	0.832(0.012)	0.832(0.012)
			$1 \times 10^{-6}$	causal	0.210(0.013)	0.198(0.013)	0.227(0.013)	0.230(0.013)
			$5 \times 10^{-8}$	causal	0.077(0.008)	0.072(0.008)	0.089(0.009)	0.088(0.009)
			0.05	null	0.049(0.001)	0.049(0.001)	0.050(0.001)	0.050(0.001)
			0.001	null	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)
		25%	0.05	causal	0.997(0.002)	0.997(0.002)	0.997(0.002)	0.995(0.002)
			0.001	causal	0.832(0.012)	0.830(0.012)	0.849(0.011)	0.844(0.011)
			$1 \times 10^{-6}$	causal	0.225(0.013)	0.209(0.013)	0.243(0.014)	0.245(0.014)
			$5 \times 10^{-8}$	causal	0.090(0.009)	0.081(0.009)	0.098(0.009)	0.093(0.009)
			0.05	null	0.050(0.001)	0.050(0.001)	0.050(0.001)	0.050(0.001)
			0.001	null	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)
		10%	0.05	causal	0.997(0.002)	0.997(0.002)	0.998(0.001)	0.998(0.001)
			0.001	causal	0.893(0.010)	0.892(0.010)	0.894(0.010)	0.894(0.010)
			$1 \times 10^{-6}$	causal	0.344(0.015)	0.334(0.015)	0.357(0.015)	0.365(0.015)
			$5 \times 10^{-8}$	causal	0.140(0.011)	0.129(0.011)	0.154(0.011)	0.160(0.012)
0.05	null		0.051(0.001)	0.051(0.001)	0.051(0.001)	0.050(0.001)		
0.001	null		0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )		
$1 \times 10^{-6}$	null		0(NA)	0(NA)	0(NA)	0(NA)		
$5 \times 10^{-8}$	null		0(NA)	0(NA)	0(NA)	0(NA)		



Table S.1.13 (Continued)

<b>N</b>	<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>set</b>	<b>ATT</b>	<b>LogR</b>	<b>MLM</b>	<b>LTMLM</b>
		1%	0.05	causal	1(NA)	1(NA)	1(NA)	1(NA)
			0.001	causal	0.978(0.005)	0.977(0.005)	0.973(0.005)	0.980(0.004)
			$1 \times 10^{-6}$	causal	0.640(0.015)	0.623(0.015)	0.613(0.015)	0.658(0.015)
			$5 \times 10^{-8}$	causal	0.397(0.015)	0.378(0.015)	0.376(0.015)	0.419(0.016)
			0.05	null	0.051(0.001)	0.050(0.001)	0.051(0.001)	0.051(0.001)
			0.001	null	0.001( $2 \times 10^{-4}$ )	0.001( $2 \times 10^{-4}$ )	0.001( $2 \times 10^{-4}$ )	0.001( $2 \times 10^{-4}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)
		0.1%	0.05	causal	1(NA)	1(NA)	1(NA)	1(NA)
			0.001	causal	0.996(0.002)	0.996(0.002)	0.995(0.002)	0.997(0.002)
			$1 \times 10^{-6}$	causal	0.868(0.011)	0.860(0.011)	0.797(0.013)	0.880(0.01)
			$5 \times 10^{-8}$	causal	0.685(0.015)	0.664(0.015)	0.584(0.016)	0.702(0.014)
			0.05	null	0.050(0.001)	0.050(0.001)	0.049(0.001)	0.050(0.001)
			0.001	null	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )	0.001( $1 \times 10^{-4}$ )
			$1 \times 10^{-6}$	null	0(NA)	0(NA)	0(NA)	0(NA)
			$5 \times 10^{-8}$	null	0(NA)	0(NA)	0(NA)	0(NA)

**Table S.1.13. Percentage of SNPs achieving alpha levels for simulated genotypes and phenotypes generated from a logistic distribution.** Results are the same simulations as described in Table S.1.12

Table S.1.14

<b>M</b>	<b>Prev</b>	<b>Set</b>	<b>Statistic</b>	<b>ATT Liability</b>	<b>MM Liability</b>	<b>ATT</b>	<b>ATT+PCs</b>	<b>LogR</b>	<b>LogR+PCs</b>	<b>MLM</b>	<b>LTMLM</b>
<b>1000</b>	50%	causal	average	25.684 (1.137)	29.055 (1.26)	16.234 (0.723)	15.775 (0.705)	15.566(0.674)	15.14(0.658)	17.480(0.780)	17.412(0.775)
		null	average	1.698 (0.015)	1.792 (0.017)	1.444 (0.011)	1.425 (0.010)	1.433(0.010)	1.415(0.010)	1.472(0.011)	1.471(0.011)
		all	average	1.938 (0.021)	2.065 (0.023)	1.592 (0.014)	1.569 (0.013)	1.574(0.013)	1.552(0.013)	1.632(0.015)	1.630(0.015)
		all	$\lambda_{GC}$	1.274 (0.018)	1.278 (0.019)	1.214 (0.017)	1.224 (0.017)	1.214(0.017)	1.223(0.017)	1.234(0.017)	1.226(0.016)
	25%	causal	average	30.169 (1.211)	33.771 (1.343)	19.277 (0.771)	18.581 (0.754)	18.47(0.721)	17.822(0.706)	20.493(0.82)	20.642(0.831)
		null	average	1.821 (0.017)	1.917 (0.019)	1.551 (0.012)	1.507 (0.011)	1.538(0.012)	1.495(0.011)	1.571(0.012)	1.577(0.013)
		all	average	2.104 (0.023)	2.236 (0.025)	1.728 (0.015)	1.678 (0.014)	1.707(0.015)	1.658(0.014)	1.761(0.016)	1.768(0.016)
		all	$\lambda_{GC}$	1.290 (0.018)	1.288 (0.019)	1.256 (0.016)	1.245 (0.017)	1.255(0.016)	1.245(0.017)	1.241(0.015)	1.243(0.015)
	10%	causal	average	34.865 (1.45)	37.861 (1.551)	22.838 (0.961)	21.406 (0.878)	21.618(0.881)	20.336(0.808)	23.865(1.022)	24.661(1.064)
		null	average	1.973 (0.020)	2.039 (0.021)	1.664 (0.014)	1.583 (0.012)	1.647(0.014)	1.568(0.012)	1.668(0.014)	1.695(0.015)
		all	average	2.302 (0.027)	2.397 (0.028)	1.876 (0.018)	1.781 (0.016)	1.846(0.017)	1.756(0.015)	1.890(0.019)	1.925(0.019)
		all	$\lambda_{GC}$	1.301 (0.017)	1.302 (0.018)	1.271 (0.016)	1.285 (0.015)	1.270(0.016)	1.284(0.015)	1.251(0.017)	1.270(0.016)

