

## Research Article

# Fast and Accurate Genome-Wide Association Test of Multiple Quantitative Traits

Baolin Wu <sup>1</sup> and James S. Pankow<sup>2</sup>

<sup>1</sup>Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN, USA

<sup>2</sup>Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA

Correspondence should be addressed to Baolin Wu; [baolin@umn.edu](mailto:baolin@umn.edu)

Received 27 September 2017; Accepted 24 January 2018; Published 18 March 2018

Academic Editor: Zoran Bursac

Copyright © 2018 Baolin Wu and James S. Pankow. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Multiple correlated traits are often collected in genetic studies. By jointly analyzing multiple traits, we can increase power by aggregating multiple weak effects and reveal additional insights into the genetic architecture of complex human diseases. In this article, we propose a multivariate linear regression-based method to test the joint association of multiple quantitative traits. It is flexible to accommodate any covariates, has very accurate control of type I errors, and offers very competitive performance. We also discuss fast and accurate significance  $p$  value computation especially for genome-wide association studies with small-to-medium sample sizes. We demonstrate through extensive numerical studies that the proposed method has competitive performance. Its usefulness is further illustrated with application to genome-wide association analysis of diabetes-related traits in the Atherosclerosis Risk in Communities (ARIC) study. We found some very interesting associations with diabetes traits which have not been reported before. We implemented the proposed methods in a publicly available R package.

## 1. Introduction

Over the past ten years, many epidemiologic studies have used genome-wide association studies (GWAS) to identify genetic components of many complex human diseases. These large cohort studies often collected a broad array of correlated traits that often reflect common physiological processes. By jointly analyzing these correlated traits, we can often gain more power by aggregating multiple weak effects and shed light on the mechanisms underlying complex human diseases [1].

There have been many methods proposed recently to detect SNP association with multiple correlated traits (see, e.g., [2–13]). A direct approach is based on the minimum trait  $p$  value [6], which typically requires permutations to compute significance  $p$  value. A related approach is the trait-based association test using an extended Simes procedure (TATES; [10]) that combines the univariate trait  $p$  values while correcting for the correlations among the multivariate traits. Various dimension reduction methods that summarize the multivariate traits into a univariate outcome are also

proposed, which then apply the traditional univariate association test. Examples include the principal component analysis (PCA) [2], principal components of heritability (PCH) [3], and averaging longitudinally observed traits [7, 14]. PCA is an unsupervised dimension reduction and the top PC may not necessarily reflect the association signal. Sample splitting is typically used in PCH for significance calculations and may lead to loss of power.

Multivariate trait testing methods generally perform better than univariate analysis-based approach [15]. Among the multivariate testing methods, a popular approach is the canonical correlation analysis (CCA) [4, 16, 17], which is fast to compute but not flexible and is unable to accommodate covariates. Liu et al. [5] proposed the GEE model [18] to jointly analyze one continuous and one binary trait. In Avery et al. [19] and He et al. [11], GEE-based marginal generalized linear modeling of multivariate traits is adopted for efficient multitrait association testing. Schifano et al. [20] proposed a closely related GEE-based scaled marginal association test of multiple secondary continuous traits. Sitlani et al. [13] explored the GEE modeling of longitudinally measured traits

for association test. These GEE-based methods typically explicitly avoided modeling the trait correlations. Another set of multivariate approaches is based on the inverted regression of genotypes to test the overall trait effects. For example, the proportional odds regression modeling of genotypes was proposed as a convenient approach to testing multitrait associations [8, 21, 22]. A related adjacent category logistic regression of genotypes was proposed by Wu and Pankow [12]. Inverted regression approach does not easily accommodate imputed SNPs and has generally used the “best-guess” genotypes, which is known to be leading to a loss of power. In contrast, the multivariate trait regression approach can easily test imputed SNPs by using the imputation dosage as covariate.

In this article, we explore an alternative multivariate regression framework to explicitly model the trait correlation and adjust for covariates to test multitrait associations. We compute the analytical  $p$  values for the proposed tests based on the  $F$ -distributions that offer very accurate type I error control with good finite sample performance. We also exploit the parallel nature of genome-wide association test to develop very efficient numerical algorithms that are extremely scalable to genome-wide association tests of millions of SNPs. We demonstrate through extensive numerical studies that the proposed methods have very competitive performance compared to existing methods. We further illustrate the usefulness of the proposed methods through an application to genome-wide association study of multiple diabetes-related glycemic traits.

## 2. Methods

We first discuss a multivariate linear regression-based framework for modeling the multiple quantitative traits and then derive the Wald type statistics for testing multitrait associations.

**2.1. Multivariate Linear Regression Model.** Consider  $m$  continuous traits  $Y = (y_1, \dots, y_m)^T$ , a covariate vector  $X = (x_1, \dots, x_p)^T$  of length  $p$  (which could contain an ancestry indicator or principal components), and a genotype score  $G$  coding the number of minor alleles. Consider the multivariate normal trait model:

$$Y = \beta_0 + \beta_X X + G\beta_1 + \epsilon, \quad (1)$$

where  $\beta_0$  is a vector of length  $m$ ,  $\beta_X$  is an  $m \times p$  matrix,  $\beta_1$  is a vector of length  $m$ , and the random error  $\epsilon$  is of length  $m$  and is assumed to follow a zero mean multivariate normal distribution with covariance  $\Sigma$ ,  $\epsilon \sim N(0, \Sigma)$ . Multivariate trait association amounts to testing  $H_0 : \beta_1 = 0$ . Here we have assumed the same covariates for all traits, which is the case for our ARIC study GWAS example (see Application to ARIC GWAS of Glycemic Traits) and many typical GWAS. In the supplementary materials (available here), we discuss the possible scenario with different covariates for each trait. The trait model (1) is a multivariate linear model (MLM; see, e.g., [23, chapter 8] and [24, chapter 9]).

