Published in final edited form as:

Genet Epidemiol. 2017 July ; 41(5): 413–426. doi:10.1002/gepi.22045.

A Novel Association Test for Multiple Secondary Phenotypes from a Case-Control GWAS

Debashree Ray and

Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA

Saonli Basu

Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA

Summary

In the past decade, many genome wide association studies (GWASs) have been conducted to explore association of single nucleotide polymorphisms (SNPs) with complex diseases using a case-control design. These GWASs not only collect information on the disease status (primary phenotype, D) and the SNPs (genotypes, X), but also collect extensive data on several risk factors and traits. Recent literature and grant proposals point towards a trend in re-using existing large case-control data for exploring genetic associations of some additional traits (secondary phenotypes, Y) collected during the study. These secondary phenotypes may be correlated, and a proper analysis warrants a multivariate approach. Commonly used multivariate methods are not equipped to properly account for the non-random sampling scheme. Current *ad hoc* practices include analyses without any adjustment, and analyses with D adjusted as a covariate. Our theoretical and empirical studies suggest that the type I error for testing genetic association of secondary traits can be substantial when X as well as Y are associated with D , even when there is no association between X and Y in the underlying (target) population. Whether using D as a covariate helps maintain type I error or not depends heavily on the disease mechanism and the underlying causal structure (which is often unknown). To avoid grossly incorrect inference, we have proposed proportional odds model adjusted for propensity score (POM-PS). It uses a proportional odds model for X on Y and adjusts estimated conditional probability of being diseased as a covariate. We demonstrate the validity and advantage of POM-PS, and compare to some existing methods in extensive simulation experiments mimicking plausible scenarios of dependency among Y , X , and D . Finally, we use POM-PS to jointly analyze four adiposity traits using a type 2 diabetes (T2D) case-control sample from the population-based Metabolic Syndrome in Men (METSIM) Study. Only POM-PS analysis of the T2D case-control sample seems to provide valid association signals.

Supplemental Data

Supplemental Data include additional figures, tables and proofs of results, and can be found with this article at the Genetic Epidemiology website on Wiley Online Library.

Web Resources

R software for POM-PS can be found at <https://github.com/RayDebashree/POM-PS>.

Keywords

Case-control design; Cross-phenotype association; GWAS; Joint modeling; METSIM; Multiple traits; Multivariate analysis; Propensity Score; Proportional Odds Model; Secondary traits; Stratification Score

1. Introduction

In an effort to better understand the genetic variants associated with complex diseases, several GWASs have been conducted, where extensive data on genetic variants and several quantitative phenotypes are often collected. In a case-control GWAS, the primary goal is to analyze association of the complex disease-status (primary phenotype, D) with a single SNP (genotype, X) or multiple SNPs such as a gene (e.g., Pan, 2009; Lee et al., 2012; Ray et al., 2015). Leveraging an existing case-control sample for a secondary study of some additional traits collected during the study can save considerable time and money, especially for a sequence-based GWAS. These traits analyzed in a secondary association study are known as secondary traits/phenotypes (Y). For example, Teslovich et al. (2010) meta-analyzed lipid traits in $> 100,000$ individuals from 46 GWASs, including samples from 9 case-control studies of diseases such as T2D, hypertension, and coronary artery disease. However, any inference using secondary phenotypes warrants attention towards the ascertainment bias that may arise from the over-representation of cases in the case-control (non-random) sample (Ma and Carroll, 2016). Further, many secondary phenotypes are often correlated and a joint analysis may improve power over independent analysis of each trait (O'Reilly et al., 2012; Basu et al., 2013; Ray et al., 2016, and references therein). In this paper, we aim to study the performances of some current approaches and propose a new approach for testing association of one or more secondary traits from a case-control GWAS.

The tests of association of multiple secondary traits Y with a single SNP X from a case-control GWAS can be categorized into two groups. The first group of tests consists of joint analysis methods that assume random ascertainment of samples. Some examples are Multivariate Analysis of Variance (MANOVA), the Sum of Squared Score (SSU) test (Pan, 2009; Yang and Wang, 2012), MultiPhen (O'Reilly et al., 2012) and GEMMA (Zhou and Stephens, 2014). Ray et al. (2016) proposed a data-adaptive test USAT and compared statistical powers of these methods on a random sample of unrelated individuals. A convenient way of using these methods for a case-control design is to adjust the case-control status D as a covariate. One may also naively ignore the sampling design and use an unadjusted approach. Following Jiang et al. (2006), we refer to these D -adjusted and unadjusted approaches as *ad hoc* methods. On the other hand, the second group of tests consists of methods that specifically account for the case-control (non-random) ascertainment of samples.

Much of the methodological research has focused on the unbiased estimation of the genetic effect of a single secondary phenotype Y , especially a binary one (e.g., Wang and Shete, 2011). Richardson et al. (2007) proposed a case-control-stratum-weighted logistic regression for estimating the association between X and a binary Y . Other weighted estimating equation approaches were recently proposed by Song et al. (2016); Xing et al. (2016).

Monsees et al. (2009) investigated the consequences of univariate *ad hoc* approaches. Lin and Zeng (2009) developed SPREG, a regression method based on retrospective likelihood function to properly account for the case-control sampling. Lutz et al. (2014) argued that SPREG can encounter numerical issues in practice, and proposed a computationally faster hypothesis testing approach using proportional odds likelihood. Further generalizations of Lin and Zeng's likelihood approach were proposed by Ghosh et al. (2013); Wei et al. (2013). He et al. (2012) used a Gaussian copula approach to jointly model Y and D , while Tchetgen (2014) considered a careful re-parameterization of the conditional model for Y given D .

In the current literature, little attention has been paid on testing genetic association of multiple secondary phenotypes. Most of the above methods cannot readily accommodate multiple secondary traits Y . Lutz et al. (2014)'s method, when extended to multiple phenotypes, is equivalent to MultiPhen with D adjusted as a covariate. He et al. (2012) claimed that their Gaussian copula method can accommodate multiple secondary traits. Liu and Leal (2012) used a prospective likelihood approach to jointly model multiple traits conditional on the ascertainment. Schifano et al. (2013) proposed SMAT, which is described in section 2.

In this article, we consider 6 possible directed acyclic graphs (DAGs) (Monsees et al., 2009) to depict directional dependencies between Y , X and D in the population of interest (Figure 1). We use simple probability rules to theoretically infer about independence of X and Y in the sample under the null for each DAG. We simulate datasets under all 6 DAGs. Recent articles (e.g., He et al., 2012; Schifano et al., 2013), too, have not explored the breadth of simulation settings that we have considered in this paper. Additionally, we found DAG 'F' (a DAG not explored by other articles) can lead to several spurious signals. In particular, SPREG (for a single trait) and SMAT (for multiple traits) fail to maintain desired error level. Our real data analysis further emphasizes this major shortcoming of current practices.