Table S.1.14 (Continued)

<b>M</b>	<b>Prev</b>	<b>Set</b>	<b>Statistic</b>	<b>ATT Liability</b>	<b>MM Liability</b>	<b>ATT</b>	<b>ATT+PCs</b>	<b>LogR</b>	<b>LogR+PCs</b>	<b>MLM</b>	<b>LTMLM</b>
<b>10000</b>	50%	causal	average	26.781 (1.104)	27.44 (1.129)	16.898 (0.702)	16.815 (0.698)	16.229(0.659)	16.156(0.655)	17.279(0.725)	17.26(0.725)
		null	average	1.114 (0.002)	1.114 (0.002)	1.078 (0.002)	1.083 (0.002)	1.074(0.002)	1.078(0.002)	1.078(0.002)	1.076(0.002)
		all	average	1.14 (0.003)	1.140 (0.003)	1.094 (0.002)	1.099 (0.002)	1.089(0.002)	1.094(0.002)	1.095(0.002)	1.092(0.002)
		all	$\lambda_{GC}$	1.046 (0.005)	1.042 (0.005)	1.039 (0.005)	1.047 (0.005)	1.039(0.005)	1.046(0.005)	1.039(0.005)	1.035(0.005)
	25%	causal	average	27.208 (1.154)	27.938 (1.178)	17.573 (0.726)	17.293 (0.696)	16.856(0.679)	16.61(0.654)	17.976(0.749)	18.077(0.758)
		null	average	1.112 (0.002)	1.11 (0.002)	1.076 (0.002)	1.078 (0.002)	1.071(0.002)	1.073(0.002)	1.074(0.002)	1.073(0.002)
		all	average	1.138 (0.003)	1.137 (0.003)	1.092 (0.002)	1.094 (0.002)	1.087(0.002)	1.089(0.002)	1.091(0.002)	1.090(0.002)
		all	$\lambda_{GC}$	1.048 (0.005)	1.043 (0.005)	1.040 (0.004)	1.045 (0.004)	1.040(0.004)	1.044(0.004)	1.040(0.004)	1.039(0.004)
	10%	causal	average	35.911 (1.500)	36.598 (1.525)	24.379 (1.026)	24.116 (1.014)	23.127(0.944)	22.894(0.934)	24.987(1.071)	25.399(1.091)
		null	average	1.161 (0.003)	1.156 (0.003)	1.112 (0.002)	1.115 (0.002)	1.107(0.002)	1.110(0.002)	1.108(0.002)	1.111(0.002)
		all	average	1.196 (0.003)	1.192 (0.003)	1.136 (0.003)	1.138 (0.003)	1.129(0.002)	1.132(0.002)	1.131(0.003)	1.135(0.003)
		all	$\lambda_{GC}$	1.059 (0.005)	1.050 (0.004)	1.051 (0.005)	1.059 (0.004)	1.050(0.005)	1.058(0.004)	1.044(0.004)	1.047(0.004)

**Table S.1.14. Complete results on real genotypes and simulated phenotypes.** Results include results from Table 4 but we also report ATT and MLM statistics computed using the underlying liability. We report average  $\chi^2$  statistics for simulations with real genotypes and simulated phenotypes averaged across 100 simulations for each parameter setting.

