Boston University

OpenBU **http://open.bu.edu**

Boston University Theses & Dissertations Boston University Theses & Dissertations

2013

Multiple phenotype modeling in pleiotropic effect studies of quantitative trait loci

https://hdl.handle.net/2144/13154

Downloaded from DSpace Repository, DSpace Institution's institutional repository

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

MULTIPLE PHENOTYPE MODELING IN PLEIOTROPIC EFFECT STUDIES OF QUANTITATIVE TRAIT LOCI

by

QIONG LOUIE-GAO

B.S., Central China Normal University, 2002
M.S., University of Toledo, 2007

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2013

© Copyright by
QIONG LOUIE-GAO
2013

Approved by

First Reader ______________________________

Serkalem Demissie, Ph.D.
Associate Professor of Biostatistics

Second Reader ______________________________

L. Adrienne Cupples, Ph.D.
Professor of Biostatistics

Third Reader ______________________________

Howard Cabral, Ph.D.
Professor of Biostatistics

ACKNOWLEDGMENTS

First, I wish to express my deepest gratitude to my advisor, Dr. Serkalem Demissie. Her patience, motivation, knowledge and enthusiasm have helped me throughout my Ph.D study and research at Boston University. Without her guidance, support and encouragement, I could not have been able to complete this thesis.

My sincere thanks also go to my committee members, Dr. L. Adrienne Cupples, Dr. Howard Cabral, Dr. Gheorghe Doros, and Dr. Qiong Yang for joining my committee, reviewing my thesis, and providing valuable comments.

A special thank you to my best friends, Yanyan, Vivian, Na and Siyan, and my colleagues Yanhua Zhou and Dr. Ching-Ti Liu, for all their help and support.

Last but not least, I want to share the joy with my parents, Daoxiong and Aixian, and my husband, Anderson. Thanks to them for giving me generous love, care, encourage and support. In addition, thanks to God for bringing me an angelic little boy, Nathan, who makes me appreciate life more every day.

MULTIPLE PHENOTYPE MODELING IN PLEIOTROPIC EFFECT STUDIES OF QUANTITATIVE TRAIT LOCI

(Order No.)

QIONG LOUIE-GAO

Boston University Graduate School of Arts and Sciences, 2013

Major Professor: Serkalem Demissie, Associate Professor of Biostatistics

ABSTRACT

Pleiotropy refers to the shared effects of a gene or genes on multiple phenotypes, a major reason for genetic correlation between phenotypes. For example, for osteoporosis, bone mineral densities at different skeletal sites may share common genetic factors; thus, examining the shared effects of genes may enable more effective fracture treatments. To date, methods are not available for estimating and testing the pleiotropic effects of single nucleotide polymorphisms (SNPs) in genetic association studies. In this dissertation, we explore two types of methods to evaluate the SNP-specific pleiotropic effect based on multivariate techniques. First, we propose two approaches based on variance components (VC) analysis for family-based studies, which quantify and test the pleiotropic effect by examining the contribution of specific genetic marker(s) to polygenic correlation or covariance of traits. Second, we propose a multivariate linear regression approach for population-based studies with samples of families or unrelated subjects. This method partitions the specific effect of the marker(s) from phenotypic covariance. We evaluate the performance of our proposed methods in simulation studies, compare them to existing

multivariate analysis methods and illustrate their application using real data to assess candidate SNPs for osteoporosis-related phenotypes in the Framingham Osteoporosis Study. In contrast to existing methods, our newly proposed approaches allow the quantification of pleiotropic effects. The bootstrap resampling percentile method is used to construct confidence intervals for statistical hypothesis testing. Simulation results suggest that the VC-based approaches are affected by the polygenic correlation level. The covariance analysis approach outperforms the VC-based approaches, with unbiased estimates and better power, which remain consistent regardless of the polygenic correlation. In addition, the covariance analysis approach is simple to implement and can be applied to both family data and genetically unrelated data. Using simulation, we also show that existing methods, such as MANOVA, can have high rejection rates when a SNP has a large effect on a single trait, which prevent us from using them for pleiotropic effect analysis. In summary, this dissertation introduces promising new approaches in multiple phenotypic models for SNP-specific pleiotropic effect.

TABLE OF CONTENTS

Chapter 1: Introduction 1

1.1 Motivation and Background 1

1.2 Purpose of the Study 3

1.3 Definitions of Symbols 6

Chapter 2: Literature Review 9

2.1 Generalized Estimating Equations 10

2.2 Multivariate Analysis of Variance 12

2.3 Principal Components Analysis 13

2.4 Conditional Analysis 15

2.5 Structural Equation Modeling 16

2.6 O'Brien Method 19

2.7 Summary 20

Chapter 3: Variance Components Analysis-Based Approaches on Pleiotropic Effects 23

3.1 Introduction 23

3.2 Background 25

3.2.1 QTL Model for a Univariate Trait 26

3.2.2 QTL Model for Multivariate Traits 27

3.3 Analysis of Pleiotropic Effect Using Genetic Correlation 28

3.3.1 Proposed Method 30

3.3.2 Evaluation of the Proposed Method Using Simulation Studies 32

3.3.3 Simulation Results 38

3.4 Analysis of Pleiotropic Effect Using Genetic Covariance 42

3.4.1 Proposed Method 42

3.4.2 Evaluation of the Proposed Method Using Simulation 44

3.4.3 Simulation Results 47

3.5 Comparison between Polygenic Genetic Correlation and Genetic Covariance Approaches 51

3.6 Summary 64

Chapter 4: Covariance Analysis on Pleiotropic Effect 65

4.1 Introduction 65

4.2 Single SNP Analysis 66

4.2.1 Proposed Method 66

4.2.2 Evaluation of the Proposed Method Using Simulation Studies 68

4.2.3 Simulation Results 71

4.3 Multiple SNPs Analysis 93

4.3.1 Proposed Method 93

4.3.2 Evaluation of the Proposed Method Using Simulation Studies....................... 95

4.3.3 Simulation Results 97

4.4 Summary 102

Chapter 5: Compare Methods in the Context of Pleiotropic Effects 104

5.1 Comparison of Covariance Analysis, Multivariate Analysis Of Variance and Conditional Analysis Approaches Using Unrelated Data 104