Based on these findings, we propose an alternative hypothesis testing approach, called POM-PS (proportional odds model adjusted for propensity score). Unlike Lutz et al. (2014), which adjusts D as a covariate, we consider covariate adjustment of the estimated propensity score in the proportional odds model for X on Y . We have shown that POM-PS alleviates the confounding effect of D on Y , and helps maintain correct type I error. Unlike existing approaches, POM-PS has proper type I error irrespective of the underlying causal structure, and has comparable statistical power to detect non-zero genetic effects. It has a known asymptotic distribution under the null, can be used to test a single or multiple secondary traits from a case-control GWAS and does not require knowledge of disease prevalence.

2. Material & Methods

2.1 Notation

Consider a case-control GWAS on n unrelated individuals, sampled based on the disease status D (primary outcome). Sample individuals with $D = 1$ are cases (n_1 of them) while the rest ($n_0 = n - n_1$) are controls. Let S denote the sampling indicator. For simplicity, we assume that the sampling strategy depends on the disease status only. For a specific unit (individual) in the population, $S = 1$ if it has been sampled. Consequently, all the individuals

not included in the sample have $S = 0$. Each sampled individual has observations on K correlated traits, and $p (\gg n)$ SNPs. Let $\mathbf{Y}_k$ and $\mathbf{D}$ be the $n \times 1$ vectors of k -th secondary trait and the primary outcome respectively, and $\mathbf{Y}$ be the $n \times K$ matrix of secondary traits for all sampled individuals. We are interested in testing the null hypothesis that for a given SNP, all the K secondary traits have zero genetic effect. We assume an additive genetic model where $\mathbf{X}$ is the $n \times 1$ vector of minor allele counts (0, 1 or 2).

2.2 Directed Acyclic Graphs depicting association scenarios

We consider six possible scenarios for the joint distribution of $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{D}$ in the underlying population (Figure 1). When a case-control sampling mechanism is used to draw a sample with observations on multiple phenotypes and genotypes, the underlying (target) population from which the sample is drawn may not be equivalent to the sampling population. The sampling population is defined as “a (potentially hypothetical) population from which the observed sample may be viewed as being representative” (Haneuse et al., 2009). The observed sample can be considered as a random sample from this sampling population. Our null hypothesis is that the SNP is not associated with any of the secondary traits in the underlying population. However, inference based on the observed case-control sample pertain to the sampling population and may not be generalizable to the underlying population.

Mathematical results in the Appendix (proofs in Supplementary S1) theoretically justify conditions and scenarios where the *ad hoc* methods will or will not provide valid tests of association. By ‘valid’, we mean a test that has controlled type I error at a given significance level. In summary, over-representation of cases do not affect type I error estimates of *ad hoc* methods (unadjusted and $\mathbf{D}$ -adjusted approaches) for DAGs ‘A’, ‘B’, ‘C’ and ‘D’ (Result 1). For DAG ‘E’, $\mathbf{D}$ -adjusted methods show inflated type I error (Result 2), while for DAG ‘F’, unadjusted methods lead to false association signals (Result 3). Few real data examples for DAGs ‘E’ and ‘F’ are provided in Supplementary S3.

2.3 Existing tests for multiple phenotype analysis

2.3.1 MANOVA—The multiple multivariate linear regression (MMLR) model for the association of K correlated traits and the SNP is given by

$$\mathbf{Y}_{n \times K} = \mathbf{X}_{n \times 1} \boldsymbol{\beta}'_{1 \times K} + \mathbf{Z}_{n \times q} \boldsymbol{\Phi}_{q \times K} + \boldsymbol{\mathcal{E}}_{n \times K} \quad (1)$$

where $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_K)$ is the vector of fixed unknown genetic effects corresponding to the K correlated traits for a specific SNP, $\mathbf{Z}$ is the matrix of q covariates (including intercept), $\boldsymbol{\Phi}$ is the matrix of covariate effects (nuisance parameters), and $\boldsymbol{\mathcal{E}}$ is an error matrix from $N_K(\mathbf{0}, \boldsymbol{\Sigma})$. This corresponds to the unadjusted method. For $\mathbf{D}$ -adjusted method, the covariate matrix $\mathbf{Z}$ will additionally include the vector of primary outcomes $\mathbf{D}$. For testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$ (no secondary trait is associated with the SNP), the likelihood ratio test (LRT) is equivalent to the MANOVA test statistic, which has a χ_K^2 distribution under H_0 .

2.3.2 SSU—Apart from multivariate models, one may use marginal models

$$\mathbf{Y}_k = \beta_{M,k} \mathbf{X} + \mathbf{Z} \Phi_k + \varepsilon_k, \varepsilon_k \sim N_n(0, \sigma^2 \mathbf{I}_n) \quad (2)$$

where $\beta_{M,k}$ is the parameter associated with the SNP effect on the k -th trait, $\mathbf{Z}$ is the matrix of q covariates, Φ_k is the $q \times 1$ vector of covariate effects. The null hypothesis for the k -th model is that the k -th trait is not associated with the SNP. Under H_0 (the global null hypothesis of no association), the SSU test $T_S = \mathbf{U}_M' \mathbf{U}_M \stackrel{approx}{\sim} a\chi_d^2 + b$, where $\mathbf{U}_M$ is the score vector obtained from (2), and the parameters a, b, d are estimated from $\text{Cov}(\mathbf{U}_M)$.

2.4 Existing test for multiple secondary phenotype analysis

2.4.1 SMAT—When multiple phenotypes are positively correlated and measure the same underlying trait in the same direction, Schifano et al. (2013) proposed the use of stratum weighted estimating equations in order to estimate and test the shared common effect of SNPs. These weights account for the case-control ascertainment and require the knowledge of population prevalence of the primary disease. SMAT tests for the common effect of SNP using a 1 degree of freedom test while allowing for phenotype-specific covariate effects.

2.5 Extension of existing tests for single secondary phenotype analysis

2.5.1 Proportional Odds Model (POM)—In the context of testing genetic association of a single secondary trait Y , Lutz et al. (2014) proposed modeling of $P[X/Y, D]$ and $P[X/D]$ (under H_0) using proportional odds model. They suggested using the LRT statistic

$$-2\log \left(\frac{\prod_{i=1}^n P[X_i|D_i]}{\prod_{i=1}^n P[X_i|Y_i, D_i]} \right) \sim \chi_1^2. \text{ This can be readily extended to incorporate } K \text{ secondary traits where the LRT statistic is distributed as } \chi_K^2 \text{ under } H_0.$$

2.6 An alternative association test using propensity score

Among the *ad hoc* methods, there is no uniformly valid hypothesis testing approach for secondary traits (refer Appendix). Inflated type I errors in DAGs ‘E’ and ‘F’ result from the confounding effect of the primary outcome (D) on the secondary phenotypes (Y). Confounding variables are commonly encountered in observational studies, and propensity scores (proposed by Rosenbaum and Rubin, 1983) are widely used to reduce effects of confounding variables. In the context of prospective observational studies, the propensity score is defined as the conditional probability of treatment assignment given some observed covariates. For a retrospective study like ours, Epstein et al. (2007); Allen and Satten (2011) defined the conditional probability of case status as the ‘stratification score’ since they used this score to assign cases and controls to different strata. We, however, will not use this conditional probability to do stratification or matching of any kind, and hence prefer to use the term ‘propensity score’ instead of ‘stratification score’ to avoid confusion.