Table S.1.15

<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>ATT+PCs</b>	<b>LogR</b>	<b>LogR +PCs</b>	<b>MLM</b>	<b>LTMLM</b>
<b>1000</b>	50%	0.05	0.689(0.016)	0.717(0.015)	0.614 (0.017)	0.602(0.017)	0.614(0.017)	0.603(0.017)	0.630(0.017)	0.625(0.017)
		0.001	0.504(0.017)	0.533(0.017)	0.409 (0.017)	0.408(0.017)	0.405(0.017)	0.403(0.017)	0.415(0.017)	0.415(0.017)
		$1 \times 10^{-6}$	0.320(0.016)	0.348(0.016)	0.235 (0.015)	0.227(0.014)	0.229(0.014)	0.218(0.014)	0.256(0.015)	0.256(0.015)
		$5 \times 10^{-8}$	0.289(0.015)	0.309(0.016)	0.185 (0.013)	0.178(0.013)	0.179(0.013)	0.172(0.013)	0.205(0.014)	0.200(0.014)
		0.05	0.095(0.001)	0.099(0.001)	0.084 (0.001)	0.083(0.001)	0.084(0.001)	0.083(0.001)	0.085(0.001)	0.085(0.001)
		0.001	0.002(0.001)	0.022(0.001)	0.014 (0.001)	0.013( $5 \times 10^{-4}$ )	0.013(0.001)	0.013( $5 \times 10^{-4}$ )	0.014(0.001)	0.014(0.001)
		$1 \times 10^{-6}$	0.007( $4 \times 10^{-4}$ )	0.009( $4 \times 10^{-4}$ )	0.003 ( $3 \times 10^{-4}$ )	0.003( $2 \times 10^{-4}$ )	0.003( $3 \times 10^{-4}$ )	0.003( $2 \times 10^{-4}$ )	0.004( $3 \times 10^{-4}$ )	0.004( $3 \times 10^{-4}$ )
		$5 \times 10^{-8}$	0.005( $3 \times 10^{-4}$ )	0.006( $3 \times 10^{-4}$ )	0.002 ( $2 \times 10^{-4}$ )	0.002( $2 \times 10^{-4}$ )	0.002( $2 \times 10^{-4}$ )	0.002( $2 \times 10^{-4}$ )	0.003( $2 \times 10^{-4}$ )	0.003( $2 \times 10^{-4}$ )
	25%	0.05	0.737(0.015)	0.750(0.014)	0.682 (0.016)	0.671(0.016)	0.680(0.016)	0.671(0.016)	0.689(0.016)	0.692(0.016)
		0.001	0.579(0.017)	0.603(0.016)	0.491 (0.017)	0.482(0.017)	0.486(0.017)	0.479(0.017)	0.498(0.017)	0.496(0.017)
		$1 \times 10^{-6}$	0.405(0.016)	0.437(0.017)	0.294 (0.016)	0.278(0.015)	0.281(0.015)	0.266(0.015)	0.314(0.016)	0.311(0.016)
		$5 \times 10^{-8}$	0.336(0.016)	0.387(0.016)	0.231 (0.014)	0.216(0.014)	0.218(0.014)	0.208(0.014)	0.248(0.015)	0.251(0.015)
		0.05	0.102(0.001)	0.104(0.001)	0.091 (0.001)	0.09(0.001)	0.091(0.001)	0.090(0.001)	0.091(0.001)	0.092(0.001)
		0.001	0.023(0.001)	0.026(0.001)	0.016 (0.001)	0.015(0.001)	0.016(0.001)	0.015(0.001)	0.017(0.001)	0.018(0.001)
		$1 \times 10^{-6}$	0.009( $4 \times 10^{-4}$ )	0.01( $4 \times 10^{-4}$ )	0.005 ( $3 \times 10^{-4}$ )	0.004( $3 \times 10^{-4}$ )	0.005( $3 \times 10^{-4}$ )	0.004( $3 \times 10^{-4}$ )	0.005( $3 \times 10^{-4}$ )	0.005( $3 \times 10^{-4}$ )
		$5 \times 10^{-8}$	0.006( $3 \times 10^{-4}$ )	0.007( $4 \times 10^{-4}$ )	0.003 ( $3 \times 10^{-4}$ )	0.003( $2 \times 10^{-4}$ )	0.003( $2 \times 10^{-4}$ )	0.003( $2 \times 10^{-4}$ )	0.004( $3 \times 10^{-4}$ )	0.004( $3 \times 10^{-4}$ )

Table S.1.15 (Continued)

<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>ATT+PCs</b>	<b>LogR</b>	<b>LogR +PCs</b>	<b>MLM</b>	<b>LTMLM</b>
	10%	0.05	0.730(0.015)	0.740(0.015)	0.678(0.016)	0.678(0.016)	0.678(0.016)	0.678(0.016)	0.684(0.016)	0.694(0.016)
		0.001	0.575(0.017)	0.593(0.016)	0.498(0.017)	0.492(0.017)	0.494(0.017)	0.490(0.017)	0.510(0.017)	0.517(0.017)
		$1 \times 10^{-6}$	0.410(0.016)	0.435(0.017)	0.332(0.016)	0.328(0.016)	0.324(0.016)	0.316(0.016)	0.340(0.016)	0.343(0.016)
		$5 \times 10^{-8}$	0.374(0.016)	0.388(0.016)	0.271(0.015)	0.265(0.015)	0.261(0.015)	0.251(0.015)	0.276(0.015)	0.281(0.015)
		0.05	0.106(0.001)	0.107(0.001)	0.096(0.001)	0.095(0.001)	0.095(0.001)	0.094(0.001)	0.095(0.001)	0.097(0.001)
		0.001	0.027(0.001)	0.028(0.001)	0.019(0.001)	0.016(0.001)	0.019(0.001)	0.016(0.001)	0.020(0.001)	0.020(0.001)
		$1 \times 10^{-6}$	0.01(4x10 <sup>-4</sup> )	0.011(5x10 <sup>-4</sup> )	0.007(4x10 <sup>-4</sup> )	0.005(3x10 <sup>-4</sup> )	0.006(3x10 <sup>-4</sup> )	0.005(3x10 <sup>-4</sup> )	0.007(4x10 <sup>-4</sup> )	0.007(4x10 <sup>-4</sup> )
		$5 \times 10^{-8}$	0.008(4x10 <sup>-4</sup> )	0.008(4x10 <sup>-4</sup> )	0.005(3x10 <sup>-4</sup> )	0.004(3x10 <sup>-4</sup> )	0.005(3x10 <sup>-4</sup> )	0.003(3x10 <sup>-4</sup> )	0.005(3x10 <sup>-4</sup> )	0.005(3x10 <sup>-4</sup> )

Table S.1.15 (Continued)