5.2 Comparison of Covariance Analysis, Variance Components Analysis-Based Approaches and Multivariate Analysis Of Variance using Family Data 113

5.3 Summary 120

Chapter 6: Application to Real Data 122

6.1 Introduction 122

6.2 The Framingham Osteoporosis Study (FOS) 123

6.3 Pleiotropic Effect Using Unrelated Data 124

6.4 Pleiotropic Effect Using Family Data 130

6.5 Summary 136

Chapter 7: Discussion and Conclusions 137

7.1 Discussion 137

7.1.1 The Genetic Correlation and Genetic Covariance Approaches 137

7.1.2 Covariance Analysis 138

7.1.3 Comparison of Approaches 139

7.1.4 Application to Real Data 140

7.2 Limitation and Future Work 141

7.3 Conclusions 142

APPENDIX 143

REFERENCES 170

CURRICULUM VITAE 180

LIST OF TABLES

Table 3. 1 Pairs of SNP effects on bivariate traits in simulations of VC analysis 33

Table 3. 2 Empirical bootstrap distribution comparing the genetic correlation and covariance approaches 53

Table 3. 3 Simulation distribution and empirical bootstrap distribution comparison for the genetic correlation and covariance approaches.. 56

Table 3. 4 Simulation distribution comparing the genetic correlation and covariance approaches under different polygenic correlation levels 59

Table 4. 1 Pairs of SNP effects on bivariate traits in simulations of CovA...................... 69

Table 4. 2 Simulation distribution and empirical bootstrap distribution comparison for estimate and variance of pleiotropic effect in CovA 84

Table 4. 3 Empirical type I error and power of single SNP pleiotropic effect comparing normal approximation approach and bootstrapping confidence interval approaches 88

Table 4. 4 Empirical type I error and power results comparing different numbers of bootstrapping samples in CovA 90

Table 4. 5 Pairs of two SNPs effects on bivariate traits in simulations of CovA 96

Table 4. 6 Empirical type I error and power of two SNPs pleiotropic effects comparing normal approximation approach and bootstrapping confidence interval approaches 100

Table 4. 7 Power comparing single SNP pleiotropic effect and two SNPs pleiotropic effects using bootstrapping PCTL method. 102

Table 5. 1 Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, GEE and conditional analysis using unrelated data 109
Table 5. 2 Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, genetic correlation and covariance approaches and MANOVA using unrelated data 115
Table 5. 3 Simulation distribution comparing CovA, and genetic correlation and covariance approaches for two SNPs pleiotropic effects with the same or opposite direction 117
Table 6. 1 SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA for unrelated subjects in Framingham Osteoporosis-BMDs study. 127
Table 6. 2 SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study 133
Table S 1 Simulation distribution and empirical bootstrap distribution comparison for the genetic correlation and covariance approaches... 143
Table S 2 Simulation distribution comparing the genetic correlation and covariance approaches under different environmental correlation levels 145
Table S 3 Simulation distribution comparing the genetic correlation and covariance approaches under different MAFs 148
Table S 4 Estimate, percent bias, empirical type I error and power comparing different polygenic and environmental correlations for a SNP Pleiotropic Effect in CovA 149

Table S 5 Estimates, percent biased, empirical type I error and power comparing different MAF for a SNP Pleiotropic Effect in CovA 152

Table S 6 Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data................................ 153

Table S 7 SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study 161

LIST OF FIGURES

Figure 2. 1 Example of path diagram and structural equation modeling. 18

Figure 3. 1 True and estimated values of the difference between polygenic correlations from bivariate polygenic models with and without adjustment for a SNP 39

Figure 3. 2 True and estimated values of the difference between polygenic covariances from bivariate polygenic models with and without adjustment for a SNP. 48

Figure 4. 1 Pleiotropic SNP effect and individual SNP effects on traits using CovA 72

Figure 4. 2 Histogram examining the distribution of test statistic in CovA using 1000 simulation samples 74

Figure 4. 3 Q-Q plot comparing the distribution of test statistic in CovA using 1000 simulation samples to a standard normal distribution.. 76

Figure 4. 4 Q-Q plot comparing the distribution of test statistic in CovA using 1000 bootstrap re-samplings method in 1000 simulation samples to a standard normal distribution 82

Figure 4. 5 Two SNPs joint pleiotropic effects and individual pleiotropic effects using CovA. 98

Figure 6. 1 Estimated SNP-specific pleiotropic effect, and SNP individual effects on FNBMD and LSBMD using CovA for unrelated subjects in Framingham Osteoporosis-BMDs study. 126

Figure 6. 2 Estimated SNP-specific pleiotropic effect, and SNP individual effects on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation

approaches for family related subjects in Framingham Osteoporosis-BMDs study.
.... 132

LIST OF ABBREVIATIONS

BC	Bias Corrected
BCa	Bias Corrected accelerated
BMD	Bone Mineral Density
CIs	Confidence Intervals
CovA	Covariance Analysis
CSA	Cross Sectional Area
FHS	Framingham Heart Study
FIML	Full Information Maximum Likelihood
FNBMD	BMD at the femoral neck
FOS	Framingham Osteoporosis Study
GEE	Generalized Estimating equations
GLMs	Generalized Linear Models
GWAS	Genome-wide Association Studies
LD	Linkage Disequilibrium
LE	Linkage Equilibrium
LRT	Likelihood ratio test
LSBMD	BMD at the lumbar spine
MAF	Minor Allele Frequency
MANOVA	Multivariate Analysis of Variance
PC	Principal Components

PCTL	Percentile
QTL	Quantitative Trait Locus
SAS	Statistical Analysis System
SEM	Structural Equation Modeling
SHARe	SNP Health Association Resource
SNP	Single Nucleotide Polymorphism
SOLAR	Sequential Oligogenic Linkage Analysis Routines
VC	Variance Components

Chapter 1: Introduction

1.1 Motivation and Background

Pleiotropy, the shared effects of a gene or genes on multiple phenotypic traits, is a common focus of genetic studies. For example, osteoporosis is a common disease characterized by low bone mass and loss of bone tissue that may lead to increased susceptibility to fracture, particularly in the hip, spine and wrist (Kanis 2002). According to the 2004 Surgeon General's Report, osteoporosis has become the most important public health threat in the United States. Ten million Americans over 50 years of age are estimated to have osteoporosis and another 34 million are estimated to have reduced bone mass which puts them at increased risk of fractures (Blake and Fogelman 2007). Previous studies use bone mineral density (BMD) as a proxy for osteoporosis because a low BMD is known to contribute to increased risk of bone fragility (Kanis et al. 2006). The position statement of the International Society for Clinical Densitometry (ISCD) suggests that a combination of BMD assessments in more than one region of the body could improve risk assessment in clinic practice (Lewiecki et al. 2004).