For our primary case-control study, any or all of the measured baseline covariates (including the secondary phenotypes Y) can potentially be associated with the case-control status D , which can confound the association between Y and X . Without loss of generality, we assume

Y are the only measured covariates in the case-control study. The propensity score $PS = P(D = 1|Y, S = 1)$ is then modeled using the logistic model: $\text{logit}(PS) = \gamma_0 + \gamma'Y$. This PS need not correspond to any population quantity as long as the relationship between D and Y is modeled correctly (Allen and Satten, 2011). By modeling the probability of case-control status (D) as a function of observed covariates (Y), it can be shown that Y is independent of D given PS (Supplementary S4). If there are other measured covariates C influencing D , then PS is modeled as $PS = P(D = 1|Y, C, S = 1)$. The question that now arises is how to use this propensity score to remove the confounding effect of case-control status in the test for genetic association of secondary traits.

In the current literature for prospective studies, there are four propensity score methods, which are traditionally used to estimate average treatment effect by comparing treated and untreated groups in a non-randomized study. First, propensity score matching involves creating matched sets of treated and untreated subjects that share a similar value of the propensity score (Rosenbaum and Rubin, 1983, 1985). There is no universally accepted choice for making such matched sets, and the observations within the matched sample are no longer independent (Austin, 2011). Second, propensity score stratification involves using estimated scores to rank subjects, who are then grouped into subsets based on thresholds defined apriori. The choice of threshold is non-trivial (Austin, 2011). Third, inverse probability weighting (IPW) using propensity score is a form of model-based direct standardization (Rosenbaum, 1987), where each subject's weight is equal to the inverse of estimated propensity score. The fourth approach involves adjusting propensity score as a covariate in the regression model.

One can potentially use any of these propensity score approaches in the context of a retrospective study such as ours. For the matching and the stratification methods, after one has obtained a matched sample or a stratified sample, one still needs to devise ways (a non-trivial task) to use such a sample for jointly testing genetic association of multiple correlated traits on a genome-wide scale (Schifano et al., 2015). For the IPW method, a propensity score based weight can potentially be used in the frameworks described by Richardson et al. (2007) (for a single trait) and Schifano et al. (2013) (for multiple traits). However, our preliminary studies show inflated type I error for this approach in some scenarios (Supplementary S5).

POM-PS: A proportional odds model adjusted for propensity-score—For testing association of a SNP with secondary traits from a case-control study, we propose to use covariate adjustment of propensity score. Specifically, we use inverse-regression framework (or POM) to regress the SNP on the secondary phenotypes Y and the estimated propensity score (which depends on Y and other baseline covariates). POM does not require normality of traits and can flexibly incorporate both binary and continuous secondary phenotypes. We can now use LRT for testing association of X and Y . The covariate adjustment approach does not induce dependence among the observations (unlike matching or stratification approaches) and hence the LRT has an asymptotic χ^2_K distribution. We can conveniently use known R functions (e.g., `polr()` in MASS package) to implement this simple straightforward method. Unlike SMAT software, POM-PS can also test genetic association

of a single secondary phenotype, and does not require knowledge of population prevalence of the primary disease.

3. Results

3.1 Simulation Setup

To simulate data under DAGs 'A', 'B', 'C', 'E', we followed the model (S5) (Supplementary S6). Similarly, for DAGs 'D' and 'F', we used the model (S1) (Supplementary S6). In other words, we first simulated X taking values 0, 1, 2 with probabilities $(1-f)^2, 2f(1-f), f^2$ respectively. $f=0.2$ is the minor allele frequency (MAF) of the SNP. Depending on the DAG, either Y or D is simulated next. Conditional on X , we simulated Y for a fixed K using the simulation model $Y = \beta_0^* \mathbf{1} + \beta^* X + \Phi^* D + \epsilon$, where the vectors $Y, \mathbf{1}, \beta^*, \Phi^*, \epsilon$ are K -dimensional. We took $\beta_0^* = 0, \Phi^* = \phi^* \mathbf{1}$ and simulated errors ϵ from $N_K(0, \sigma^2 R(\rho))$, where $R(\rho)$ is a compound symmetry (CS) correlation matrix. The genetic effect size for an associated trait was chosen such that the SNP explained $h\%$ ($0 < h < 1$) of the total phenotypic variance. Specific choices of $h, \beta^*, \phi^*, \sigma^2$ and ρ for each simulation are given in their respective sections. The disease status is simulated as:

$\text{logit}(P(D=1)) = \alpha_0^* + (\alpha^*)' Y + \alpha_x^* X$. Parameter α_0^* is chosen such that the baseline disease prevalence in the underlying population is $\pi = 0.1$ (common disease). We took $\alpha^* = \alpha^* \mathbf{1}$. Within each DAG, $N \geq 21,000$ individuals representing the population are simulated, out of which N_0 (N_1) are controls (cases). $n_0 (\ll N_0)$ controls and $n_1 (\ll N_1)$ cases are randomly sampled from the population to form our case-control sample. For each scenario, N_R independent replications are considered. The different parameter values in each simulation scenario are summarized in Table 1. We are interested in testing $H_0 : \beta^* = \mathbf{0}$. All the analyses were implemented in R 3.0.1. For SMAT, the R package SMAT 0.98 is used with exchangeable working correlation and two choices of the estimated disease prevalence: $\pi_{SMAT} = 0.1$ and 0.2.

3.2 Simulation 1: K = 2 traits

3.2.1 Type I error comparison & bias evaluation—Here both traits have zero genetic effects in the population ($\beta_1^* = 0 = \beta_2^*$). We simulate $N_R = 10,000$ null datasets with $n_0 = n_1 = 2,000$ unrelated individuals. At level α , we estimate type I error as the proportion of null datasets with p-value $\leq \alpha$. Level $\alpha = 0.01$ is sufficient to demonstrate the relative type I error performance of these methods since the performance of the basic model framework of each method at a genome-wide level can be found in existing literature (Supplementary S6).

Table 2 shows the type I error estimates for each of the methods under the 6 possible DAGs for 4 values of trait correlation ρ : -0.8, -0.2, 0.2, 0.8. All methods, except SMAT and POM-PS, have been analyzed twice: once without considering any adjustment for the non-random sampling scheme, and once with adjustment for D as a covariate. As expected from Result 1, all methods (irrespective of any adjustment for ascertainment) maintain proper type I error under DAGs 'A', 'B', 'C' and 'D'. For DAG 'E', the methods that treat the ascertained sample as if it were a random sample from the underlying population do not exhibit inflated type I error while analyses with covariate adjustment for D suffer from inflated type I errors. SMAT seems to have a slightly inflated error level. Exactly opposite behavior is observed

under DAG ‘F’. As indicated by Result 3, the methods adjusted for D maintain proper error level across all values of correlation while the unadjusted methods have severely inflated type I errors. SMAT, too, has noticeably high type I error estimates even when its assumption about common genetic effects and positively correlated traits are not violated. POM-PS, however, exhibits correct type I error estimates under all scenarios.