<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>ATT+PCs</b>	<b>LogR</b>	<b>LogR +PCs</b>	<b>MLM</b>	<b>LTMLM</b>	
<b>10000</b>	50%	0.05	0.713(0.015)	0.72(0.015)	0.653(0.016)	0.652(0.016)	0.652(0.016)	0.651(0.016)	0.656(0.016)	0.657(0.016)	
		0.001	0.535(0.017)	0.551(0.017)	0.427(0.017)	0.428(0.017)	0.422(0.017)	0.426(0.017)	0.432(0.017)	0.433(0.017)	
		$1 \times 10^{-6}$	0.356(0.016)	0.365(0.016)	0.245(0.015)	0.244(0.015)	0.238(0.015)	0.237(0.015)	0.248(0.015)	0.245(0.015)	
		$5 \times 10^{-8}$	0.306(0.015)	0.311(0.016)	0.187(0.013)	0.191(0.013)	0.181(0.013)	0.178(0.013)	0.190(0.013)	0.189(0.013)	
		0.05	0.058(3x10 <sup>-4</sup> )	0.058(3x10 <sup>-4</sup> )	0.057(3x10 <sup>-4</sup> )	0.057(3x10 <sup>-4</sup> )	0.056(3x10 <sup>-4</sup> )	0.057(3x10 <sup>-4</sup> )	0.057(3x10 <sup>-4</sup> )	0.057(3x10 <sup>-4</sup> )	0.056(3x10 <sup>-4</sup> )
	0.001	0.004(9x10 <sup>-5</sup> )	0.004(9x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(8x10 <sup>-5</sup> )	0.003(8x10 <sup>-5</sup> )	
	$1 \times 10^{-6}$	0.001(4x10 <sup>-5</sup> )	0.001(4x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	
	$5 \times 10^{-8}$	0.001(4x10 <sup>-5</sup> )	0.001(4x10 <sup>-5</sup> )	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	
	25%	0.05	0.725(0.015)	0.726(0.015)	0.635(0.017)	0.633(0.017)	0.635(0.017)	0.632(0.017)	0.635(0.017)	0.635(0.017)	0.635(0.017)
		0.001	0.529(0.017)	0.538(0.017)	0.440(0.017)	0.438(0.017)	0.438(0.017)	0.437(0.017)	0.448(0.017)	0.445(0.017)	
$1 \times 10^{-6}$		0.352(0.016)	0.362(0.016)	0.266(0.015)	0.261(0.015)	0.256(0.015)	0.254(0.015)	0.270(0.015)	0.267(0.015)		
$5 \times 10^{-8}$		0.303(0.015)	0.309(0.016)	0.209(0.014)	0.206(0.014)	0.198(0.014)	0.197(0.014)	0.215(0.014)	0.216(0.014)		
0.05		0.057(3x10 <sup>-4</sup> )	0.057(3x10 <sup>-4</sup> )	0.056(3x10 <sup>-4</sup> )	0.056(3x10 <sup>-4</sup> )	0.055(3x10 <sup>-4</sup> )	0.056(3x10 <sup>-4</sup> )	0.056(3x10 <sup>-4</sup> )	0.056(3x10 <sup>-4</sup> )	0.055(3x10 <sup>-4</sup> )	
0.001		0.004(9x10 <sup>-5</sup> )	0.004(9x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	0.003(7x10 <sup>-5</sup> )	
$1 \times 10^{-6}$		0.001(5x10 <sup>-5</sup> )	0.001(5x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	0.001(3x10 <sup>-5</sup> )	
$5 \times 10^{-8}$	0.001(4x10 <sup>-5</sup> )	0.001(4x10 <sup>-5</sup> )	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)		

Table S.1.15 (Continued)

<b>M</b>	<b>Prev</b>	<b><math>\alpha</math></b>	<b>ATT Liability</b>	<b>MLM Liability</b>	<b>ATT</b>	<b>ATT+PCs</b>	<b>LogR</b>	<b>LogR +PCs</b>	<b>MLM</b>	<b>LTMLM</b>
	10%	0.05	0.752(0.014)	0.756(0.014)	0.711(0.015)	0.711(0.015)	0.709(0.015)	0.706(0.015)	0.713(0.015)	0.715(0.015)
		0.001	0.597(0.016)	0.603(0.016)	0.519(0.017)	0.528(0.017)	0.518(0.017)	0.523(0.017)	0.520(0.017)	0.517(0.017)
		$1 \times 10^{-6}$	0.437(0.017)	0.438(0.017)	0.329(0.016)	0.330(0.016)	0.322(0.016)	0.322(0.016)	0.328(0.016)	0.336(0.016)
		$5 \times 10^{-8}$	0.380(0.016)	0.388(0.016)	0.281(0.015)	0.275(0.015)	0.266(0.015)	0.264(0.015)	0.283(0.015)	0.284(0.015)
		0.05	0.061( $3 \times 10^{-4}$ )	0.06( $3 \times 10^{-4}$ )	0.059( $3 \times 10^{-4}$ )	0.059( $3 \times 10^{-4}$ )	0.058( $3 \times 10^{-4}$ )	0.059( $3 \times 10^{-4}$ )	0.058( $3 \times 10^{-4}$ )	0.058( $3 \times 10^{-4}$ )
		0.001	0.005( $1 \times 10^{-4}$ )	0.005( $1 \times 10^{-4}$ )	0.004( $8 \times 10^{-5}$ )	0.004( $8 \times 10^{-5}$ )	0.003( $8 \times 10^{-5}$ )	0.003( $8 \times 10^{-5}$ )	0.004( $8 \times 10^{-5}$ )	0.004( $8 \times 10^{-5}$ )
		$1 \times 10^{-6}$	0.001( $5 \times 10^{-5}$ )	0.001( $5 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )
		$5 \times 10^{-8}$	0.001( $5 \times 10^{-5}$ )	0.001( $5 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )	0.001( $3 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )	0.001( $3 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )	0.001( $4 \times 10^{-5}$ )

**Table S.1.15. Percentage of SNPs achieving alpha levels for real genotypes and simulated phenotypes.** Results include results for Table 4, but we also report results for ATT and MLM computed using the underlying liability.