In humans, BMD is a highly heritable phenotype (Eisman 1999; Ralston 2002). For example, the heritability of BMD at the spine and hip is estimated to range from 70% to 85% (Ralston 2002). Several studies examine the relationship between genetic factors and osteoporosis related phenotypes (Deng et al. 1999; Livshits et al. 1999; Deng et al. 2002; Ralston 2002; Karasik et al. 2002; Karasik et al. 2010). Some of these studies have found that BMD at different skeletal sites may share common genetic factors (Deng et al.

1999; Livshits et al. 1999; Deng et al. 2002; Karasil et al. 2002). Examining the shared effects of genes between osteoporosis-related traits will result in a better understanding of the mechanism of risk and may make a great contribution to more effective fracture treatments.

Pleiotropic effects are the main causes for the existence of a genetic correlation between traits (Falconer and Mackay 1996). We are concerned with what factors and to what extent result in sets of correlated traits. Thus, "Does a quantitative trait locus (QTL) have shared effects on multiple traits?" is an interesting question and is particularly important in genetics research, where QTL refers to a region of a chromosome within which one or more genes contributes to the variation observed at a quantitative trait (Gardner and Latta 2007). A comprehensive understanding of pleiotropy may contribute to disease treatment and medicine development. Davignon (2004) pointed out that pleiotropic effects may relate to the primary mechanism of a drug whose action could be undesirable, neutral or beneficial. For example, "pleiotropic effects of statins include improvement of endothelial dysfunction, increased nitric oxide bioavailability, antioxidant properties, inhibition of inflammatory responses, and stabilization of atherosclerotic plaques. These and several other emergent properties could act in concert with the potent low-density lipoprotein cholesterol-lowering effects of statins to exert early as well as lasting cardiovascular protective effects. Understanding the pleiotropic effects of statins is important to optimize their use in treatment and prevention of cardiovascular disease" (Davignon 2004).

A QTL with pleiotropic effects may be detected by using multivariate statistical analysis, which deals with multiple dependent variables and exploits some inherent interdependent information from genetic and phenotypic correlations between traits (Korol et al. 1995; Knott and Haley 2000; Gilbert and Le Roy 2003). Multivariate analysis provides potentially more insight into the nature of the genetic correlations between different traits. Jiang and Zeng (1995) suggest that analyzing multiple traits simultaneously by fitting a pleiotropic QTL in multivariate approaches would increase power if the true model examines pleiotropy.

1.2 Purpose of the Study

Earlier studies have put a great deal of effort into multivariate methods for quantitative variables, such as Multivariate Analysis of Variance (MANOVA) and Generalized Estimating Equations (GEE); however, their utilities of applying to pleiotropic effect analysis is not always clear. For example, in statistical hypothesis testing of a single pleiotropic QTL effect on two correlated traits (Y_1 and Y_2), one can apply bivariate linear models

$$Y_1 = \mu_1 + \beta_1 QTL + e_1; \quad (1.1)$$

$$Y_2 = \mu_2 + \beta_2 QTL + e_2,$$

where μ_i is the mean if trait i; β_i is the effect of the QTL on the ith trait; e_i is a residual error for trait i, $E(e_i) = 0$, $Cov(e_i, e_j) = \sigma_{ij}$, $i, j = 1,2$. The hypothesis is to determine if

a QTL has shared effects on multiple traits. Symbolically, the hypotheses can be expressed as

$$H_0\colon \beta_1 = 0 \ or\ \beta_2 = 0 \ \ vs\ \ H_1\colon \beta_1 \neq 0 \ and\ \beta_2 \neq 0$$

The null hypothesis is compound, including three cases:

$$\beta_1 = 0 \ and\ \beta_2 = 0;$$

$$\beta_1 = 0 \ and\ \beta_2 \neq 0;$$

$$\beta_1 \neq 0 \ and\ \beta_2 = 0.$$

We reject the null hypothesis if QTL has shared effects on all traits. However, the hypotheses of some existing methods are

$$H_0\colon \beta_1 = 0 \ and\ \beta_2 = 0 \ \ vs\ \ H_1\colon \beta_1 \neq 0 \ or\ \beta_2 \neq 0$$

With these methods, it is impossible to judge if the rejection comes from the real pleiotropic effect or a strong effect on a single trait only. If any association between trait(s) and gene(s) is very strong, these methods cannot conclude that there is pleiotropy because the possibility that the gene may only affect one of the traits.

Therefore, the purpose of this dissertation is to develop effective statistical approaches to evaluate pleiotropic effects based on multivariate techniques, as presented in the following chapters:

Chapter 2 provides in-depth descriptions of five current methods of multivariate analysis. We discuss their pros and cons and probe their underlying problems in pleiotropic effect analysis.

Chapter 3 investigates variance components analysis of shared genetic and environmental correlations in bivariate polygenic models. We provide insight into marker specific pleiotropic effects by partitioning the specific effect of the marker(s) from the polygenic genetic variance components. We then propose polygenic genetic correlation and genetic covariance analyses approaches for family-based studies by examining the contribution of specific genetic marker(s) to polygenic correlation and covariance of traits. We test and quantify the pleiotropic effect of a marker by comparing polygenic correlations or covariances from polygenic models with and without adjustment for the marker. We evaluate the performance of these approaches by using simulation studies with respect to bias, power and type I error with family data.

Chapter 4 investigates multivariate linear regression models and explores marker(s) specific pleiotropic effects by partitioning the contribution of specific genetic marker(s) from the phenotypic covariance. We develop a covariance analysis (CovA) which can be used for population-based studies with samples of families or unrelated subjects. We test and quantify the pleiotropic effects of specific marker(s) by using the regression coefficients from multivariate analysis and the covariance between marker(s). We assess bias, type I error and power of CovA by using simulation studies with unrelated subjects.