The relative behavior of these approaches remain unaffected for a disease with low baseline prevalence ($\pi = 0.03$) in the underlying population (Supplementary S8). In the presence of unmeasured covariates affecting the disease status, even though the propensity score is not modeled correctly, POM-PS continues to have controlled type I error (Supplementary S9).

To show that POM-PS maintains proper type I error when applied to a single secondary phenotype, we performed type I error analysis using only $K = 1$ trait. In terms of validity of inference, our results (Supplementary S7) indicate that POM-PS is also preferable to existing methods for a single secondary trait such as Lin and Zeng (2009); Lutz et al. (2014).

To study the possible source(s) of inflated type I errors in the existing methods, we calculated the biases and the distortion in the variances of estimated genetic effects (Supplementary S10). As theoretically expected from Result 1, DAGs ‘A’–‘D’ do not reflect biases or distorted variances for any method. For DAGs ‘E’ and ‘F’, we do not observe any distortion in the variances of the estimated genetic effects. However, the biases in estimated genetic effects are very high in situations where inflated type I errors are observed.

3.3 Simulation 2: $K = 5$, 20 traits

3.3.1 Type I error performance of POM-PS—To explore if POM-PS continues to have controlled type I error rates for larger number of traits, we simulate data for $K = 5$ or 20 traits with pairwise correlations $\rho = 0.2, 0.4, 0.6$. We consider 5,000 null datasets with 2,000 cases and 2,000 controls, and an error threshold $\alpha = 0.01$ (Supplementary S6). The type I error estimates, reported in Table 3, show that POM-PS provides a valid test of the null hypothesis of no association irrespective of the causal mechanism among Y , X and D .

3.3.2 Power comparison of POM-PS with existing methods—We simulate 500 datasets for each scenario with 200 cases and 200 controls (see Supplementary S6). The pairwise correlation ρ is varied to have values 0.2, 0.4 or 0.6. We chose $h = 0.005$, which gave effect size 0.395 for an associated trait. Two scenarios of association are considered: 20% (low) or 60% (moderate number) of traits are associated with the SNP. Since the existing methods do not always have correct type I error, empirical powers based on corrected critical values are compared (see Supplementary S6). We have presented results for DAGs ‘E’ and ‘F’ only. Since LRT based on POM (or MultiPhen) and MANOVA have similar behavior when testing association of a single SNP with multiple traits, only MANOVA is presented.

For DAG ‘E’, powers of unadjusted and D -adjusted MANOVA tend to be similar for larger K and stronger ρ (Figure 2). In most situations, MANOVA (both the unadjusted and the D -adjusted versions) and POM-PS dominate over other methods. Like DAG ‘E’, in most situations under DAG ‘F’, MANOVA (both versions) and POM-PS dominate over other

methods (Figure 3). One must keep in mind that the dependency among Y , X and D may not be known apriori and using any existing test may not provide valid test for H_0 . In this regard, POM-PS seems to be the only approach that provides valid test, and is powerful in detecting cross-phenotype association.

4. Real Data Analysis

4.1 METSIM Study: Description

The METSIM study is a large, single-site, longitudinal study designed to investigate cardiovascular diseases, T2D and insulin resistance in Finnish men. It includes 10,197 men (aged 45–73 years) randomly selected from Kuopio, Finland. Participants were genotyped with the Illumina OmniExpress GWAS chip and the Illumina exome chip. Details on laboratory measurements and genotyping can be found in Stan áková et al. (2009), Stan áková et al. (2009).

We examine four adiposity traits from the first visit: body mass index (BMI), waist circumference (WC), hip circumference (HC) and waist-to-hip ratio (WHR). The pairwise trait correlations are within (0.4, 0.9). We excluded individuals with any missing adiposity traits, leaving 9,964 unrelated participants. However, this sample is a random sample from the population. To select a case-control sample from this cohort, we consider the T2D status. Adiposity related traits may be associated with T2D and may share common SNPs (Thorleifsson et al., 2009). We create a T2D case-control sample using all 1,381 individuals with diagnosed T2D (cases) and 1,381 randomly selected controls. It is of interest to conduct a GWAS of the four adiposity traits based on this T2D case-control sample.

The purpose of this real data analysis is not to detect novel signals from this artificially constructed case-control sample. We aim to show how multivariate analysis of quantitative phenotypes from a case-control sample using some existing methods may yield spurious association signals. We also analyzed the original METSIM cohort (a random sample from the underlying population), the results of which can act as a benchmark for the case-control sample results in the sense that it is not expected to discover true signals from the smaller sample that are not found in the cohort.

All traits are adjusted for age and the residuals are inverse normalized. Adiposity traits are usually adjusted for sex as well; however all participants are males. We analyzed the resulting marginal normal traits and tested for association with genotypes while adjusting the first 3 principal components from the genetic data to account for possible population stratification and relatedness among the individuals. We analyzed 604,495 SNPs with MAF $\geq 1\%$, and consider a significance threshold of 5×10^{-7} (see Supplementary S14).

4.2 METSIM Study: Analysis Results

First, we conduct a GWAS using our case-control sample of size 2,762. All four inverse-normalized adiposity trait residuals are jointly analyzed using MANOVA (both with and without adjustment for T2D), SMAT (with prevalence estimate 0.2) and our method POM-PS. Figure 4 shows the Manhattan plots for these approaches. To provide a benchmark for comparison, Figure 5 shows the Manhattan plots for the unadjusted MANOVA and the

unweighted SMAT analyses of the entire METSIM cohort of size 9,964. At level 5×10^{-7} , unweighted SMAT detected six significant SNPs, and unadjusted MANOVA additionally found five significant SNPs from the cohort (Table S8).

Although the QQ-plots (Figure S2) and the values of genomic control inflation factor suggest little evidence of systematic bias of the methods, the Manhattan plots show that both versions of MANOVA, and SMAT detect SNPs from the case-control sample that are not found significant in the cohort. It is noteworthy that these SNPs are not detected as significant by POM-PS (Table 4). For all the SNPs listed in Table 4, we find consistency of $p_{\text{POM-PS}}$ from the case-control sample with p_{MANOVA} from the cohort. The other methods exhibit stronger p-values in the case-control sample compared to the cohort even though lower power is expected for the case-control sample owing to the much reduced sample size. Further, we found no known association result for any of these SNPs or any known SNP in linkage disequilibrium (LD, $r^2 > 0.5$) with these SNPs (Table S9). These SNPs have nominal p-values in the GIANT study meta-analyses of anthropometric traits (e.g., Locke et al., 2015). Thus, the signals found by unadjusted MANOVA, T2D-adjusted MANOVA and SMAT from the case-control sample are potentially spurious signals (Table S10). Only POM-PS analysis of the case-control sample does not yield any of these potentially false signals and seems to reflect a picture that one would obtain if it were a random sample. Moreover, POM-PS is computationally advantageous over other methods. For instance, SMAT and POM-PS respectively took 172 and 52 minutes to test 10,530 single SNP associations from chromosome 22 on an Intel(R) Xeon(R) CPU X5660 @2.80GHz processor.