Category	Controlling $\lambda_{GC}$	$\alpha$	ATT	ATT+PCs	LogR	LogR+PCs	MLM	LTMLM
Genome Wide	No	0.05	0.093(5x10 <sup>-4</sup> )	0.068(4x10 <sup>-4</sup> )	0.093(5x10 <sup>-4</sup> )	0.067(4x10 <sup>-4</sup> )	0.067(4x10 <sup>-4</sup> )	0.07(4x10 <sup>-4</sup> )
Genome Wide	No	0.001	0.006(1x10 <sup>-4</sup> )	0.003(9x10 <sup>-5</sup> )	0.006(1x10 <sup>-4</sup> )	0.003(9x10 <sup>-5</sup> )	0.003(9x10 <sup>-5</sup> )	0.003(1x10 <sup>-4</sup> )
Genome Wide	No	1x10 <sup>-6</sup>	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
Genome Wide	No	5x10 <sup>-8</sup>	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
Published	No	0.05	0.747(0.050)	0.733(0.051)	0.747(0.050)	0.747(0.050)	0.707(0.053)	0.747(0.05)
Published	No	0.001	0.427(0.057)	0.387(0.056)	0.427(0.057)	0.387(0.056)	0.387(0.056)	0.360(0.055)
Published	No	1x10 <sup>-6</sup>	0.107(0.036)	0.080(0.031)	0.107(0.036)	0.067(0.029)	0.080(0.031)	0.093(0.034)
Published	No	5x10 <sup>-8</sup>	0.053(0.026)	0.040(0.023)	0.053(0.026)	0.027(0.019)	0.027(0.019)	0.053(0.026)
Genome Wide	Yes	0.05	0.053(4x10 <sup>-4</sup> )	0.053(4x10 <sup>-4</sup> )	0.053(4x10 <sup>-4</sup> )	0.053(4x10 <sup>-4</sup> )	0.053(4x10 <sup>-4</sup> )	0.053(4x10 <sup>-4</sup> )
Genome Wide	Yes	0.001	0.002(7x10 <sup>-5</sup> )	0.002(7x10 <sup>-5</sup> )	0.002(7x10 <sup>-5</sup> )	0.002(7x10 <sup>-5</sup> )	0.002(7x10 <sup>-5</sup> )	0.002(7x10 <sup>-5</sup> )
Genome Wide	Yes	1x10 <sup>-6</sup>	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
Genome Wide	Yes	5x10 <sup>-8</sup>	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)	0(NA)
Published	Yes	0.05	0.720(0.052)	0.693(0.053)	0.720(0.052)	0.707(0.053)	0.693(0.053)	0.733(0.051)
Published	Yes	0.001	0.320(0.054)	0.347(0.055)	0.320(0.054)	0.333(0.054)	0.347(0.055)	0.347(0.055)
Published	Yes	1x10 <sup>-6</sup>	0.040(0.023)	0.053(0.026)	0.040(0.023)	0.067(0.029)	0.040(0.023)	0.053(0.026)
Published	Yes	5x10 <sup>-8</sup>	0.013(0.013)	0.013(0.013)	0.013(0.013)	0.013(0.013)	0.013(0.013)	0.013(0.013)

**Table S.1.16. Proportion of SNPs achieving alpha levels for WTCCC2 MS data set.** The number of known associated SNPs that are significant for LTMLM but not MLM (or vice versa) after controlling for  $\lambda_{GC}$  are 3(0) at  $\alpha = 0.05$ , 1(1) at  $\alpha = 0.001$ , 1(0) at  $\alpha = 1x10^{-6}$ , 0(0) at  $\alpha = 5x10^{-8}$ .



SNP Set	ATT	PCA	MLM	LTMLM
<b>Genome Wide AVERAGE</b>	1.38	1.16	1.14	1.17
<b>Genome Wide LD Score INTERCEPT</b>	1.29	1.09	1.08	1.10
<b>Published SNPs AVERAGE</b>	11.64	9.97	9.92	10.59
<b>Published SNPs/Genome Wide Average</b>	8.44	8.61	8.67	9.03
<b>Published SNPs/LD score INTERCEPT</b>	9.06	9.17	9.20	9.66

**Table S.1.17: Results on WTCCC2 MS data set with calibration via LD Score regression.** We report the genome wide  $\chi^2$  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	LTMLM REML
<b>Genome Wide AVERAGE</b>	8188	1.16	1.11	1.14	1.14
<b>Published SNPs AVERAGE</b>		8.94	8.26	8.76	8.82
<b>Published SNPs/Genome Wide Average</b>		7.73	7.45	7.71	7.73
<b>Genome Wide AVERAGE</b>	10034	1.38	1.14	1.17	1.24
<b>Published SNPs AVERAGE</b>		11.64	9.92	10.59	11.04
<b>Published SNPs/Genome Wide Average</b>		8.44	8.67	9.03	8.92
<b>Genome Wide AVERAGE</b>	15633	3.95	1.23	1.08	1.50
<b>Published SNPs AVERAGE</b>		18.54	11.30	5.76	13.76
<b>Published SNPs/Genome Wide Average</b>		4.69	9.20	5.32	9.15

**Table S.1.18: 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). The additional column is for the LTMLM REML statistic calculated using the REML estimate of  $h^2$ . LTMLM using the REML estimate for  $h^2$  produces inflated test statistics and it is not recommended.

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

**Table S.1.19: 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).

## Appendix 2 Supplemental Figures for Chapter 2

M	N	Sibling Set	Prevalence	HE	HE Cut-off	REML	REML Cut-Off	
50000	1000	Unbiased	50%	0.372(0.023)	0.55(0.014)	0.332(0.021)	0.517(0.015)	
			10%	0.317(0.011)	0.493(0.008)	0.276(0.011)	0.509(0.01)	
			1%	0.332(0.008)	0.41(0.007)	0.329(0.008)	0.489(0.014)	
		Concord	50%	1.000(set)	1.000(set)	1.571(0.000)	1.000(set)	
			10%	1.000(set)	1.000(set)	1.052(0.000)	1.000(set)	
			1%	0.961(0.002)	1.000(set)	0.552(0.000)	0.552(0.000)	
		Discord	50%	-2.789(0.002)	-3.039(0.002)	0.000(0.000)	0.000(0.000)	
			10%	-1.863(0.002)	-2.03(0.002)	0.000(0.000)	0.000(0.000)	
			1%	-0.974(0.001)	-1.064(0.001)	0.000(0.000)	0.000(0.000)	
	5000	5000	Unbiased	50%	0.354(0.008)	0.549(0.006)	0.311(0.008)	0.512(0.006)
				10%	0.312(0.005)	0.506(0.003)	0.277(0.005)	0.514(0.004)
				1%	0.343(0.004)	0.458(0.004)	0.365(0.004)	0.493(0.009)
Concord			50%	1.000(set)	1.000(set)	1.007(0.000)	1.000(set)	
			10%	1.000(set)	1.000(set)	0.673(0.000)	0.692(0.000)	
			1%	0.871(0.003)	1.000(set)	0.352(0.000)	0.362(0.000)	
Discord			50%	-2.154(0.002)	-3.028(0.001)	0.000(0.000)	0.000(0.000)	
			10%	-1.439(0.001)	-2.028(0.001)	0.000(0.000)	0.000(0.000)	
			1%	-0.748(0.001)	-1.064(0.000)	0.000(0.000)	0.000(0.000)	

**Table S2.1.** The  $h^2$  estimates for simulated genotypes and phenotypes using varying levels of ascertainment bias, number of individuals, and family ascertainment.