Chapter 5 compares our newly proposed approaches with other multivariate statistical analysis methods for pleiotropic effects analysis, focusing on bias, power and Type I error for family data and unrelated data.

Chapter 6 applies our proposed methods to assess candidate SNPs that are likely to contribute shared effects in relation to osteoporosis-related phenotypes as a real data example. In order to evaluate the pleiotropic effect of 38 SNPs which are associated at $p < .001$ with both FN and LS BMD (Karasik et al. 2010), we apply (1) CovA to a sample of genetically unrelated subjects and (2) CovA, genetic correlation and covariance approaches to a sample of family related subjects.

Chapter 7 discusses the overall research results, limitations and future work.

1.3 Definitions of Symbols

Symbols employed in this study are defined below:

ρ_P---- A phenotypic correlation between traits, which shows how strongly traits are related (Searle 1961). It can be partitioned into genetic components and environmental components (Mahaney et al. 1995).

ρ_g---- A polygenic genetic correlation between traits, which represents the extent of any genetic influences on the traits (Neale and Maes 1996).

ρ_{g^*}---- A residual polygenic correlation between traits, which represents the extent of shared residual additive genetic influences on the traits which excludes the genetic effects contributed by the major gene.

ρ_e---- An environmental correlation between traits, which represents the extent of shared environmental influences on the traits.

ρ_{qtl}---- The additive genetic correlation between the traits due to the effects of the QTL, which is a measure of shared major gene effects near the region.

ρ_{g_diff}---- A difference of the polygenic correlation in the model without adjustment for the SNP effect to the residual polygenic correlation in the model with adjustment for the SNP effect.

h_{ri}^2---- Heritability for trait *i*, which is the proportion of phenotypic variance due to genetic factors for trait *i*.

$h_{ri^*}^2$---- Residual heritability for trait *i*, which is the fraction of phenotypic variance due to genetic factors after the QTL effect has been accounted for.

h_{qi}^2---- QTL heritability for trait *i*, which is the proportion of variance attributable to the QTL effect.

cov_P---- Phenotypic covariance between traits.

cov_g---- A polygenic genetic covariance between traits, which represents the extent of any genetic influences on the traits.

cov_{g^*}---- A residual polygenic covariance between traits, which represents the extent of shared residual additive genetic influences on the traits which excludes the genetic effects contributed by the major gene.

cov_e---- Environmental covariance between traits, which represents the extent of shared environmental influences on the traits.

cov_{qtl}---- The genetic covariance between the traits due to the effects of the QTL, which is a measure of shared major gene effects near the region.

cov_{g_diff}---- A difference of the polygenic covariance in the model without adjustment for the SNP effect to the residual polygenic covariance in the model with adjustment for the SNP effect.

C_P---- Pleiotropic effect(s). A single gene or genes having effects on more than one trait.

Chapter 2: Literature Review

Multivariate analysis is a statistical approach for dealing with more than one response variable. These variables might correlate with each other; hence, we need to take into account their statistical dependence. In contrast to univariate analysis which has only one dependent variable, multivariate analysis has more response variables, resulting in a structure of several regression equations. Multivariate analysis has a number of advantages. First, since more information is analyzed simultaneously, a multivariate model may provide more significant results (Stearns et al. 2005), indicating greater power of the test (Schork et al. 1994; Korol et al. 2001). Second, it can increase precision of the parameter estimation (Jiang and Zeng 1995; Korol et al. 1995; Gilbert and Le Roy 2004; Stearns et al. 2005). Third and most important to genetic research, multivariate analysis can improve the detection of potential QTLs whose effects are too small to be found in single-trait analyses, and it can allow testing hypotheses involving multiple traits. These facilitate the investigation of genetic mechanisms such as pleiotropy and multiple linked QTLs (Jiang and Zeng 1995; Mangin et al. 1998; Williams et al. 1999; Gerbens et al. 2001; Grindflek et al. 2001; Ovilo et al. 2002; Stearns et al. 2005).

Existing literature and research have developed a number of multivariate methods to simultaneously analyze the traits of interest. In this chapter, we present brief reviews of the literature for five commonly used statistical methods: generalized estimating equations (GEE); multivariate analysis of variance (MANOVA); principal components (PC) analysis; structural equation modeling (SEM) and conditional analysis.

2.1 Generalized Estimating Equations

Liang and Zeger proposed the method of Generalized Estimating Equations (GEEs) in 1986 (Liang and Zeger, 1986). This approach is typically used in longitudinal (Zorn 2001; Ballinger 2004) and clustered (Stoner 2002) studies that primarily focus on a single trait with multiple observations, such as a single outcome measured repeatedly on the same subject. GEE uses quasi-likelihood for modeling correlated responses which could be either continuous or discrete. The GEE approach offers many advantages to researchers who are interested in modeling correlated data (Zorn 2001). First, the outcome variable could take a wide range of forms in data collection. Second, the GEE models allows for substantial flexibility in specifying the working correlation structures to account for the within-subject correlations and offer the potential for valuable substantive insight into the nature of that correlation. In addition, even if a correlation structure is misspecified, the estimation of population-averaged regression coefficients is still consistent under certain regularity conditions. Moreover, GEE models are available in many current software packages and the result is easy to interpret, similar to the commonly used models for uncorrelated data.

Take osteoporosis as an example, the multiple measurements from the vertebral levels of the same patient tend to be correlated. Let Y_j be the vertebral measurement of BMD at the jth level ($j = 1, \cdots, p$). In addition to being influenced by environmental effects, the measurements are governed by some genetic factors. If the genotypes of a major genetic marker are observable, we incorporate this marker into the GEE model and test if the observed genetic marker is associated with the observed traits. A special data

structure is needed for multiple measurements. First, we put values for trait variables for each subject into one column and name it as Y. Second, we create a dummy variable *TRAIT*. An unstructured correlation matrix is specified for each subject. GEE regression model can be expressed as

$$Y_j = \beta_0 + \beta_1 SNP + \beta_2 TRAIT_j + \beta_3 (SNP * TRAIT_j) + \beta_4 X_1 \cdots + \beta_{k+3} X_k + e_j \ , \tag{2.2}$$

where SNP (single nucleotide polymorphism, a most common type of sequence variation) denotes the genetic marker with observed genotypes which has a fixed effect on Y_j; $X_1, \cdots, X_k$ are non-genetic k predictor variables or covariates; $\beta_0, \cdots, \beta_{k+1}$ are regression coefficients; e_j is a random error, $E(e_j) = 0$, $Cov(e_i, e_j) = \ \sigma_{ij}, i, j = 1, \cdots, p$.