Given observations for  $n$  unrelated individuals, for individual  $i$ , denote  $Y_i$  as the outcome,  $X_i$  as the covariate, and

$G_i$  as the genotype score. Denote  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ ,  $\mathbf{X} = (X_1, \dots, X_n)^T$ ,  $\mathbf{G} = (G_1, \dots, G_n)^T$ , and design matrix  $\mathbf{Z} = (\mathbf{1}_n, \mathbf{X}, \mathbf{G})$  of dimension  $n \times (p + 2)$ , where  $\mathbf{1}_n = (1, \dots, 1)^T$  is a column vector of  $n$  ones.

Denote the  $m \times (p + 2)$  parameter matrix  $\beta = (\beta_0, \beta_X, \beta_1)$ . We can check that the maximum likelihood estimators (MLEs) are (see, e.g., [23, p. 294])

$$\begin{aligned} \hat{\beta} &= \mathbf{Y}^T \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}, \\ \hat{\Sigma} &= \frac{1}{n} (\mathbf{Y} - \mathbf{Z} \hat{\beta})^T (\mathbf{Y} - \mathbf{Z} \hat{\beta}). \end{aligned} \quad (2)$$

**2.2. Conducting Multivariate Association Tests.** Denote the vector operator  $\text{vec}()$ , which stacks the columns of a matrix into a vector. Denote  $\mathbf{A} = \mathbf{Z}^T \mathbf{Z}$ . For the MLEs (2) of the MLM model (1), we can check that (see, e.g., [23, p. 296])

$$\begin{aligned} E[\text{vec}(\hat{\beta})] &= \text{vec}(\beta), \\ \text{Cov}[\text{vec}(\hat{\beta})] &= \mathbf{A}^{-1} \otimes \Sigma, \end{aligned} \quad (3)$$

where  $\otimes$  denotes the Kronecker product and  $n\hat{\Sigma}$  independently follows a Wishart distribution,  $W_m(\Sigma, n - p - 2)$ , with  $n - p - 2$  degrees of freedom (DFs) and scale matrix  $\Sigma$ .

Define the  $n \times (p + 1)$  design matrix  $\mathbf{Z}_0 = (\mathbf{1}_n, \mathbf{X})$  and the corresponding  $n \times n$  hat matrix  $H = \mathbf{Z}_0 (\mathbf{Z}_0^T \mathbf{Z}_0)^{-1} \mathbf{Z}_0^T$ . Let  $\mathbf{P} = \mathbf{I} - H$  and  $G_e = \mathbf{P}\mathbf{G}$ . Here  $\mathbf{I}$  is an  $n \times n$  identity matrix. We can check that

$$\hat{\beta}_1 = \frac{\mathbf{Y}^T G_e}{G_e^T G_e}, \quad (4)$$

$$\text{Cov}(\hat{\beta}_1) = (G_e^T G_e)^{-1} \Sigma.$$

We test the multitrait association with the following Wald statistic:

$$Q = (G_e^T G_e) \hat{\beta}_1^T \hat{\Sigma}^{-1} \hat{\beta}_1. \quad (5)$$

Note that  $\hat{\beta}_1$  and  $\hat{\Sigma}$  are independent. Under the null hypothesis,  $((n - p - 1 - m)/mn)Q$  follows the  $F$ -distribution with  $(m, n - p - 1 - m)$  DFs (see, e.g., [25, p. 541]).

In the supplementary materials, we analytically show that the CCA test approach [4] is equivalent to a Score test statistic under the MLM model (1) when there are no covariates other than the genotype. Therefore, the proposed MLM-based Wald test can be treated as a natural and flexible generalization of the CCA: (I) it can accommodate any covariates; (II) it is based on the more powerful Wald test instead of the Score test for an association test of quantitative traits; (III) it has an exact  $F$ -distribution for the multivariate normally distributed traits and hence has very accurate control of type I errors for any sample sizes without the need of asymptotic approximation; and (IV) it is very fast to compute (see next section for details) and extremely scalable to genome-wide association tests of millions of SNPs.

When genetic effects are similar across traits, we can further improve the multivariate association test power using

a test statistic with 1-DF following the lines of O'Brien [26], which performed a Wald test of linear combinations of  $\beta_1$ . We can derive similar Wald tests under the MLM (1) (see supplementary materials for technical details). When the genotype effects are the same across different traits, we study the following test statistic:

$$T = \frac{\mathbf{1}_m^T \widehat{\Sigma}^{-1} \widehat{\beta}_1}{\sqrt{\mathbf{1}_m^T \widehat{\Sigma}^{-1} \mathbf{1}_m}}, \quad (6)$$

where  $\mathbf{1}_m$  is an  $m \times 1$  column vector of ones. When the scaled genotype effects are the same across different traits, we study the following test statistic:

$$T' = \frac{S^T \widehat{\Sigma}^{-1} \widehat{\beta}_1}{\sqrt{S^T \widehat{\Sigma}^{-1} S}}, \quad (7)$$

where  $S$  is a column vector of estimated standard errors:  $S = \sqrt{\text{diag}(\widehat{\Sigma})}$ .

Under the null hypothesis, both  $T$  and  $T'$  follow the asymptotic standard normal distribution. To improve the finite sample performance, we can compare  $((n-p-1-m)/n)T$  and  $((n-p-1-m)/n)T'$  to a  $t$ -distribution with  $(n-p-1-m)$ -DF.