<b>M</b>	<b>N</b>	<b>Input</b> <i><math>h^2_{narrow}</math></i>	<b>Set</b>	<b>ATT</b>	<b>MLM</b>	<b>LT-Fam</b>	<b>LTMLM</b>
10000	1000	0.5	Causal	69.951(2.567)	34.12(1.294)	50.106(1.938)	61.185(2.428)
			Null	1.5(0.002)	0.742(0.001)	0.999(0.001)	1.333(0.002)
			All	1.569(0.004)	0.775(0.002)	1.048(0.003)	1.392(0.004)
			All GC	1.509(0.004)	0.745(0.004)	1.006(0.005)	1.337(0.012)
		0.25	Causal	70.219(2.546)	34.154(1.276)	49.922(1.896)	63.001(2.511)
			Null	1.499(0.002)	0.746(0.001)	0.998(0.001)	1.334(0.002)
			All	1.568(0.004)	0.779(0.002)	1.047(0.003)	1.396(0.004)
			All GC	1.507(0.003)	0.75(0.003)	1.003(0.005)	1.342(0.014)
		0.4	Causal	65.094(2.281)	31.086(1.117)	45.903(1.676)	50.333(1.977)
			Null	1.499(0.002)	0.734(0.001)	0.999(0.001)	1.265(0.002)
			All	1.563(0.004)	0.764(0.002)	1.043(0.003)	1.314(0.003)
			All GC	1.503(0.004)	0.737(0.003)	0.998(0.005)	1.272(0.015)
		0.6	Causal	66.751(2.466)	32.239(1.218)	47.241(1.835)	55.73(2.235)
			Null	1.496(0.002)	0.737(0.001)	0.996(0.001)	1.298(0.002)
			All	1.561(0.004)	0.768(0.002)	1.043(0.003)	1.352(0.003)
			All GC	1.505(0.004)	0.74(0.003)	1.004(0.005)	1.3(0.013)
		0.75	Causal	69.296(2.435)	33.636(1.221)	49.247(1.822)	58.542(2.279)
			Null	1.497(0.002)	0.74(0.001)	0.997(0.001)	1.305(0.002)
			All	1.564(0.004)	0.773(0.002)	1.045(0.003)	1.362(0.003)
			All GC	1.503(0.004)	0.744(0.004)	0.995(0.005)	1.315(0.014)

**Table S2.2** Above is a comparison of concordant simulated sibling at a prevalence of 1% with all different values of the input  $h^2_{narrow}$ , where the true value is 0.50.

GRM cut-off	Controls	Cases	Total	Category	ATT	MLM	LT-Fam
<b>0.05</b>	1778	339	2144	mean $\chi^2$	1.044(0.002)	1(0.002)	1.019(0.002)
				$\lambda_{GC}$	1.056(0.008)	1.011(0.006)	1.029(0.007)
	1318	339	1684	mean $\chi^2$	1.135(0.002)	0.95(0.002)	0.995(0.002)
				$\lambda_{GC}$	1.144(0.01)	0.958(0.005)	1.003(0.009)
	1778	94	1900	mean $\chi^2$	1.236(0.002)	0.973(0.002)	0.994(0.002)
				$\lambda_{GC}$	1.239(0.01)	0.979(0.007)	0.995(0.008)
1318	94	1440	mean $\chi^2$	1.337(0.002)	0.82(0.001)	0.98(0.002)	
			$\lambda_{GC}$	1.336(0.013)	0.822(0.005)	0.98(0.009)	
<b>0.025</b>	1778	339	2144	mean $\chi^2$	1.044(0.002)	1(0.002)	1.019(0.002)
				$\lambda_{GC}$	1.056(0.008)	1.011(0.006)	1.03(0.007)
	968	339	1335	mean $\chi^2$	1.241(0.002)	0.921(0.002)	1.034(0.002)
				$\lambda_{GC}$	1.253(0.018)	0.926(0.005)	1.044(0.015)
	1778	206	2012	mean $\chi^2$	1.087(0.002)	0.998(0.002)	1.016(0.002)
				$\lambda_{GC}$	1.092(0.008)	1.003(0.005)	1.02(0.007)
	968	206	1202	mean $\chi^2$	1.233(0.002)	0.901(0.001)	0.96(0.002)
				$\lambda_{GC}$	1.233(0.013)	0.903(0.005)	0.957(0.01)

**Table S2.3** Genome wide include analysis of all SNPs was performed to test the calibration of the different statistics. Real genotypes and phenotypes from JHS T2D samples with different family ascertainment bias (note there are 28 individuals with unknown status which is why the total isn't just the sum of cases and controls). To mimic family based ascertainment: controls that were related to cases were removed, cases that did not have a case relative were removed, or both.

<b>GRM cut-off</b>	<b>Controls</b>	<b>Cases</b>	<b>Total</b>	<b>HE</b>	<b>HE Cut-Off</b>	<b>REML</b>	<b>REML Cut-Off</b>
<b>0.05</b>	1778	339	2144	0.277	0.257	0.357	0.313
	1318	339	1684	1.000	1.000	1.525	1.000
	1778	94	1900	1.000	1.000	2.798	1.000
	1318	94	1440	1.000	1.000	4.017	1.000
<b>0.025</b>	1778	339	2144	0.277	0.222	0.357	0.921
	968	339	1335	1.000	1.000	1.294	1.000
	1778	206	2012	0.766	0.865	0.910	1.000
	968	206	1202	1.000	1.000	1.726	1.000

**Table S2.4** JHS T2D  $h^2$  estimates on the liability under the same family based ascertainment as described in table 1.

## Appendix 3 Chapter 3 Detailed proof of the 2-Step BMA method and Algorithm Steps

*Step 1:*

The posterior samples of  $(\alpha_x, \gamma, \alpha_y, \beta)$  are obtained by iteratively sampling from  $P(\alpha_x | X, C)$ ,  $P(\gamma | \alpha_x, X, C)$ ,  $P(\alpha_y | D, f(\alpha_x, \gamma))$ , and  $P(\beta | \alpha_y, D, f(\alpha_x, \gamma))$ . Where our observed data in a retrospective model consists of  $D=(X, C, Y)$ .