The Manhattan plots for the cohort in Figure 5 show some significant associations in chromosomes 2, 6 and 16 (refer Table S8), which no method could detect from the case-control sample. To verify if the non-detection of these cohort-significant SNPs from our case-control sample is, indeed, due to the small sample size and is not an artifact of the particular sample we sampled from the original cohort, we conducted GWAS of 100 random samples (of the same size as the case-control sample) drawn from the cohort. The very low proportions (< 0.02) of random samples with significant findings for these SNPs substantiate our belief (Table S11).

5. Discussion

In this paper, we investigate the performances of several methods for testing the association of a SNP with multiple quantitative secondary phenotypes in case-control studies (non-randomized studies). Currently there are very few methods that focus on the joint analysis of multiple secondary traits, some of which we discuss in section 2. Our results show that no single method can maintain correct type I error under different scenarios of dependency among the secondary phenotypes, the SNP and the primary disease status (case-control variable). Severely inflated error levels in some association scenarios indicate that the association signals detected in GWASs of secondary traits using these methods may be spurious.

We consider six underlying association scenarios depicted by DAGs 'A'-'F' and establish theoretical conditions as to when *ad hoc* approaches provide valid tests of the null. For DAG 'F', adjustment of the primary disease status as a covariate is required to maintain type I error. On the other hand, for DAG 'E', such an adjustment can, in fact, inflate the type I error considerably. A background knowledge about the causal structure of the problem is necessary for deciding whether the primary disease status should be adjusted in the model or not (Hernán et al., 2002). SMAT, which was specifically designed to take the ascertainment bias into account, does not provide valid test of the null when the secondary phenotypes are affected by the primary disease outcome (DAG 'F').

Building on our theoretical and empirical studies, we propose a new method POM-PS, which uses a proportional odds model for X with estimated propensity score adjusted as a covariate. The propensity score is modeled as the conditional probability of being a case given the secondary phenotypes and other measured non-genetic covariates. Note that this propensity score needs to be calculated only once for a given GWAS. Our choice of adjusting this propensity score eliminates the confounding effect of primary disease status on the secondary traits as evident from our empirical results (section 3) as well as theoretical results (Supplementary S4). Unlike existing methods, POM-PS does not require knowledge of the underlying causal structure to provide valid test of association. Simulation experiments show that it maintains correct type I error under all DAGs 'A'-'F' (for both rare and common diseases, and in the presence of unmeasured covariates affecting disease status). Additionally, POM-PS is powerful in detecting cross-phenotype associations. The real data study using METSIM adiposity traits establishes that association analysis of multiple traits from a case-control sample using POM-PS can potentially reflect a picture that one would obtain if it were a random sample from the underlying (target) population. The reverse regression framework of POM-PS does not require multivariate normality assumption for the traits (Supplementary S11), which is often difficult to ensure, and allows us to accommodate both binary and continuous traits. It is also robust to moderate non-linearity and non-additivity of the primary trait (Supplementary S12). The fast & simple implementation of POM-PS using known software makes it applicable to large case-control GWASs with one or many secondary traits.

Finally, the simulation scenarios used to establish the performance of POM-PS are not exhaustive. For example, we have assumed no interactive effect of X and Y in the population model for D . In the presence of outliers, MANOVA or POM is substantially inflated for low MAF of the SNP (O'Reilly et al., 2012), and hence trait outliers may significantly affect POM-PS. We simulated data for an additive model only and did not consider non-additive genetic model or interactions. We intend to study how power of POM-PS would be affected in such situations. It is also of interest to develop an approach for testing association of a set of multiple SNPs with multiple secondary traits, which, to our knowledge, remains unexplored. Further, POM-PS currently analyzes secondary phenotypes when the primary trait is binary. For primary trait with three or more categories, one may employ a proportional odds model for the propensity score (Joffe and Rosenbaum, 1999). Lin et al. (2013) and more recently Tao et al. (2015) proposed likelihood-based methods when individuals with extreme values of continuous traits are sampled.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the NIH grant R01-DA033958, and the Doctoral Dissertation Fellowship of the University of Minnesota Graduate School. This work was carried out in part using computing resources at the Dept. of Psychology, University of Minnesota and at the Center for Statistical Genetics, Dept. of Biostatistics, University of Michigan. We are grateful to Dr. Michael Boehnke and Dr. Markku Laakso for kindly providing access to the METSIM data, and allowing us to analyze it. We thank the referees for a prompt and careful review of our work.

Appendix

Result 1

Under the null hypothesis of no association in the underlying population, the secondary phenotypes and the genotype are independently distributed in the sampling population for DAGs ‘A’, ‘B’, ‘C’, ‘D’. As a consequence, existing methods for joint association analysis of multiple phenotypes, irrespective of whether case-control status is adjusted or not, yield valid inference.

Result 2

Under the null hypothesis of no association in the underlying population, $\mathbf{Y}$ and $\mathbf{X}$, conditional on $\mathbf{D}$, are not independently distributed in the sampling population for DAG ‘E’. Consequently, existing methods for joint association analysis of multiple phenotypes that adjust case-control status as a covariate do not yield valid inference.

Result 3

Under the null hypothesis of no association in the underlying population, the secondary phenotypes and the genotype are not independently distributed in the sampling population for DAG ‘F’. As a consequence, unadjusted methods for joint association analysis of multiple phenotypes do not yield valid inference. However, in the sampling population, the secondary phenotypes and the genotype are independent conditional on the case-control status, and hence methods that adjust case-control status as a covariate yield valid inference.

References

- Allen AS, Satten GA. Control for confounding in case-control studies using the stratification score, a retrospective balancing score. *Am J Epidemiol.* 2011; 173:752–760. [PubMed: 21402731]
- Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivar Behav Res.* 2011; 46:399–424.
- Basu S, Zhang Y, Ray D, Miller MB, Iacono WG, McGue M. A rapid gene-based genome-wide association test with multivariate traits. *Hum Hered.* 2013; 76:53–63. [PubMed: 24247328]
- Epstein MP, Allen AS, Satten GA. A simple and improved correction for population stratification in case-control studies. *Am J Hum Gen.* 2007; 80:921–930.
- Ghosh A, Wright FA, Zou F. Unified analysis of secondary traits in case-control association studies. *J Am Stat Assoc.* 2013; 108:566–576.