The hypothesis of model (2.2) in GEE is to test the main effect and interaction effect, assessing the association between multiple BMD values and the SNP.

$$H_0: \beta_1 = \beta_3 = 0 \quad vs \quad H_1: Not\ H_0.$$

It is a chi-square test with 2 degrees of freedom.

In multiple observations of the same trait, pleiotropy is defined as the shared effects of a gene or genes on multiple measurements. GEE leaves open the possibility of rejection the null hypothesis when the association between the SNP and any BMD measurement(s) is very strong. Therefore, rejecting of H_0 does not necessarily imply pleiotropy.

2.2 Multivariate Analysis of Variance

Multivariate Analysis of Variance (MANOVA) is another traditional method of a population-based multivariate approach. It is a generalized form of analysis of variance (ANOVA), considering situations where multiple outcomes are measured one or more times on the same subject. Each outcome represents the measurement of a different characteristic or repeated measures on the same trait.

In osteoporosis study, BMD and cross sectional area (CSA) are two correlated osteoporosis-related phenotypes. Let Y_j denote the value of jth trait (j=1,2). If we can observe the genotypes of a major genetic marker, the multivariate regression models incorporate it as a fixed effect.

$$Y_1 = \beta_{10} + \beta_{11}SNP + \beta_{12}X_1 + \cdots + \beta_{1,k+1}X_k + e_1, \qquad (2.3)$$

$$Y_2 = \beta_{20} + \beta_{21}SNP + \beta_{22}X_1 + \cdots + \beta_{2,k+1}X_k + e_2,$$

where SNP denotes the genetic marker with observed genotypes; $X_1, \cdots, X_k$ are non-genetic k predictor variables or covariates; $\beta_{i,0}, \cdots, \beta_{i,k+1}$ are regression coefficients; e_i is a random error, $E(e_i) = 0$, $Cov(e_i, e_j) = \sigma_{ij}$, $i, j = 1,2$. The hypotheses in MANOVA are:

$$H_0: \beta_{11} = \beta_{21} = 0 \quad vs \quad H_1: \beta_{11} \neq 0 \; or \; \beta_{21} \neq 0.$$

We reject the null hypothesis if β_{11} or β_{21} , or both β_{11} and β_{21}, does not equal to zero. In MANOVA, we use a multivariate F value (Wilks' λ) rather than a univariate F value (French et al. 2006), and the Hotelling's T^2 test in MANOVA is analogous to the univariate t test in ANOVA (Carey 1998).

It is noteworthy that Chidambaram concluded there are two advantages of MANOVA compared to multiple ANOVAs. It accounts for multiple testing and thus does not inflate the type I error rate, and it also accounts for the relationship between dependent variables.

However, the results of a MANOVA simply measure the degree of relationship between a set of traits and a single SNP. It does not tell that how many traits and which traits are influenced by this SNP. In addition, similar to GEE, a significant MANOVA result does not necessarily imply pleiotropy.

2.3 Principal Components Analysis

Principal components (PCs) analysis describes the variation in a set of multivariate data in terms of a group of uncorrelated variables (Pearson 1901; Hotelling 1933). Its basic idea is to explain variation within and covariation between a large numbers of observed variables with a smaller numbers of principal components. PCs analysis reduces dimensionality by rejecting low variance features and then performs a univariate analysis of a synthetic linear combination of the outcomes in which most of the information from the data is summarized (Gilbert and Roy 2003). Weller et al. (1996) recommended applying PCs analysis on multi-trait detection of QTLs. First, conduct PCs analysis on multiple quantitative traits. PCs are calculated from the correlation matrix of the traits where data all individuals are assumed to be independent. Suppose PC_t is the *t*th synthetic combination of the traits, named the *t*th PC. One can apply a linear regression

model to estimate and test for the association between an observed major genetic marker and the PCs of phenotypes rather than the individual phenotypes.

$$PC_t = \beta_{t0} + \beta_{t1}SNP + \beta_{t2}X_1 \cdots + \beta_{t,k+1}X_k + e_t \; , \qquad (2.4)$$

where SNP denotes the genetic marker with observed genotypes; $X_1, \cdots, X_k$ are non-genetic k predictor variables or covariates; $\beta_{t0}, \cdots, \beta_{t,k+1}$ are regression coefficients for the *t*th PC; e_t is a random error which is assumed to be distributed as $N(0, \sigma_{e_t}^2)$. The statistical testing is to assess the association between each combined trait represented by PC_t and the SNP. The hypotheses of the *t*th PC can be expressed as

$$H_0: \beta_{t1} = 0 \quad vs \quad H_1: \; \beta_{t1} \neq 0.$$

Wu (2009) proposed two extensions to the variable reduction approach: canonical correlation analysis (CCA) using sample splitting and principal component of QTL heritability (PCQH) using sample splitting, which are similar in spirit to principal components of phenotypes. The samples are split into two subsets without “crossing”, one for obtaining the weight vector and the other one for association testing.

Stearns et al. (2005) pointed out there is a clear gain of power with the variable reduction techniques over univariate models to detect loci on the original traits since fewer tests need to be performed. The disadvantage of using variable reduction in QTL analysis is that in terms of traits, the interpretation of the estimated effects might be not easy (Stearns et al. 2005; Koshkoih 2006). Because the combined variables are composed of a specific partition of the phenotypic covariance, they may not reflect the associated QTL covariance (Gilbert and Le Roy 2003; Stearns et al. 2005). Even if there is an

association between the QTL and a combined variable, it is ambiguous how many traits or which traits share the pleiotropic effects.

2.4 Conditional Analysis

Because of the pleiotropic effects of genes, there exists some degree of correlation among multiple quantitative traits. QTL changes on one trait might result in simultaneous responses on other related traits (Li et al. 2006). Conditional analysis uses one trait (Y_2) as a covariate in the analysis of another trait (Y_1), and then the contribution of Y_2 can be excluded from the variation of Y_1 (Zhu 1995; Liu et al. 2008). The remaining variation of Y_1, defined as conditional variation, indicates the extra effect of genes that are independent of Y_2 (Atchley and Zhu 1997; Liu et al. 2008).