*J*

$$P(\alpha_x, \gamma | D, \beta, \alpha_y) = P(\alpha_x, \gamma | X, C)$$

A1:  $(Y, \beta, \alpha_y)$  are independent of  $(\alpha_x, \gamma)$  given PS

Let's define  $D^*=(X, C)$  to denote the data in the PS model, now invoking Bayes rule we can set up the posterior predictive distributions of  $\alpha_x$  and  $\gamma$ .

$$P(\alpha_x, \gamma | D^*) = \frac{P(D^* | \alpha_x, \gamma)P(\gamma | \alpha_x)P(\alpha_x)}{P(D^*)} \sim P(D^* | \alpha_x, \gamma)P(\gamma | \alpha_x)P(\alpha_x)$$

The posteriors for  $\alpha_x$  and  $\gamma$  are then iteratively estimated drawn using MCMC. We assume a flat prior on  $P(\alpha_x)$ .

*Step 2:*

The outcome model is then conditional on the posterior predictive distribution of the PS and A1. The posterior predictive distribution of the PS is a deterministic function of the posterior distributions of  $\alpha_x$  and  $\gamma$  ( $PS = f(\alpha_x, \gamma)$ ) as seen in the equation above. For simplicity, we can now assume that all of our distributions are conditional on the observed data, the exposure and set of confounders.

$$P(\alpha_y, \beta, f(\alpha_x \gamma), X, C | Y) = \frac{P(Y | \alpha_y, \beta, f(\alpha_x \gamma), X, C) P(\alpha_x, \gamma | X, C) P(\beta | \alpha_y, X, C) P(\alpha_y | X, C)}{P(Y)}$$

We've factored the priors into it's different terms.

$$\sim P(Y | \alpha_y, \beta, f(\alpha_x \gamma), X, C) P(\alpha_x, \gamma | X, C) P(\beta | \alpha_y, X, C) P(\alpha_y | X, C)$$

Then by sampling over the posterior predictive distribution of the propensity score we integrate over:

$$P(\alpha_y, \beta | Y) \sim \int \sum_{\alpha_x} P(Y | \alpha_y, \beta, f(\alpha_x \gamma), X, C) P(\alpha_x \gamma | X, C) P(\beta | \alpha_y, X, C) P(\alpha_y | X, C) d\gamma$$

We again assume a flat prior on  $P(\alpha_y)$ , implicitly using the assumption and cutting the feedback between fitting the outcome and PS models. We can sample from the posteriors of  $\alpha_y, \beta$  iteratively sampling using MCMC methods to get their posterior predictive distributions. Both the priors on the sets of coefficients  $\gamma$  and  $\beta$  are flat.

### Algorithm Steps

The 2-step BMA approach separately fits the exposure and outcome model, where the outcome model is estimated conditional on the PS estimates calculated in the exposure model. This processes is iteratively computed and then averaged over both the outcome and exposure model space to get the desired treatment estimate.

1. Propose a step for the exposure model:

$$g_x(E[X_i | C_i]) = \sum_{k=0}^p \alpha_x \gamma_k C_{i,k}$$

$\alpha_x^{(t)}$  ->  $\alpha_x'$  by including or excluding a covariate from the propensity score model.  
Criteria for stepping  $\alpha_x^{(t+1)} = \alpha_x'$ , accept with probability:



- a.  $\exp(\text{BIC}(\alpha^{(t)}) - \text{BIC}(\alpha'))$ 
  - i. BIC is effectively a weighted log likelihood that penalizes for more parameters:  $\text{BIC} \sim -2 * \log(\hat{L}) + 2 * p \ln(n)$  (Raftery Sociological Methodology, 1995; Lefebvre stats in med et al., 2014)
2. Update the exposure model covariates by running an MCMC regression and taking the posterior mean of the new coefficients,  $\gamma^{(t+1)}$ . Currently this is done with a probit link and regressing on X, implemented with other links possible though.
  - a. Use this to calculate the  $\text{PS}(\gamma, \mathbf{C} | \alpha_x^{(t+1)})$  to be included as quintile covariates in the outcome model.
3. Propose a step for the outcome model

$$g_y(E[Y_i | X_i, C_i]) = \beta_0 + \beta_X X_i + h(\text{PS}(\gamma, \alpha_x, C_i); \beta) + \sum_{k=0}^p \alpha_y \beta_k C_{i,k}$$

$\alpha_y^{(t)} \rightarrow \alpha_y'$  by including or excluding a covariate from model. Criteria for stepping  $\alpha_x^{(t+1)} = \alpha_x'$ , accept with probability:

- a.  $\exp(\text{BIC}(\alpha^{(t)}) - \text{BIC}(\alpha'))$
4. Update the outcome model,
 
$$g_y(E[Y_i | X_i, C_i]) = \beta_0 + \beta_X X_i + h(\text{PS}(\gamma, \alpha, C_i); \beta) + \sum_{k=0}^p \alpha \beta_k C_{i,k}$$
 covariates by running an MCMC regression and taking the posterior mean of the new coefficients,  $(\beta, \delta)^{(t+1)}$ . (Currently this is done logit link and regressing on Y, implemented with other links possible though.)
5. Calculate the ACE

$$\widehat{\text{ACE}} = \sum_{t=1}^{\text{MCMC iter}} \left\{ P(\alpha | \text{Data}) * \frac{1}{n} (\hat{Y}^t | X = 1, C = c) - (\hat{Y}^t | X = 0, C = c) \right\}$$

# Bibliography

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., Vilhjalmsson, 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., Vilhjalmsson, 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. Chen, G.-B. (2014). Estimating heritability of complex traits from genome-wide association studies using IBS-based Haseman-Elston regression. *Frontiers in Genetics* 5.
20. 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.
21. 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.
22. 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.
23. Heyde, C.C. (1997). *Quasi-likelihood and its application: a general approach to optimal parameter estimation.* (Springer).
24. 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.
25. Devlin, B., and Roeder, K. (1999). Genomic control for association studies. *Biometrics* 55, 997-1004.
26. Lee, S., Harold, D., and Nyholt, D. (2013). Genetic and Environmental Risk for Alzheimer's Disease Consortium. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet* 22, 832-841.
27. 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.
28. 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.
29. Golan, D., Lander, E.S., and Rosset, S. (2014). Measuring missing heritability: Inferring the contribution of common variants. *Proceedings of the National Academy of Sciences* 111, E5272-E5281.
30. 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.
31. 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. *Nat Genet*, in press <http://biorxiv.org/content/early/2014/02/21/002931short>.
32. Visscher, P.M., Hill, W.G., and Wray, N.R. (2008). Heritability in the genomics era--concepts and misconceptions. *Nat Rev Genet* 9, 255-266.
33. Plomin, R., Haworth, C.M.A., and Davis, O.S.P. (2009). Common disorders are quantitative traits. *Nat Rev Genet* 10, 872-878.