- Haneuse S, Schildcrout J, Crane P, Sonnen J, Breitner J, Larson E. Adjustment for selection bias in observational studies with application to the analysis of autopsy data. *Neuroepidemiology*. 2009; 32:229–239. [PubMed: 19176974]
- He J, Li H, Edmondson AC, Rader DJ, Li M. A gaussian copula approach for the analysis of secondary phenotypes in case-control genetic association studies. *Biostatistics*. 2012; 13:497–508. [PubMed: 21933777]
- Hernán MA, Hernández-Díaz S, Werler MW, Mitchell A. Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *Am J Epidemiol*. 2002; 155:176–184. [PubMed: 11790682]
- Jiang Y, Scott AJ, Wild CJ. Secondary analysis of case-control data. *Statist Med*. 2006; 25:1323–1339.
- Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. *Am J Epidemiol*. 1999; 150:327–333. [PubMed: 10453808]
- Lee S, Wu M, Lin X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*. 2012; 13:762–775. [PubMed: 22699862]
- Lin DY, Zeng D. Proper analysis of secondary phenotype data in case-control association studies. *Genet Epidemiol*. 2009; 33:256–265. [PubMed: 19051285]
- Lin DY, Zeng D, Tang ZZ. Quantitative trait analysis in sequencing studies under trait-dependent sampling. *Proc Natl Acad Sci*. 2013; 110:12247–12252. [PubMed: 23847208]
- Liu DJ, Leal SM. A flexible likelihood framework for detecting associations with secondary phenotypes in genetic studies using selected samples: application to sequence data. *Eur J Hum Genet*. 2012; 20:449–456. [PubMed: 22166943]
- Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015; 518:197–206. [PubMed: 25673413]
- Lutz SM, Hokanson JE, Lange C. An alternative hypothesis testing strategy for secondary phenotype data in case-control genetic association studies. *Front Genet*. 2014; 5:188. [PubMed: 25071819]
- Ma Y, Carroll RJ. Semiparametric estimation in the secondary analysis of case-control studies. *J R Stat Soc Series B Stat Methodol*. 2016; 78:127–151. [PubMed: 26834506]
- Monsees GM, Tamimi RM, Kraft P. Genome-wide association scans for secondary traits using case-control samples. *Genet Epidemiol*. 2009; 33:717–728. [PubMed: 19365863]
- O'Reilly P, Hoggart C, Pomyen Y, Calboli C, et al. MultiPhen: Joint model of multiple phenotypes can increase discovery in GWAS. *PLoS One*. 2012; 7:e34861. [PubMed: 22567092]
- Pan W. Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genet Epidemiol*. 2009; 33:497–507. [PubMed: 19170135]
- Ray D, Li X, Pan W, Pankow J, Basu S. A Bayesian partitioning model for detection of multilocus effects in case-control studies. *Hum Hered*. 2015; 79:69–79. [PubMed: 26044550]
- Ray D, Pankow J, Basu S. USAT: a unified score-based association test for multiple phenotype-genotype analysis. *Genet Epidemiol*. 2016; 40:20–34. [PubMed: 26638693]
- Richardson DB, Rzehak P, Klenk J, Weiland SK. Analyses of case-control data for additional outcomes. *Epidemiology*. 2007; 18:441–445. [PubMed: 17473707]
- Rosenbaum PR. Model-based direct adjustment. *J Am Stat Assoc*. 1987; 82:387–394.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55.
- Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat*. 1985; 39:33–38.
- Schifano ED., Bar, H., Harel, O. *Innovative Statistical Methods for Public Health Data*. Springer; 2015. Methods for analyzing secondary outcomes in public health case-control studies; p. 3-15.
- Schifano ED, Li L, Christiani DC, Lin X. Genome-wide association analysis for multiple continuous secondary phenotypes. *Am J Hum Genet*. 2013; 92:744–759. [PubMed: 23643383]
- Song X, Ionita-Laza I, Liu M, Reibman J, We Y. A general and robust framework for secondary traits analysis. *Genetics*. 2016; 202:1329–1343. [PubMed: 26896329]

- Stan áková A, Javorsk'y M, Kuulasmaa T, Haffner SM, Kuusisto J, Laakso M. Changes in insulin sensitivity and insulin release in relation to glycemia and glucose tolerance in 6416 Finnish men. *Diabetes*. 2009; 58:1212–1221. [PubMed: 19223598]
- Stan áková A, Kuulasmaa T, Paananen J, Jackson AU, Bonnycastle LL, Collins FS, Boehnke M, Kuusisto J, Laakso M. Association of 18 confirmed susceptibility loci for type 2 diabetes with indices of insulin release, proinsulin conversion, and insulin sensitivity in 5327 non-diabetic Finnish men. *Diabetes*. 2009; 58:2129–2136. [PubMed: 19502414]
- Tao R, Zeng D, Franceschini N, North KE, Boerwinkle E, Lin DY. Analysis of sequence data under multivariate trait-dependent sampling. *J Am Stat Assoc*. 2015; 110:560–572. [PubMed: 26366025]
- Tchetgen EJT. A general regression framework for a secondary outcome in case–control studies. *Biostatistics*. 2014; 15:117–128. [PubMed: 24152770]
- Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, et al. Biological, clinical, and population relevance of 95 loci for blood lipids. *Nature*. 2010; 466:707–713. [PubMed: 20686565]
- Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, Styrkarsdottir U, Gretarsdottir S, Thorlacius S, Jonsdottir I, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet*. 2009; 41:18–24. [PubMed: 19079260]
- Wang J, Shete S. Estimation of odds ratios of genetic variants for the secondary phenotypes associated with primary diseases. *Genet Epidemiol*. 2011; 35:190–200. [PubMed: 21308766]
- Wei J, Carroll RJ, Müller UU, Keilegom IV, Chatterjee N. Robust estimation for homoscedastic regression in the secondary analysis of case–control data. *J R Stat Soc Series B Stat Methodol*. 2013; 75:185–206. [PubMed: 23637568]
- Xing C, McCarthy JM, Dupuis J, Adrienne Cupples L, Meigs JB, Lin X, Allen AS. Robust analysis of secondary phenotypes in case-control genetic association studies. *Statist Med*. 2016; 35:4226–4237.
- Yang Q, Wang Y. Review article: Methods for analyzing multivariate phenotypes in genetic association studies. *J Probab Stat*. 2012; 2012:13.
- Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat Methods*. 2014; 11:407–409. [PubMed: 24531419]

Figure 1.

The directed acyclic graphs (DAGs) showing 6 possible scenarios of association of X (genotype), Y (multiple secondary phenotypes), D (primary trait or disease status) and S (sampling indicator) in the underlying population. These scenarios have been considered following Monsees et al. (2009). For simplicity, it is assumed that the sampling mechanism depends only on the case-control variable D . The direction of the arrows indicate cause-effect relationships. Without an arrow between X and Y , the DAGs represent null scenarios. The alternative scenario is represented by an arrow from X to Y . For ease of notation, the distribution of any random variable/vector (discrete or continuous) is denoted using square braces $[\cdot]$. For example, $[X]$ denotes the distribution (probability mass function) of X , $[Y]$ denotes the joint distribution (probability density function) of Y , $[Y|X]$ denotes the conditional distribution of Y given X and so on.

Figure 2.

The empirical powers of the existing approaches along with our method POM-PS under DAG 'E' at 1% level. MANOVA-adj (SSU-adj) means the method MANOVA (SSU) where D is adjusted as a covariate. $K = 5$ or 20 traits have been simulated at different within trait correlation values $\rho = 0.2, 0.4, 0.6$. For each value of K and ρ , there were $N_R = 500$ datasets consisting of 200 cases and 200 controls. Same effect size of 0.395 (proportion of variance explained is 0.5%) was used for the traits that are associated. The power is plotted along y-axis while the fraction of traits associated with the SNP is plotted along x-axis.

Figure 3.