**2.3. Efficient Computation of GWAS Wald Test Statistics.** For a typical GWAS with millions of SNPs, rather than fitting a MLM for each SNP, we developed very efficient algorithm to estimate the MLMs for all SNPs using matrix decomposition tricks following the line of Voorman et al. [27] as follows. For  $\mathbf{Z}_0$ , denote its singular value decomposition (SVD) as  $\mathbf{Z}_0 = \mathbf{U}\mathbf{D}\mathbf{V}^T$ , where  $\mathbf{U}$  is an  $n \times (p+1)$  matrix with orthogonal columns,  $\mathbf{D}$  is a  $(p+1) \times (p+1)$  diagonal matrix, and  $\mathbf{V}$  is a  $(p+1) \times (p+1)$  orthogonal matrix. The null MLM hat matrix can then be computed as  $H = \mathbf{U}\mathbf{U}^T$ , and  $G_e = G - \mathbf{U}(\mathbf{U}^T G)$ . Denote the null MLM residual matrix as  $\mathbf{E} = \mathbf{Y} - \mathbf{U}(\mathbf{U}^T \mathbf{Y})$ , and let  $\mathbf{V}_0 = \mathbf{E}^T \mathbf{E}$ . In (4), we have shown that the genotype effect can be efficiently computed as  $\widehat{\beta}_1 = \mathbf{Y}^T G_e / (G_e^T G_e)$ . We can then compute the covariance matrix MLE as  $\widehat{\Sigma} = \mathbf{V}_0/n - (G_e G_e^T) \widehat{\beta}_1 \widehat{\beta}_1^T / n$ . Here both  $\mathbf{V}_0$  and  $\mathbf{U}$  just need to be precomputed once and can be stored for use with all SNPs. Operationally we can also apply the popular PLINK tool [28] to test multitrait association. We first obtain the residuals of multivariate traits and genotypes adjusting for all covariates. We then input the residuals into the PLINK CCA test approach [4]. Technically, we need to adjust the PLINK output  $p$  value using an  $F$ -distribution with different DFs (see supplementary materials for technical details).

### 3. Results

**3.1. Simulation Studies.** We consider three forms of Wald statistics:  $Q$  is the omnibus test, and  $T$  and  $T'$  are the 1-DF test assuming common or common scaled effects. The GEE-based approaches of He et al. [11] are computationally very efficient, have been shown to appropriately control the type I errors, and have the overall best detection power compared to

the other methods (e.g., TATES of [10] and other univariate test-based methods) in extensive numerical studies. Here we compared the proposed methods to their GEE score tests, denoted as  $(Q_s, T_s, T'_s)$ , which are the  $m$ -DF omnibus test and 1-DF tests assuming a common effect or common scaled effect.

We consider a standard normal covariate  $X_1$  and a Bernoulli covariate  $X_2$  with probability of 0.5. The SNP genotype score  $G$  is simulated from a Binomial distribution,  $\text{Binom}(2, f_0)$ , where the minor allele frequency (MAF)  $f_0 = p_0 + p_1 X_2$ . Here  $X_2$  is essentially a population indicator and we have simulated SNPs under population stratification.

We conducted simulations for testing  $m = 2, 4, 8$  related traits of 1,000 unrelated individuals, respectively. Each time, we simulate the  $m$  traits from a multivariate normal distribution with a compound symmetry correlation matrix with correlation  $\rho$ . The first trait has a variance of 2 and all the other traits have unit variance. We set  $E(Y_i) = 1 + 0.5X_1 + 0.5X_2 + \gamma_i G$  for  $i = 1, 3, \dots, m-1$ , and  $E(Y_k) = 1 + X_1 + X_2 + \gamma_k G$  for  $k = 2, 4, \dots, m$ .

We used 10 million experiments to evaluate the type I error and  $10^5$  experiments to evaluate the power under various combinations of  $(\gamma_1, \dots, \gamma_m)$ . We conducted simulations for  $p_0 = (0.1, 0.3)$ ,  $p_1 = 0.1$ , and  $\rho = 0, 0.2, 0.5, 0.8$ . Here we report the results for  $m = 2, 8$ ,  $\rho = 0, 0.5$ , and  $p_0 = 0.1$ . The conclusions remain the same for other settings (data not shown).

Tables 1 and 2 summarize the estimated type I errors. Overall, the type I errors are well controlled for the proposed methods, while the GEE score tests are conservative, especially for large number of traits ( $m = 8$ ). In general, the proposed Wald tests follow the exact  $F$ -distribution under the null hypothesis and hence the type I errors are well controlled under all settings. The GEE tests rely on the large-sample asymptotic distribution and therefore generally we need large sample size to have better control of type I errors, especially for a larger number of traits (containing more model parameters).

Tables 3 and 4 summarize the power for  $m = 2$  and  $m = 8$ , respectively.  $T$  is the most powerful when  $\gamma_j$  are close to each other, and  $T'$  is the most powerful when  $\gamma_j/\sigma_j$  are close to each other. In general, the proposed MLM-based Wald tests perform better than the corresponding GEE-based score tests, especially when testing a large number of traits. This agrees with the general principle that the Wald test is typically more powerful than the GEE-based test.

The chi-square statistic  $((n-p-1)/n)Q$  is commonly used in practice and referred to an  $m$ -DF chi-square distribution to compute multitrait association test's  $p$  values, which can lead to significantly inflated type I errors at stringent genome-wide significance levels. Figure 1 shows the ratio of actual significance level of Wald test's  $p$  values computed using the chi-square distribution and  $F$ -distribution, respectively. We can see that the type I error based on the chi-square distribution is inflated: more so for larger number of traits, smaller significance level, and smaller sample size. For example, when testing  $m = 8$  traits with  $p = 2$  covariates and  $n = 500$  samples, under genome-wide significance level  $5 \times 10^{-8}$ , the

TABLE 1: Type I error of multitrait tests ( $m = 2, p_0 = 0.1$ ) divided by the nominal significance level  $\alpha$ . The MAFs of SNP are 0.1 and 0.2 in the two populations, respectively.  $Q$  is the  $m$ -DF omnibus Wald test;  $T$  and  $T'$  are the 1-DF Wald tests assuming a common or common scaled effect. ( $Q_s, T_s, T'_s$ ) are the corresponding GEE-based  $m$ -DF omnibus test and 1-DF tests assuming a common effect or common scaled effect.

$\alpha$	$\rho = 0$			$\rho = 0.5$		
	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-5}$	$10^{-4}$	$10^{-3}$
$Q_s$	0.69	0.79	0.89	0.67	0.79	0.89
$T_s$	0.74	0.85	0.93	0.71	0.83	0.92
$T'_s$	0.74	0.85	0.89	0.71	0.83	0.92
$Q$	1.04	1.00	1.00	1.03	1.01	1.00
$T$	0.98	0.99	1.01	0.97	0.99	1.00
$T'$	0.96	0.98	1.00	0.96	0.98	0.99

TABLE 2: Type I error divided by the nominal significance level  $\alpha$  for multitrait tests ( $m = 8, p_0 = 0.1$ ).