The multiple linear regression model used in conditional analysis is

$$Y_1 = \beta_0 + \beta_1 SNP + \beta_2 Y_2 + \beta_3 X_1 + \cdots + \beta_{k+2} X_k + e, \tag{2.5}$$

where Y_1 denotes the trait as dependent variable and Y_2 denotes a second trait as a covariate. SNP denotes the genetic marker with observed genotypes and $X_1, \cdots, X_k$ are non-genetic k predictor variables or covariates. β_1 is the effect of the SNP of interest, adjusted for Y_2 and other covariates; β_2 is the effect of Y_2 on the response Y_1, adjusted for the SNP effect and non-genetic covariates; $\beta_3, \cdots, \beta_{k+2}$ are regression coefficients of non-genetic covariates; e is a random error which is assumed to be distributed as $N(0, \sigma_e^2)$.

Li et al. (2006) reminded that one has to determine which trait to appropriately use as a covariate. For instance, it is inappropriate to use a trait with smaller SNP effect as the dependent variable. This decision can be informed from the known relationships among the traits from previous studies or can be carried out systematically from a set of unconditional univariate analyses:

$$Y_1 = \beta_{10} + \beta_{11}SNP + \beta_{12}X_1 + \cdots + \beta_{1,k+1}X_k + e_1, \quad (2.6)$$

$$Y_2 = \beta_{20} + \beta_{21}SNP + \beta_{22}X_1 + \cdots + \beta_{2,k+1}X_k + e_2. \quad (2.7)$$

A substantial change in the regression estimates of a marker between the uncondtitional model (2.6) and the conditional model (2.5) reveals a presence of pleiotropic effect (Li et al. 2006). However, this approach does not focus on the contribution of marker(s) to the phenotypic covariance. Therefore, it does not quantify and test pleiotropic effect.

2.5 Structural Equation Modeling

Structural Equation Modeling (SEM) includes a type of analysis referred to , also known as path analysis, a straightforward extension of multiple regressions. SEM can include both manifest (observed) variables as well as latent (unobserved) variables whereas path analysis in the usual sense only models manifest variables. The purpose of SEM and path models is to account for variation or covariation of the measured variables and to distinguish direct from indirect effects in mediation models. According to Wright (1934) and Li et al. (2006), one important feature of SEM is that it is flexible to

distinguish direct and indirect effects of a QTL on a trait, and then it is feasible to calculate and compare the relative strengths of effects along different paths. Regression coefficients between the exogenous (or independent) and the endogenous (or dependent) variables present the effect of a direct path from a QTL to a trait (Shipley 2000; Li et al. 2006).

Li et al. (2006) proposed a methodology of using SEM for the analysis of multilocus and multitrait genetic data in the following steps:

1. Identify SNPs for each individual phenotype based on a linear model

$$Y_1 = \beta_{10} + \beta_{11} SNP + \beta_{12} X_1 + \cdots + \beta_{1,k+1} X_k + e_1, \quad (2.8)$$

$$Y_2 = \beta_{20} + \beta_{21} SNP + \beta_{22} X_1 + \cdots + \beta_{2,k+1} X_k + e_2. \quad (2.9)$$

where Y_1 and Y_2 denote traits; SNP denotes the genetic marker with observed genotypes; $X_1, \cdots, X_k$ are non-genetic k predictor variables or covariates; $\beta_{i0}, \cdots, \beta_{i,k+1}$ are regression coefficients; e_i is a residual error which is assumed to be distributed as $N(0, \sigma_{e_i}^2)$, $i, = 1,2$.

2. Perform conditional analysis using one trait as a covariate in the analysis of another trait.

$$Y_1 = \beta_0 + \beta_1 SNP + \beta_2 Y_2 + \beta_3 X_1 + \cdots + \beta_{k+2} X_k + e, \quad (2.10)$$

where Y_1 denotes the trait as dependent variable; Y_2 denotes the trait as a covariate; β_1 is the effect of the SNP of interest, adjusted for Y_2 and non-genetic covariates; β_2 is the effect of Y_2 on the response Y_1, adjusted for the SNP effect and non-genetic covariates. The trait to use as covariates should be chosen carefully. The SNP effect on the

dependent variable should not be smaller than that on the covariate.The comparison of results from unconditional and conditional models can give a first insight into the causal relationship among the phenotypes.

3. Construct initial SEM by including SNPs identified in step 1 and 2. The conditioning trait to the response is also added. Figure 2.1 shows the path diagram of a mediation model.

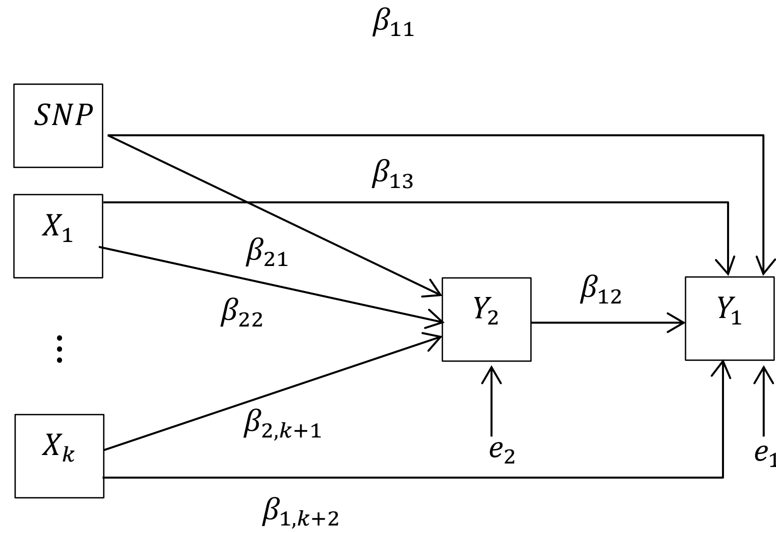

Figure 2. 1 Example of path diagram and structural equation modeling.