The empirical powers of the existing approaches along with our method POM-PS under DAG 'F' at 1% level. MANOVA-adj (SSU-adj) means the method MANOVA (SSU) where D is adjusted as a covariate. $K = 5$ or 20 traits have been simulated at different within trait correlation values $\rho = 0.2, 0.4, 0.6$. For each value of K and ρ , there were $N_R = 500$ datasets consisting of 200 cases and 200 controls. Same effect size of 0.395 (proportion of variance explained is 0.5%) was used for the traits that are associated. The power is plotted along y-axis while the fraction of traits associated with the SNP is plotted along x-axis.

Figure 4. Manhattan Plots for the SNPs with MAF $\geq 1\%$ on chromosomes 1–22. Genome wide association analysis is based on the case-control sample of size 2,762. Red horizontal line corresponds to negative log-transformed significance level of 5×10^{-7} .

(a) (Unadjusted) MANOVA for entire cohort

(b) (Unweighted) SMAT for entire cohort

Figure 5.

Manhattan Plots for the SNPs with MAF $\geq 1\%$ on chromosomes 1–22. Genome wide association analysis is based on the entire cohort of size 9,964. Red horizontal line corresponds to negative log-transformed significance level of 5×10^{-7} .

Table 1

List of parameter values considered under different scenarios (DAGs) and trait models in our simulation experiments in sections 3.2–3.3.

K	Parameters	$\mathbf{A}$	$\mathbf{B}$	$\mathbf{C}$	$\mathbf{D}$	$\mathbf{E}$	$\mathbf{F}$
2	f	0.2	0.2	0.2	0.2	0.2	0.2
	β_0^*	0	0	0	0	0	0
	h (null trait)	0	0	0	0	0	0
	h (non-null trait)	0.002	0.002	0.002	0.002	0.002	0.002
	ϕ^*	0	0	0	3.5	0	3.5
	σ	9.98	9.98	9.98	9.98	9.98	9.98
	α_0^*	-2.197	-2.197	-2.197	-2.197	-2.197	-2.197
	α^*	0	0	log(3)	0	log(3)	0
	α_x^*	0	log(1.5)	0	0	log(1.5)	log(1.5)
	N_R (null datasets)	10^4	10^4	10^4	10^4	10^4	10^4
	N_R (non-null datasets)	500	500	500	500	500	500
	$l_U = u$	2000	2000	2000	2000	2000	2000
5	f	0.2	0.2	0.2	0.2	0.2	0.2
	β_0^*	0	0	0	0	0	0
	h (null trait)	0	0	0	0	0	0
	h (non-null trait)	0.005	0.005	0.005	0.005	0.005	0.005
	ϕ^*	0	0	0	3.5	0	3.5
	σ	9.95	9.95	9.95	9.95	9.95	9.95
	α_0^*	-2.197	-2.197	-2.197	-2.197	-2.197	-2.197
	α^*	0	0	log(3)	0	log(3)	0
	α_x^*	0	log(1.5)	0	0	log(1.5)	log(1.5)
	N_R (null datasets)	5000	5000	5000	5000	5000	5000
	N_R (non-null datasets)	500	500	500	500	500	500

200	200	200	200	200	200	(non-null)	$u_1 = u$
2000	2000	2000	2000	2000	2000	(null)	$u_1 = u_0$
,F,	,E,	,D,	,C,	,B,	,A,	Parameters	K

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Estimated type I errors of the afore mentioned association tests for $K = 2$ correlated traits for disease prevalence $\pi = 0.1$. Type I error rate is calculated as the proportion of null datasets (out of 10, 000) with p-value ≤ 0.01 . Event [S] means the case-control sample is analyzed without any adjustment for ascertainment. Event [S, D] means the analyses are based on the case-control sample and adjusted for D. Although SMAT and POM-PS consider adjustment for ascertainment, their type I error estimates are placed under [S] to indicate that these methods have not considered adjustment of D as covariate in their respective models. The inflated errors are **boldfaced**. The very conservative errors are in italics.

DAG	Method	$\rho = -0.8$			$\rho = -0.2$			$\rho = 0.2$			$\rho = 0.8$		
		[S]	[S, D]	[S]	[S, D]	[S]	[S, D]	[S]	[S, D]	[S]	[S, D]	[S]	[S, D]
A	MANOVA	0.0104	0.0105	0.0101	0.0102	0.0114	0.011	0.0105	0.0105	0.0105	0.0105	0.0105	0.0105
	SSU	0.0103	0.01	0.0113	0.0113	0.0119	0.0118	0.0099	0.0101	0.0099	0.0101	0.0099	0.0101
	LRT-POM	0.0111	0.0109	0.0103	0.0103	0.0112	0.0112	0.0104	0.0105	0.0104	0.0105	0.0104	0.0105
	SMAT ($\pi_{SMAT}=0.1$)	0.0096	—	0.0115	—	0.0114	—	0.0115	—	0.0115	—	0.0115	—
	SMAT ($\pi_{SMAT}=0.2$)	0.0099	—	0.0106	—	0.0115	—	0.0114	—	0.0114	—	0.0114	—
	POM-PS	0.0103	—	0.0088	—	0.0114	—	0.0094	—	0.0094	—	0.0094	—
B	MANOVA	0.0105	0.0115	0.0113	0.0104	0.0103	0.0101	0.0101	0.0101	0.0101	0.0101	0.0101	0.0101
	SSU	0.0112	0.0105	0.0105	0.0096	0.0086	0.008	0.0105	0.0107	0.0105	0.0107	0.0105	0.0107
	LRT-POM	0.0098	0.0103	0.0109	0.0113	0.0101	0.0103	0.0103	0.0108	0.0103	0.0108	0.0103	0.0108
	SMAT ($\pi_{SMAT}=0.1$)	0.0109	—	0.0106	—	0.0099	—	0.0115	—	0.0115	—	0.0115	—
	SMAT ($\pi_{SMAT}=0.2$)	0.0104	—	0.0112	—	0.009	—	0.012	—	0.012	—	0.012	—
	POM-PS	0.0089	—	0.011	—	0.0099	—	0.009	—	0.009	—	0.009	—
C	MANOVA	0.0101	0.0101	0.0116	0.0113	0.011	0.0102	0.0097	0.0102	0.0097	0.0102	0.0097	0.0102
	SSU	0.0113	0.011	0.0113	0.0117	0.0117	0.0099	0.01	0.0089	0.01	0.0089	0.01	0.0089
	LRT-POM	0.0097	0.0114	0.0108	0.011	0.0101	0.0103	0.0094	0.0096	0.0094	0.0096	0.0094	0.0096
	SMAT ($\pi_{SMAT}=0.1$)	0.0111	—	0.011	—	0.0099	—	0.0116	—	0.0116	—	0.0116	—
	SMAT ($\pi_{SMAT}=0.2$)	0.0117	—	0.0103	—	0.0101	—	0.012	—	0.012	—	0.012	—
	POM-PS	0.0095	—	0.0099	—	0.0093	—	0.0092	—	0.0092	—	0.0092	—
D	MANOVA	0.0106	0.0097	0.0094	0.0107	0.0113	0.0109	0.0077	0.0089	0.0077	0.0089	0.0077	0.0089
	SSU	0.0094	0.0093	0.0088	0.0093	0.0109	0.0117	0.01	0.0098	0.01	0.0098	0.01	0.0098
	LRT-POM	0.0103	0.0095	0.0097	0.0105	0.0106	0.0106	0.0089	0.0099	0.0089	0.0099	0.0089	0.0099