$\alpha$	$\rho = 0$			$\rho = 0.5$		
	$10^{-5}$	$10^{-4}$	$10^{-3}$	$10^{-5}$	$10^{-4}$	$10^{-3}$
$Q_s$	0.43	0.62	0.75	0.44	0.60	0.75
$T_s$	0.74	0.84	0.93	0.77	0.85	0.93
$T'_s$	0.74	0.84	0.93	0.78	0.85	0.93
$Q$	0.94	0.99	1.00	0.94	1.00	1.00
$T$	1.03	1.03	1.02	1.05	1.04	1.03
$T'$	1.03	1.03	1.03	1.03	0.99	0.99

TABLE 3: Power of multitrait tests for  $m = 2$  continuous traits ( $Y_1, Y_2$ ) under significance level  $\alpha = 10^{-4}$ . The MAFs of SNP are 0.1 and 0.2 in the two populations, respectively.  $Q$  is the  $m$ -DF omnibus Wald test;  $T$  and  $T'$  are the 1-DF Wald tests assuming common or common scaled effect. ( $Q_s, T_s, T'_s$ ) are the corresponding GEE-based  $m$ -DF omnibus test and 1-DF tests assuming a common effect or common scaled effect.  $\sigma_i$  is the standard error of  $Y_i$  and  $\gamma_i$  is the SNP coefficient,  $i = 1, 2$ . The highest powered tests are bold-faced.

$(\gamma_1, \gamma_2)$	$\left(\frac{\gamma_1}{\sigma_1}, \frac{\gamma_2}{\sigma_2}\right)$	$Q$	$T$	$T'$	$Q_s$	$T_s$	$T'_s$
$\rho = 0.5$							
(0.3, 0)	(0.21, 0)	<b>0.375</b>	0.001	0.024	0.334	0.001	0.019
(0.3, 0.1)	(0.21, 0.1)	<b>0.206</b>	0.047	0.146	0.177	0.039	0.126
(0.25, 0.18)	(0.18, 0.18)	0.180	0.221	<b>0.258</b>	0.154	0.194	0.233
(0.3, 0.25)	(0.21, 0.25)	0.523	0.617	<b>0.619</b>	0.476	0.573	0.582
(0.2, 0.2)	(0.14, 0.2)	0.179	<b>0.257</b>	0.215	0.154	0.23	0.193
(0.2, 0.25)	(0.14, 0.25)	0.410	<b>0.501</b>	0.369	0.367	0.466	0.333
(0.25, 0.25)	(0.18, 0.25)	0.449	<b>0.560</b>	0.492	0.403	0.521	0.455
(0, 0.25)	(0, 0.25)	<b>0.638</b>	0.278	0.052	0.59	0.247	0.040
(0, 0.3)	(0, 0.3)	<b>0.893</b>	0.525	0.121	0.865	0.477	0.093
(0.1, 0.25)	(0.07, 0.25)	0.465	<b>0.485</b>	0.372	0.418	0.448	0.330
(0.1, 0.3)	(0.07, 0.3)	<b>0.744</b>	0.726	0.590	0.700	0.688	0.534
(0.2, 0.3)	(0.14, 0.3)	0.845	<b>0.891</b>	0.842	0.810	0.870	0.810
$\rho = 0$							
(0.3, 0)	(0.21, 0)	<b>0.206</b>	0.026	0.063	0.178	0.020	0.051
(0.3, 0.1)	(0.21, 0.1)	0.316	0.249	<b>0.337</b>	0.278	0.215	0.304
(0.25, 0.18)	(0.18, 0.18)	0.419	0.510	<b>0.530</b>	0.376	0.471	0.494
(0.3, 0.25)	(0.21, 0.25)	0.830	0.891	<b>0.892</b>	0.796	0.868	0.870
(0.2, 0.2)	(0.14, 0.2)	0.375	<b>0.486</b>	0.462	0.333	0.449	0.427
(0.2, 0.25)	(0.14, 0.25)	0.631	<b>0.727</b>	0.677	0.584	0.692	0.636
(0.25, 0.25)	(0.18, 0.25)	0.734	<b>0.820</b>	0.801	0.690	0.792	0.771
(0, 0.25)	(0, 0.25)	<b>0.405</b>	0.249	0.134	0.36	0.217	0.107
(0, 0.3)	(0, 0.3)	<b>0.701</b>	0.485	0.29	0.657	0.437	0.235
(0.1, 0.25)	(0.07, 0.25)	<b>0.451</b>	0.385	0.165	0.406	0.356	0.140
(0.1, 0.3)	(0.07, 0.3)	<b>0.769</b>	0.639	0.301	0.728	0.605	0.257
(0.2, 0.3)	(0.14, 0.3)	0.700	<b>0.743</b>	0.545	0.655	0.713	0.500

TABLE 4: Power of multitrait tests for  $m = 8$  continuous traits under significance level  $\alpha = 10^{-4}$ . The MAFs of SNP are 0.1 and 0.2 in the two populations, respectively.  $Q$  is the  $m$ -DF omnibus Wald test;  $T$  and  $T'$  are the 1-DF Wald tests assuming common or common scaled effect. ( $Q_s, T_s, T'_s$ ) are the corresponding GEE-based  $m$ -DF omnibus test and 1-DF tests assuming a common effect or common scaled effect. The highest powered tests are bold-faced.