This path diagram corresponds to the structural equation models (2.11)

$$Y_1 = \beta_{11}SNP + \beta_{12}Y_2 + \beta_{13}X_1 + \cdots + \beta_{1,k+2}X_k + e_1, \tag{2.11}$$

$$Y_2 = \beta_{21}SNP + \beta_{22}X_1 + \cdots + \beta_{2,k+1}X_k + e_2.$$

with $cov(e_1, e_2) = 0$. Note that the model parameters are typically estimated using the observed covariance matrix; so there is no intercept term in the model (Lattin et al. 2003). Next, assess the model in terms of goodness of fit by comparing the predicted and observed covariance matrices and by significance tests for individual path coefficients. Finally, refine the model by proposing and assessing alternative models. Either the

likelihood ratio test or other model selection criteria, such as the AIC, can be used to compare models. Model refinement and assessment are often carried iteratively (Li et al. 2006; Rosa et al. 2011).

SEM provides an intuitive and precise characterization of the genetic architecture underlying complex traits. For those SNPs that are associated with multiple traits, SEM establishes the relative contributions of the direct and indirect effects (Li et al. 2006). If a SNP(s) has effects on multiple traits, SEMs may help identify the existence of pleiotropy. However, similar to the conditional analysis, it cannot quantify and test pleiotropic effects.

2.6 O'Brien Method

O'Brien (O'Brien 1984) and Wei and Johnson (Wei and Johnson 1985) recommended an approach of combining univariate test statistics of multiple phenotypes which can contain a mixture of quantitative and qualitative measures (Yang et al. 2010).

Suppose Y_j is the value of *j*th trait (*j*=1,2), one can apply a linear regression model to approximate the relationship between trait Y_j and an observed major genetic marker.

$$Y_1 = \beta_{10} + \beta_{11}SNP + \beta_{12}X_1 + \cdots + \beta_{1,k+1}X_k + e_1, \tag{2.12}$$

$$Y_2 = \beta_{20} + \beta_{21}SNP + \beta_{22}X_1 + \cdots + \beta_{2,k+1}X_k + e_2. \tag{2.13}$$

$X_1, \cdots, X_k$ are non-genetic *k* predictor variables or covariates; $\beta_{i0}, \cdots, \beta_{i,k+1}$ are regression coefficients; e_i is a residual error which is assumed to be distributed as $N(0, \sigma_{e_i}^2)$, $i, = 1,2$. Let $T = (T_1, T_2)^T$ be a vector of test statistics representing the association of each

individual trait and the genetic marker, which is assumed to follow a multivariate normal distribution with mean $\beta = (\beta_{11}, \beta_{21})^T$ and covariance matrix Σ. The hypotheses are:

$$H_0: \beta_{11} = \beta_{21} = 0 \quad vs \quad H_1: \beta_{11} \neq 0 \; or \; \beta_{21} \neq 0.$$

O'Brien proposed the linear combination of T_1, T_2 with weight $e = (1,1)^T$,

$$U = e^T \Sigma^{-1} T \,, \tag{2.14}$$

which is most powerful among a set of test statistics that are linear combinations of T_1, T_2 if $\beta_{11} = \beta_{21} > 0$. Under the null hypothesis, U is normally distributed with variance $e^T \Sigma^{-1} e$ (Wu 2009; Yang et al. 2010).

The power of the O'Brien method may be eroded if parameters being estimated are very different from each other. Therefore, Wu (2009) proposed two extended strategies, sample splitting method and cross-validation using re-simulations, improving the O'Brien approach to deal with heterogeneous β's.

These approaches can be used to integrate univariate GWAS results of multivariate phenotypes, but does not qualify them as a screening tool for pleiotropic effects. For each SNP, combining univariate test statistics approaches focus on detecting the SNP effect on at least one of the phenotypes, rather than the shared SNP effect on all of the phenotypes. Therefore, these approaches may be less optimal for the pleiotropic effect analysis when a SNP has no effect on any of phenotypes.

2.7 Summary

Unlike univariate analysis, multivariate analysis goes beyond the simple descriptive purposes and focuses more on the relationship between multiple variables.

Thus, multivariate analysis is used to investigate the contribution of a gene or genes to multiple phenotypic traits in a genetic study.

In this chapter, we have reviewed six multivariate analysis methods. GEE and MANOVA conduct an overall test of SNP effects on multiple traits based on a regression model. A similar limitation in pleiotropic effect analysis has been discussed in GEE and MANOVA. If a genetic marker is strongly associated with any trait, GEE and MANOVA possibly make a wrong conclusion of a marker specific pleiotropic effect on traits. In terms of dimension reduction approach, such as PCs analysis, it would be difficult to interpret the result of the estimated effects from the combined variables and discriminate how many trait and which traits share the genetic effects. Both SEM and conditional analysis help identify the presence of pleiotropic effects, but neither of them provides insight into the genetic influence of variation in multiple traits. Thus, they do not quantify and test pleiotropic effects of a major gene locus. As to approaches of combining univariate test statistics of multiple phenotypes, such as O'Brien method, they concentrate on assessing the genetic marker affecting at least one of the phenotypes. In sum, no previously developed methods can be directly applied for pleiotropic effect analysis.

Therefore, new methods are needed to evaluate marker specific pleiotropic effects where a strong association between a SNP and a trait should not significantly influence the overall effect. The new methods will explore the underlying principle behind pleiotropic effect by examining the contribution of specific genetic marker to genetic

correlation or genetic covariance of traits. Family data should be appropriately considered in these methods.

Several researchers have recommended a variance components analysis that partitions the observed phenotypic variance into components attributable to genetic and environmental causes (Schork 1993; Amos 1994; Blangero 1995; Lynch and Walsh, 1998). This method can localize pleiotropic QTL in a certain region with family data in a bivariate linkage analysis (Stein et al. 2004; Zhao et al. 2008). However, the new methods in this dissertation investigate the associations between a genetic marker and traits. In the next chapter, we propose a method of marker specific pleiotropic effect based on VC analysis in association studies.

Chapter 3: Variance Components Analysis-Based Approaches on Pleiotropic Effects

In this chapter, we propose two approaches based on VC analysis for family-based studies. First, we introduce some background about VC analysis. Second, we develop genetic correlation and covariance approaches and examine their performance on estimates across different residual polygenic correlations using simulation studies. Third, we compare these two approaches under different residual polygenic correlations on bias, type I error and power using the bootstrapping resampling method. Finally, we explore the impact of environmental correlation and minor allele frequency on both approaches.