DAG	Method	$\rho = -0.8$			$\rho = -0.2$			$\rho = 0.2$			$\rho = 0.8$		
		[S]	[S, D]	[S]	[S]	[S, D]	[S]	[S]	[S, D]	[S]	[S]	[S, D]	[S]
E	SMAT ($\pi_{SMAT} = 0.1$)	0.0099	—	0.0101	—	0.0118	—	0.0113	—	0.0113	—	—	—
	SMAT ($\pi_{SMAT} = 0.2$)	0.0114	—	0.0095	—	0.0112	—	0.0106	—	0.0106	—	—	—
	POM-PS	0.0108	—	0.0091	—	0.0091	—	0.0106	—	0.0106	—	—	—
	MANOVA	0.0133	0.5961	0.0094	0.4123	0.0096	0.3111	0.009	0.2103	0.009	0.2103	—	—
	SSU	0.01	0.0161	0.0092	0.0686	0.0087	0.2049	0.0101	0.2898	0.0101	0.2898	—	—
	LRT-POM	0.0148	0.5697	0.0097	0.3855	0.0089	0.2968	0.0096	0.2006	0.0096	0.2006	—	—
	SMAT ($\pi_{SMAT} = 0.1$)	0.0384	—	0.0459	—	0.0404	—	0.0317	—	0.0317	—	—	—
	SMAT ($\pi_{SMAT} = 0.2$)	0.0106	—	0.0172	—	0.014	—	0.0142	—	0.0142	—	—	—
	POM-PS	0.0109	—	0.0101	—	0.0098	—	0.0089	—	0.0089	—	—	—
	MANOVA	0.9999	0.0096	0.9817	0.009	0.9344	0.0101	0.8149	0.011	0.8149	0.011	—	—
	SSU	0.9963	0.0093	0.9864	0.0096	0.9633	0.0103	0.888	0.0108	0.888	0.0108	—	—
	LRT-POM	0.9998	0.009	0.9775	0.0102	0.9274	0.0094	0.7964	0.0122	0.7964	0.0122	—	—
F	SMAT ($\pi_{SMAT} = 0.1$)	0.806	—	0.2592	—	0.1636	—	0.1025	—	0.1025	—	—	—
	SMAT ($\pi_{SMAT} = 0.2$)	0.9967	—	0.7921	—	0.6118	—	0.4253	—	0.4253	—	—	—
	POM-PS	0.0096	—	0.0092	—	0.0089	—	0.0101	—	0.0101	—	—	—

Table 3

Estimated type I error of the proposed approach POM-PS for $K = 5$ or 20 correlated traits with pairwise correlations $\rho = 0.2, 0.4, 0.6$. Type I error rate is estimated as the proportion of null datasets (out of $5,000$) with p-value 0.01 . Only DAGs ‘E’ and ‘F’ are reported since we expect all methods to be valid under DAGs ‘A’-‘D’.

DAG	$\rho = 0.2$		$\rho = 0.4$		$\rho = 0.6$	
	$K = 5$	$K = 20$	$K = 5$	$K = 20$	$K = 5$	$K = 20$
E	0.009	0.0102	0.0122	0.0086	0.0118	0.008
F	0.0112	0.0106	0.0086	0.0094	0.0078	0.009

Table 4

METSIM Study: A summary of association p-values from the case-control sample analysis. The case-control sample has been analyzed using four different approaches: unadjusted MANOVA, T2D-adjusted MANOVA, SMAT and our method POM-PS. The SNPs listed here are the ones found significant by either of these four analysis approaches. The corresponding significant p-values are bold-faced. Notice that all methods except POM-PS detect signals based on the case-control sample. To provide a benchmark for comparison, the original cohort has also been analyzed using unadjusted MANOVA. Noticeably the p-values of POM-PS are similar to the ones obtained from the analysis of the entire cohort. In other words, POM-PS based on the selected sample seem to reflect a picture that one would obtain if it were a random sample. Here, chr is the abbreviation for chromosome number.

rsID	chr	position	MAF	Case-control sample analysis (n = 2,762)				Cohort analysis (n = 9,964)			
				Unadjusted MANOVA	T2D-adjusted MANOVA	SMAT	POM-PS	Unadjusted MANOVA	Unweighted SMAT	Known association result	
rs2034055 ^a	2	185282179	0.18	1.0 × 10 ⁻⁵	2.8 × 10 ⁻⁵	7.7 × 10⁻⁸	5.4 × 10 ⁻¹	5.3 × 10 ⁻³	4.4 × 10 ⁻⁴	Not known	
rs7585186 ^a	2	226460340	0.40	1.2 × 10 ⁻³	2.2 × 10 ⁻³	2.2 × 10⁻⁷	1.2 × 10 ⁻¹	4.4 × 10 ⁻²	2.6 × 10 ⁻²	Not known	
rs12477972 ^a	2	226482872	0.26	1.2 × 10 ⁻³	1.5 × 10 ⁻³	3.8 × 10⁻⁷	4.5 × 10 ⁻¹	1.2 × 10 ⁻¹	5.8 × 10 ⁻²	Not known	
rs11884382 ^a	2	226516346	0.21	5.8 × 10 ⁻⁶	1.1 × 10 ⁻⁵	1.1 × 10⁻⁸	1.5 × 10 ⁻¹	1.9 × 10 ⁻³	2.3 × 10 ⁻³	Not known	
rs6737960 ^a	2	226583962	0.22	4.9 × 10 ⁻⁶	5.8 × 10 ⁻⁶	1.9 × 10⁻⁸	7.3 × 10 ⁻²	1.0 × 10 ⁻³	1.1 × 10 ⁻³	Not known	
rs77045370 ^{* a}	9	6434112	0.012	2.4 × 10 ⁻⁴	3.8 × 10 ⁻⁴	2.1 × 10⁻⁷	1.1 × 10 ⁻¹	1.3 × 10 ⁻¹	2.5 × 10 ⁻²	Not known	
rs1616208 ^b	9	132298733	0.31	1.8 × 10⁻⁸	2.3 × 10 ⁻⁶	4.1 × 10 ^{-B}	3.2 × 10 ⁻³	2.3 × 10 ⁻³	5.1 × 10 ⁻²	Not known	
rs117419007 ^{† bc}	17	78021155	0.016	2.0 × 10⁻⁷	1.9 × 10⁻⁷	5.6 × 10 ⁻¹	2.4 × 10 ⁻⁵	7.2 × 10 ⁻⁵	9.8 × 10 ⁻¹	Not known	

^aSignal detected by SMAT from the selected sample.
^{*}Illumina OmniExpress Exome Chip ID is exm738338.
^bSignal detected by unadjusted MANOVA from the selected sample.
[†]Illumina OmniExpress Exome Chip ID is exm2253275.
^cSignal detected by T2D-adjusted MANOVA from the selected sample.