$(\gamma_1, \dots, \gamma_8)$	$Q$	$T$	$T'$	$Q_s$	$T_s$	$T'_s$
$\rho = 0.5$						
$\gamma_1 = 0.3, \gamma_{i>1} = 0$	<b>0.303</b>	0.001	0	0.229	0	0
$(.3, .2, .1, .05, 0, \dots, 0)$	<b>0.696</b>	0	0.008	0.599	0	0.005
$\gamma_1 = 0.2, \gamma_{i>1} = 0.15$	0.045	0.201	<b>0.220</b>	0.030	0.169	0.195
$\gamma_i = 0.15$	0.048	<b>0.237</b>	0.193	0.032	0.204	0.170
$\rho = 0$						
$\gamma_1 = 0.3, \gamma_{i>1} = 0$	<b>0.063</b>	0.001	0.004	0.043	0.001	0.002
$(.3, .2, .1, .05, 0, \dots, 0)$	<b>0.467</b>	0.156	0.224	0.372	0.102	0.152
$\gamma_1 = 0.2, \gamma_{i>1} = 0.15$	0.934	0.996	<b>0.997</b>	0.887	0.992	0.993
$\gamma_i = 0.15$	0.912	<b>0.995</b>	0.994	0.855	0.989	0.988

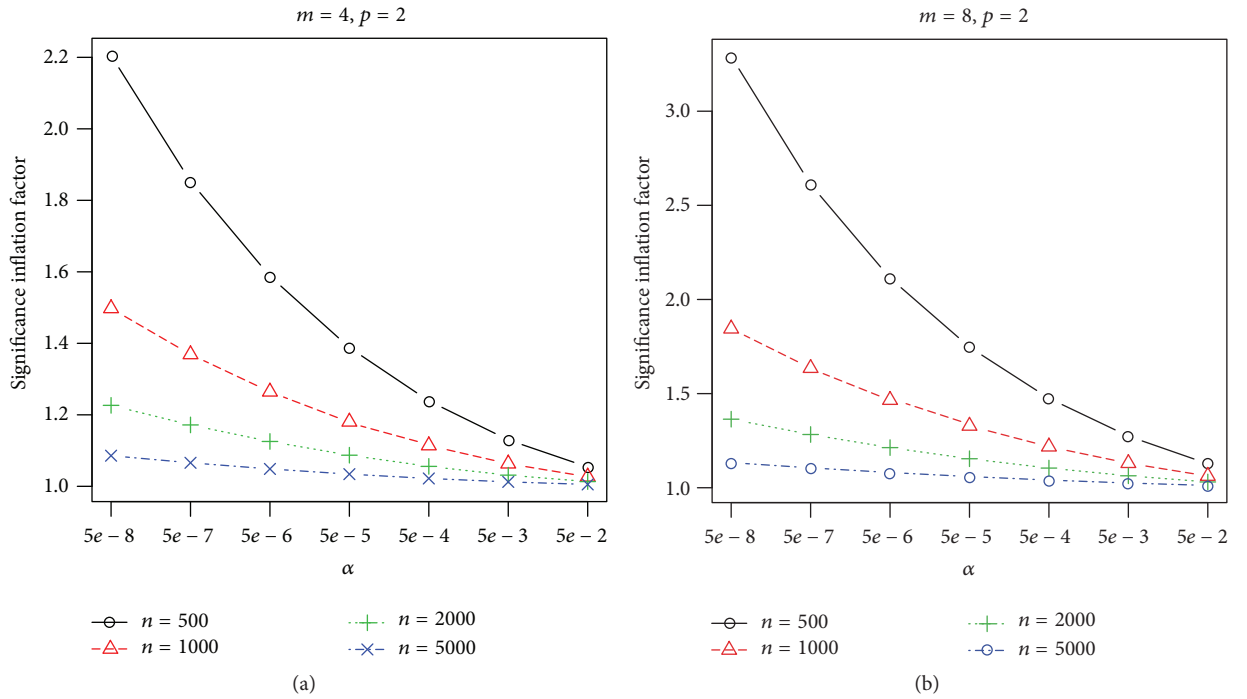


FIGURE 1: Ratio of the actual significance levels of  $m$ -DF chi-square test versus the  $F$ -test with  $(m, n - p - 1 - m)$  DFs. The  $x$ -axis is the type I error rate. (a) Shows the results for testing  $m = 4$  traits with  $p = 2$  covariates based on  $n$  individuals. (b) Shows the results for testing  $m = 8$  traits with  $p = 2$  covariates.

actual significance level of chi-square distribution  $p$  value is  $3.42 \times 5 \times 10^{-8} = 1.7 \times 10^{-7}$ . Using the chi-square distribution to compute  $p$  values will lead to very small inflation only when the sample size is large, such as in the meta-analysis of multiple GWAS studies. For typical GWAS with small-to-medium sample sizes, we recommend using the appropriate  $F$ -distribution to compute significance  $p$  values to reduce false positive findings.

**3.2. Application to ARIC GWAS of Glycemic Traits.** The Atherosclerosis Risk in Communities (ARIC) study [29] is a population-based, multicenter prospective investigation of

cardiovascular disease. Men and women aged 45–64 years at baseline were recruited from four US communities: Forsyth County, North Carolina; Jackson, Mississippi; suburban areas of Minneapolis, Minnesota; and Washington County, Maryland. A total of 15,792 individuals participated in the baseline examination during the period of 1987–1989. The vast majority of ARIC participants are of European (73%) or African (26%) ancestry. We conducted two analyses of diabetes-related glycemic traits in ARIC GWAS data, which has been imputed to around 2.5 million HapMap SNPs using MaCH [30]. We included in the analysis those common SNPs with  $MAF \geq 0.05$  and imputation score  $R^2 \geq 0.3$ .

TABLE 5: Six novel SNPs identified in the ARIC joint association test, which were not significant in the corresponding MAGIC consortium meta-analyses of fasting glucose (FG), fasting insulin (FI), and 2-hour fasting glucose (2hFG) but were significant in the MAGIC meta-analysis of fasting proinsulin (FP). We listed the ARIC joint test's  $p$  values (the proposed MLM Wald test and the GEE chi-square test) and the corresponding MAGIC consortium meta-analyses'  $p$  values for FG, FI, 2hFG, and FP.