34. 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. *Nat Genet*, in press  
<http://biorxiv.org/biorxiv/early/2014/08/09/007799fullpdf>.
35. Zhou, X., and Stephens, M. (2014). Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods* 11, 407-409.
36. Eu-ahsunthornwattana, J., Miller, E.N., Fakiola, M., Jeronimo, S.M., Blackwell, J.M., Cordell, H.J., and W.T.C.C.C. (2014). Comparison of methods to account for relatedness in genome-wide association studies with family-based data.
37. Hayeck, T.J., Zaitlen, N.A., Loh, P.-R., Vilhjalmsdottir, B., Pollack, S., Gusev, A., Yang, J., Chen, G.-B., Goddard, M.E., and Visscher, P.M. (2015). Mixed Model with Correction for Case-Control Ascertainment Increases Association Power. *The American Journal of Human Genetics* 96, 720-730.
38. Zaitlen, N., Kraft, P., Patterson, N., Pasaniuc, B., Bhatia, G., Pollack, S., and Price, A.L. (2013). Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits.
39. Thornton, T., and McPeck, M.S. (2010). ROADTRIPS: case-control association testing with partially or completely unknown population and pedigree structure. *Am J Hum Genet* 86, 172-184.
40. Pasaniuc, B., Zaitlen, N., Lettre, G., Chen, G.K., Tandon, A., Kao, W., Ruczinski, I., Fornage, M., Siscovick, D.S., and Zhu, X. (2011). Enhanced statistical tests for GWAS in admixed populations: assessment using African Americans from CARE and a Breast Cancer Consortium. *PLoS Genet* 7, e1001371.
41. Schneeweiss, S., Rassen, J.A., Glynn, R.J., Avorn, J., Mogun, H., and Brookhart, M.A. (2009). High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology (Cambridge, Mass)* 20, 512.
42. Zigler, C.M., Watts, K., Yeh, R.W., Wang, Y., Coull, B.A., and Dominici, F. (2013). Model feedback in bayesian propensity score estimation. *Biometrics* 69, 263-273.
43. Wang, C., Parmigiani, G., and Dominici, F. (2012). Bayesian effect estimation accounting for adjustment uncertainty. *Biometrics* 68, 661-671.
44. Raftery, A.E., Madigan, D., and Hoeting, J.A. (1997). Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92, 179-191.
45. Zigler, C.M., and Dominici, F. (2014). Uncertainty in propensity score estimation: Bayesian methods for variable selection and model-averaged causal effects. *Journal of the American Statistical Association* 109, 95-107.
46. Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999). Bayesian model averaging: a tutorial. *Statistical science*, 382-401.
47. McCaffrey, D.F., Ridgeway, G., and Morral, A.R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods* 9, 403.
48. Mayers, J.R., Wu, C., Clish, C.B., Kraft, P., Torrence, M.E., Fiske, B.P., Yuan, C., Bao, Y., Townsend, M.K., and Tworoger, S.S. (2014). Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nature medicine*.
49. Liu, Z., Zheng, Yan, Qi, Qibin, Hu, Frank, Kraft, Pete, Qi, Lu (2015). Plasma Metabolites and Body Mass Index in U.S. Caucasian Men and Women In Progress.
50. Rosenbaum, P.R., and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41-55.
51. Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics* 25, 1.

52. Lefebvre, G., Atherton, J., and Talbot, D. (2014). The effect of the prior distribution in the Bayesian Adjustment for Confounding algorithm. *Computational Statistics & Data Analysis* 70, 227-240.
53. Claudino, W.M., Goncalves, P.H., di Leo, A., Philip, P.A., and Sarkar, F.H. (2012). Metabolomics in cancer: a bench-to-bedside intersection. *Critical reviews in oncology/hematology* 84, 1-7.
54. Batch, B.C., Shah, S.H., Newgard, C.B., Turer, C.B., Haynes, C., Bain, J.R., Muehlbauer, M., Patel, M.J., Stevens, R.D., and Appel, L.J. (2013). Branched chain amino acids are novel biomarkers for discrimination of metabolic wellness. *Metabolism* 62, 961-969.
55. Lackey, D.E., Lynch, C.J., Olson, K.C., Mostaedi, R., Ali, M., Smith, W.H., Karpe, F., Humphreys, S., Bedinger, D.H., and Dunn, T.N. (2013). Regulation of adipose branched-chain amino acid catabolism enzyme expression and cross-adipose amino acid flux in human obesity. *American Journal of Physiology-Endocrinology and Metabolism* 304, E1175-E1187.
56. Newgard, C.B., An, J., Bain, J.R., Muehlbauer, M.J., Stevens, R.D., Lien, L.F., Haqq, A.M., Shah, S.H., Arlotto, M., and Slentz, C.A. (2009). A branched-chain amino acid-related metabolic signature that differentiates obese and lean humans and contributes to insulin resistance. *Cell metabolism* 9, 311-326.
57. Hermanussen, M., Garcia, A., Sunder, M., Voigt, M., Salazar, V., and Tresguerres, J. (2006). Obesity, voracity, and short stature: the impact of glutamate on the regulation of appetite. *European journal of clinical nutrition* 60, 25-31.
58. Holmseth, S., Lehre, K., and Danbolt, N. (2006). Specificity controls for immunocytochemistry. *Anatomy and embryology* 211, 257-266.