SNP	Chr	bp	ARIC joint test's $p$ value		MAGIC meta-analysis' $p$ value			
			Wald	GEE	FG	FI	2hFG	FP
rs4502156	15	62383155	$5.4E-09$	$7.9E-09$	$8.4E-08$	$6.7E-01$	$8.2E-05$	$3.8E-11$
rs7163757	15	62391608	$1.4E-08$	$1.8E-08$	$4.2E-07$	$5.7E-01$	$1.9E-05$	$3.9E-11$
rs8037894	15	62394264	$1.2E-08$	$1.6E-08$	$4.1E-07$	$4.8E-01$	$3.5E-05$	$8.7E-11$
rs6494307	15	62394690	$1.7E-08$	$2.1E-08$	$3.3E-07$	$4.9E-01$	$2.7E-05$	$4.1E-11$
rs7167878	15	62396189	$1.7E-08$	$2.1E-08$	$4.6E-07$	$4.5E-01$	$2.4E-05$	$4.1E-11$
rs7172432	15	62396389	$1.7E-08$	$2.2E-08$	$6.5E-07$	$3.3E-01$	$1.9E-05$	$4.3E-11$

As a proof of concept, we first analyzed four fasting glucose levels in 5947 nondiabetic ARIC white participants measured at four visits (visits 1–4) conducted approximately three years apart. The average correlation of glucose levels is 0.55. We applied an additive genetic model with imputed dosage as a covariate and adjusted for age, gender, and study center in all tests. By analyzing four fasting glucose measures jointly,  $T'$  identified 104 significant SNPs,  $T$  identified 103,  $T'_s$  identified 102,  $T_s$  identified 101, and  $Q$  and  $Q_s$  identified the same set of 95 SNPs at the genome-wide significance level  $5 \times 10^{-8}$ . Analyzing each glucose measure separately identified 34, 84, 37, and 64 genome-wide significant SNPs at visits 1, 2, 3, and 4, respectively. All the identified SNPs by different methods are genome-wide significant in the MAGIC Consortium, a meta-analysis of 21 fasting glucose GWAS which together included 46,186 nondiabetic participants [31].

Compared to  $T'_s$ , the two additional SNPs identified by  $T'$ , rs780093 and rs780094, had  $p$  values of  $4.8 \times 10^{-8}$  and  $4.8 \times 10^{-8}$  using  $T'$ . Their respective MAGIC meta-analysis'  $p$  values were  $2.9 \times 10^{-13}$  and  $2.5 \times 10^{-12}$ . Compared to  $T_s$ , the two additional SNPs identified by  $T$ , rs1260326 and rs11688384, had  $p$  values of  $4.7 \times 10^{-8}$  and  $4.0 \times 10^{-8}$  using  $T$ . Their respective MAGIC meta-analysis'  $p$  values were  $4.3 \times 10^{-13}$  and  $4.1 \times 10^{-10}$ .

Second, we jointly analyzed three distinct diabetes-related glycemic traits measured at visit 4 in 5068 nondiabetic white participants measured at visit 4 in ARIC: fasting glucose, fasting insulin, and glucose level 2 hours after an oral glucose challenge. We applied an additive genetic model with imputed dosage as a covariate and adjusted for age, gender, and study center. To account for the skewed distribution of fasting insulin, we adopted the Box-Cox transformation with an estimated power of 0.35 [32]. The three traits had an average pairwise correlation of 0.31. When analyzing fasting insulin or 2-hour glucose levels individually, we did not identify any significant SNPs at the genome-wide significance level ( $5 \times 10^{-8}$ ). For joint testing of all three traits,  $T_s$ ,  $T'_s$ ,  $T$ ,  $T'$  identified none,  $Q_s$  identified 139, and  $Q$  identified 140 genome-wide significant SNPs, among which 61 and 61 SNPs were reported as genome-wide significant in the MAGIC meta-analyses of fasting glucose, fasting insulin, or 2-hour glucose levels [31, 33].

Compared to  $Q_s$ ,  $Q$  identified two additional genome-wide significant SNPs, rs4665987 and rs853780, with  $p$  values of  $4.9 \times 10^{-8}$  and  $4.9 \times 10^{-8}$ , respectively. MAGIC meta-analysis of fasting glucose reported a  $p$  value of  $2.1 \times 10^{-38}$  for rs853780. Its MAGIC meta-analyses of fasting insulin and 2-hour glucose  $p$  values are 0.054 and 0.477, respectively. For rs4665987 (near GCKR on chromosome 2:27755825), MAGIC meta-analysis'  $p$  values for the fasting glucose, fasting insulin, and 2-hour glucose levels are  $4.6 \times 10^{-6}$ , 0.04, and  $9.3 \times 10^{-5}$ , respectively. This SNP was genome-wide significantly associated with human serum metabolite levels in a GWAS of 8330 Finnish individuals [34] and several other GWAS [35–38]. Compared to  $Q$ ,  $Q_s$  reported one additional genome-wide significant SNP, rs17540154, with  $p$  value of  $4.3 \times 10^{-8}$ . The MAGIC meta-analysis of fasting glucose reported a  $p$  value of  $8.7 \times 10^{-38}$  for rs17540154. Its MAGIC meta-analyses of fasting insulin and 2-hour glucose  $p$  values are 0.101 and 0.720, respectively.

Among the identified significant SNPs by joint testing, there were 79 novel genome-wide significant SNPs that have not been reported as significantly associated with diabetes-related fasting glucose and insulin levels before. Among them, one SNP, rs4665987, is located on chromosome 2:27755825 and 78 other SNPs are clustered on chromosomes 15:62132921 to 15:62396389. Interestingly, six of them (listed in Table 5) were genome-wide significant in the MAGIC meta-analysis of proinsulin level [39]. The list of all identified SNPs with detailed analysis' results is available in the supplementary materials.

#### 4. Discussion

So far typical effect sizes of most identified genetic variants for many diseases or traits are very small and they have only explained a very small proportion of the overall disease heritability or trait variation. It is commonly accepted that there are many more common variants with relatively small-to-medium effect sizes or rare variants with larger effect sizes yet to be discovered. To identify these additional variants, very large sample sizes will be needed. One approach is to form a consortium to facilitate meta-analysis of many studies, but development of these genetics consortia is generally

time-consuming and logistically challenging. Meanwhile the recently studied joint association test of multiple correlated traits offers an alternative approach to boost power in that it can often dramatically improve the association test power by “enlarging the sample size” through the incorporation of many correlated traits that are typically collected in most large genetic studies and may share genetic determinants. Another strategy to further improve the detection power is to use a variant-set association test, which has been proven to be very useful (see, e.g., [16, 17, 40–42]). It is worthwhile to generalize the proposed Wald tests to develop more accurate and powerful association tests of variant sets across multiple traits.

Here we have focused on testing a relatively small number of correlated quantitative traits, which have enabled us to develop accurate and powerful association tests without any asymptotic approximations as adopted in the more general though conservative GEE approach, which can be applied to any mix of quantitative and discrete traits. It will be interesting to extend the proposed methods to the phenome-wide association studies (PheWAS) with a large collection of phenotypes [43–45] and develop more powerful joint association test of quantitative and discrete traits.

In the previous discussions, we have assumed the same set of covariates across all traits. With differing covariates, we provide technical details regarding model estimation and extensive simulation studies to confirm that the proposed methods accurately control type I errors and perform favorably compared to existing methods (see the supplementary materials for complete results). In summary, we recommend the proposed multivariate linear regression-based test as a complementary approach to enhancing the power of analyzing multiple quantitative traits in unrelated individuals. Our numerical studies have suggested that the omnibus Wald test generally has robust and good performance. The 1-DF Wald tests can perform well due to reduced DFs, but they could be sensitive to the underlying assumptions. It will be worthwhile to develop adaptive and powerful tests. We have implemented the proposed methods in an R package available at <http://www.github.com/baolinwu/MTAR>. We provide some sample R codes to install and use the package in the supplementary materials. The developed algorithms are very efficient and extremely scalable to genome-wide association test.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was supported in part by NIH Grants GM083345 and CA134848. The authors are grateful to the University of Minnesota Supercomputing Institute for assistance with the computations. The ARIC study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute Contracts (HHSN268201100005C,

HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute Contract U01HG004402; and National Institutes of Health Contract HHSN268200625226C. The authors thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant no. UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

## Supplementary Materials

In the supplementary materials, we provide more details for (1) model estimation and inference with numerical results when we have different covariates for each trait, (2) 1-DF Wald test assuming similar effects across all traits, (3) relation of proposed methods to CCA (e.g., as implemented in the PLINK), (4) the list of all identified SNPs with detailed analysis results for the ARIC GWAS of glycemic traits, (5) sample codes to install and use the developed R package, and (6) simulation results illustrating the robustness of proposed tests to deviation from normality and impact of directions of marginal effects and trait correlations on the multitrait association test power. (*Supplementary Materials*)

## References

- [1] N. Solovieff, C. Cotsapas, P. H. Lee, S. M. Purcell, and J. W. Smoller, “Pleiotropy in complex traits: challenges and strategies,” *Nature Reviews Genetics*, vol. 14, no. 7, pp. 483–495, 2013.
- [2] K. Wang and D. Abbott, “A principal components regression approach to multilocus genetic association studies,” *Genetic Epidemiology*, vol. 32, no. 2, pp. 108–118, 2008.
- [3] L. Klei, D. Luca, B. Devlin, and K. Roeder, “Pleiotropy and principal components of heritability combine to increase power for association analysis,” *Genetic Epidemiology*, vol. 32, no. 1, pp. 9–19, 2008.
- [4] M. A. R. Ferreira and S. M. Purcell, “A multivariate test of association,” *Bioinformatics*, vol. 25, no. 1, pp. 132–133, 2009.
- [5] J. Liu, Y. Pei, C. J. Papasian, and H.-W. Deng, “Bivariate association analyses for the mixture of continuous and binary traits with the use of extended generalized estimating equations,” *Genetic Epidemiology*, vol. 33, no. 3, pp. 217–227, 2009.
- [6] Q. Yang, H. Wu, C.-Y. Guo, and C. S. Fox, “Analyze multivariate phenotypes in genetic association studies by combining univariate association tests,” *Genetic Epidemiology*, vol. 34, no. 5, pp. 444–454, 2010.
- [7] L. J. Rasmussen-Torvik, A. Alonso, M. Li et al., “Impact of repeated measures and sample selection on genome-wide association studies of fasting glucose,” *Genetic Epidemiology*, vol. 34, no. 7, pp. 665–673, 2010.
- [8] P. F. O’Reilly, C. J. Hoggart, Y. Pomyen et al., “MultiPhen: Joint model of multiple phenotypes can increase discovery in GWAS,” *PLoS ONE*, vol. 7, no. 5, Article ID e34861, 2012.
- [9] M. Stephens, “A Unified Framework for Association Analysis with Multiple Related Phenotypes,” *PLoS ONE*, vol. 8, no. 7, Article ID e65245, 2013.

- [10] S. van der Sluis, D. Posthuma, and C. V. Dolan, "TATES: Efficient Multivariate Genotype-Phenotype Analysis for Genome-Wide Association Studies," *PLoS Genetics*, vol. 9, no. 1, Article ID e1003235, 2013.
- [11] Q. He, C. L. Avery, and D.-Y. Lin, "A general framework for association tests with multivariate traits in large-scale genomics studies," *Genetic Epidemiology*, vol. 37, no. 8, pp. 759–767, 2013.
- [12] B. Wu and J. S. Pankow, "Statistical Methods for Association Tests of Multiple Continuous Traits in Genome-Wide Association Studies," *Annals of Human Genetics*, vol. 79, no. 4, pp. 282–293, 2015.
- [13] C. M. Sitlani, K. M. Rice, T. Lumley et al., "Generalized estimating equations for genome-wide association studies using longitudinal phenotype data," *Statistics in Medicine*, vol. 34, no. 1, pp. 118–130, 2015.
- [14] S. K. Ganesh, D. I. Chasman, and M. G. Larson, "Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations," *The American Journal of Human Genetics*, vol. 95, no. 1, pp. 49–65, 2014.
- [15] T. E. Galesloot, K. Van Steen, L. A. L. M. Kiemeny, L. L. Janss, and S. H. Vermeulen, "A comparison of multivariate genome-wide association methods," *PLoS ONE*, vol. 9, no. 4, Article ID e95923, 2014.
- [16] C. S. Tang and M. A. R. Ferreira, "A gene-based test of association using canonical correlation analysis," *Bioinformatics*, vol. 28, no. 6, Article ID bts051, pp. 845–850, 2012.
- [17] J. A. Seoane, C. Campbell, I. N. M. Day, J. P. Casas, and T. R. Gaunt, "Canonical Correlation Analysis for Gene-Based Pleiotropy Discovery," *PLoS Computational Biology*, vol. 10, no. 10, 2014.
- [18] K. Y. Liang and S. L. Zeger, "Longitudinal data analysis using generalized linear models," *Biometrika*, vol. 73, no. 1, pp. 13–22, 1986.
- [19] C. L. Avery, Q. He, K. E. North et al., "A phenomics-based strategy identifies loci on APOC1, BRAP, and PLCG1 associated with metabolic syndrome phenotype domains," *PLoS Genetics*, vol. 7, no. 10, Article ID e1002322, 2011.
- [20] E. D. Schifano, L. Li, D. C. Christiani, and X. Lin, "Genome-wide association analysis for multiple continuous secondary phenotypes," *American Journal of Human Genetics*, vol. 92, no. 5, pp. 744–759, 2013.
- [21] K. Wang, "Testing genetic association by regressing genotype over multiple phenotypes," *PLoS ONE*, vol. 9, no. 9, Article ID e106918, 2014.
- [22] X. Guo, Y. Li, X. Ding, M. He, X. Wang, and H. Zhang, "Association tests of multiple phenotypes: ATeMP," *PLoS ONE*, vol. 10, no. 10, Article ID e0140348, 2015.
- [23] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, NY, USA, 3rd edition, 2003.
- [24] J. Fox, *Applied Regression Analysis and Generalized Linear Models*, SAGE Publications, Inc., Los Angeles, CA, USA, 2nd edition, 2008.
- [25] C. R. Rao, *Linear Methods of Statistical Induction and their Applications*, Wiley, New York, NY, USA, 2nd edition, 1973.
- [26] P. C. O'Brien, "Procedures for comparing samples with multiple endpoints," *Biometrics*, vol. 40, no. 4, pp. 1079–1087, 1984.
- [27] A. Voorman, K. Rice, and T. Lumley, "Fast computation for genome-wide association studies using boosted one-step statistics," *Bioinformatics*, vol. 28, no. 14, pp. 1818–1822, 2012.
- [28] S. Purcell, B. Neale, K. Todd-Brown et al., "PLINK: a tool set for whole-genome association and population-based linkage analyses," *American Journal of Human Genetics*, vol. 81, no. 3, pp. 559–575, 2007.
- [29] The ARIC Investigators, "The atherosclerosis risk in communities (aric) study: design and objectives," *American Journal of Epidemiology*, vol. 129, no. 4, pp. 687–702, 1989.
- [30] Y. Li, C. J. Willer, J. Ding, P. Scheet, and G. R. Abecasis, "MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes," *Genetic Epidemiology*, vol. 34, no. 8, pp. 816–834, 2010.
- [31] J. Dupuis, C. Langenberg, I. Prokopenko et al., "New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk," *Nature Genetics*, vol. 42, no. 2, pp. 105–116, 2010.
- [32] G. E. P. Box and D. R. Cox, "An analysis of transformations," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 26, pp. 211–252, 1964.
- [33] R. Saxena, M. F. Hivert, and C. Langenberg, "Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge," *Nature Genetics*, vol. 42, no. 2, pp. 142–148, 2010.
- [34] J. Kettunen, T. Tukiainen, A.-P. Sarin et al., "Genome-wide association study identifies multiple loci influencing human serum metabolite levels," *Nature Genetics*, vol. 44, no. 3, pp. 269–276, 2012.
- [35] W. Tang, S. Basu, X. Kong et al., "Genome-wide association study identifies novel loci for plasma levels of protein C: The ARIC study," *Blood*, vol. 116, no. 23, pp. 5032–5036, 2010.
- [36] M. J. Barber, L. M. Mangravite, C. L. Hyde et al., "Genome-Wide Association of Lipid-Lowering Response to Statins in Combined Study Populations," *PLoS ONE*, vol. 5, no. 3, Article ID e9763, 2010.
- [37] A. Tin, O. M. Woodward, W. H. L. Kao et al., "Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele," *Human Molecular Genetics*, vol. 20, no. 20, Article ID ddr307, pp. 4056–4068, 2011.
- [38] A. Kottgen, E. Albrecht, A. Teumer et al., "Genome-wide association analyses identify 18 new loci associated with serum urate concentrations," *Nature Genetics*, vol. 45, no. 2, pp. 145–154, 2013.
- [39] R. J. Strawbridge, J. Dupuis, and I. Prokopenko, "Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes," *Diabetes*, vol. 60, no. 10, pp. 2624–2634, 2011.
- [40] M. C. Wu, P. Kraft, M. P. Epstein et al., "Powerful snp-set analysis for case-control genome-wide association studies," *American Journal of Human Genetics*, vol. 86, no. 6, pp. 929–942, 2010.
- [41] M. C. Wu, S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin, "Rare-variant association testing for sequencing data with the sequence kernel association test," *American Journal of Human Genetics*, vol. 89, no. 1, pp. 82–93, 2011.
- [42] S. Lee, G. R. Abecasis, M. Boehnke, and X. Lin, "Rare-variant association analysis: study designs and statistical tests," *American Journal of Human Genetics*, vol. 95, no. 1, pp. 5–23, 2014.
- [43] S. A. Pendergrass, K. Brown-Gentry, S. M. Dudek et al., "The use of phenome-wide association studies (PheWAS) for exploration of novel genotype-phenotype relationships and pleiotropy discovery," *Genetic Epidemiology*, vol. 35, no. 5, pp. 410–422, 2011.



- [44] S. A. Pendergrass, K. Brown-Gentry, S. Dudek et al., “Phenome-Wide Association Study (PheWAS) for detection of pleiotropy within the Population Architecture using Genomics and Epidemiology (PAGE) network,” *PLoS Genetics*, vol. 9, no. 1, Article ID e1003087, 2013.
- [45] R. M. Cronin, J. R. Field, Y. Bradford et al., “Phenome-wide association studies demonstrating pleiotropy of genetic variants within FTO with and without adjustment for body mass index,” *Applied Genetic Epidemiology*, vol. 5, article 250, 2014.