3.1 Introduction

Variance Components (VC) Analysis is one of the most popular approaches of genetic multivariate analysis for family data. This concept was introduced by R. A. Fisher in 1918. Lange and Boehnke (1983) and Boehnke et al. (1986) established the theoretical foundation of polygenic variance components analysis for multivariate traits. Almasy et al. (1997), and Almasy and Blangero (1998) later extended VC approach to multipoint linkage analysis. In addition, they built up the multivariate VC analysis method by decomposing pleiotropic effects of a major locus and the genetic effects of residual polygenes.

Today, VC analysis is widely applied to family data involving genetically related individuals (Almasy et al. 1997; Williams et al. 1999). Allison et al. (1999) pointed out

that the model can fit phenotypes with any general complex pedigree structure, but this model is sensitive to the normality assumption. Violation of multivariate normality assumption may lead to inflated type I error. Therefore, it is unwise to blindly apply the variance components analysis in QTL mapping without regard to the phenotypic distribution. When family sizes are large, the estimates of QTL location and effects are accurate (Grignola et al., 1996).

Genetic researchers often use variance components analysis in the context of linkage studies. Genetic Linkage is a term that describes the tendency of genes and other genetic markers to be inherited together because their locations are near one another on the same chromosome (Davies and Soundy 2009). Linkage analysis has been applied to estimate the genetic distance between genetic markers or between genetic markers and a trait locus in family subjects (Feingold 2001). The primary goal of linkage analysis is to localize genes influencing a specific trait on the human genome, which typically serves as the first procedure in genetic studies. Specifically, multivariate linkage analysis can localize pleiotropic QTL in a certain region for multiple correlated phenotypic traits (Stein et al. 2004; Zhao et al. 2008).

Historically, genetic association analysis usually follows linkage analysis, establishing associations between genetic polymorphisms of a particular candidate gene and phenotype(s) in a population-based study with samples of families or unrelated subjects, where the genetic marker is observable (Haines and Pericak-Vance 1998; Feingold 2001; Ralston and De Crombrugghe 2006; Lunetta 2008). Association studies

are comparatively more powerful than linkage analysis to detect small effects (Cordell and Clayton 2005; Ralston and De Crombrugghe 2006; Wu 2009).

In this research, we apply the ideas and principles of bivariate polygenic VC analysis to human genetic association studies.

3.2 Background

Variance components analysis decomposes the observed phenotypic variance into components attributable to genetic and environmental causes under a general linear model (Schork 1993; Amos 1994; Blangero 1995; Lynch and Walsh 1998). The basic genetic model is

$$Y = \mu + g + \sum_{k=1}^{m} \beta_k X_k + e, \tag{3.1}$$

where Y is the vector of trait value; μ is the overall mean; g is a random polygenic effect; β_k is the effect size of X_k $(k = 1, \cdots, m)$; X_k is the kth environmental predictor variable; and e is a residual variability uncorrelated with the genetic factors and covariates (North et al. 2003; Bauman et al. 2005).

The phenotypic variance usually consists of the genetic and environmental components

$$\sigma_P^2 = \sigma_A^2 + \sigma_D^2 + \sigma_e^2, \tag{3.2}$$

where σ^2 is a variance with subscripts P, A, D and e representing phenotypic, polygenic additive, polygenic dominant and environmental variance. This equation can be

simplified in a narrow sense of heritability by removing the dominant term (σ_D^2). An update can be obtained as

$$\sigma_P^2 = \sigma_g^2 + \sigma_e^2, \tag{3.3}$$

where σ_g^2 represents the additive genetic effects (Lange et al. 1976; Falconer 1981; Hanis et al. 1983; Cheverud 1988; Amos 1994; Allison et al. 1998; Williams et al. 1999). Here heritability (h_r^2) is defined as the proportion of phenotypic variance unexplained by covariates that can be attributed to additive genetic effects

$$h_r^2 = \sigma_g^2/\sigma_P^2. \tag{3.4}$$

3.2.1 QTL Model for a Univariate Trait

After further decomposing the genetic component into two parts, the genetic component due to the QTL effect and the residual genetic component, the polygenic model with a major gene is given by

$$Y = \mu + g^* + q + \textstyle\sum_{k=1}^{m} \beta_k X_k + e, \tag{3.5}$$

where q is a random variable indicating an unobserved major locus component in linkage analysis; g^* is a random polygenic effect. q isolates the contribution of the major gene from the polygenic background (Bauman et al. 2005). The previous genetic variance is also partitioned into two parts

$$\sigma_P^2 = \sigma_{g^*}^2 + \sigma_q^2 + \sigma_e^2, \tag{3.6}$$

where σ_q^2 represents the QTL variance and $\sigma_{g^*}^2$ represents the residual polygenic genetic variance.

The QTL heritability is defined as the percentage of variation explained by the QTL

$$h_q^2 = \sigma_q^2/\sigma_P^2. \tag{3.7}$$

3.2.2 QTL Model for Multivariate Traits

These ideas extend readily to multivariate situations. Without losing generality, we restrict our discussion to two traits in which both traits are simultaneously considered. Compared with the univariate polygenic model, the bivariate polygenic model additionally estimates the genetic and environmental correlations between both traits.

Let Y_1 and Y_2 be the corresponding random traits. The bivariate traits can be expressed as $Y = (Y_1', \ Y_2')'$ with the mean

$$E(Y) = Z\beta, \tag{3.8}$$

where $Z = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ is a block diagonal design matrix specifying the covariates, β represents the regression coefficient matrix (Bauman et al. 2005).

Lange and Boehnke (1983) and Lange (2002) derived the decomposition of the covariance matrix

$$\Omega = \begin{bmatrix} \sigma_{g1}^2 & \rho_g \sigma_{g1}\sigma_{g2} \\ \rho_g \sigma_{g1}\sigma_{g2} & \sigma_{g2}^2 \end{bmatrix} \otimes 2\Phi + \begin{bmatrix} \sigma_{q1}^2 & \rho_q \sigma_{q1}\sigma_{q2} \\ \rho_q \sigma_{q1}\sigma_{q2} & \sigma_{q2}^2 \end{bmatrix} \otimes \Pi$$

$$+ \begin{bmatrix} \sigma_{e1}^2 & \rho_e \sigma_{e1}\sigma_{e2} \\ \rho_e \sigma_{e1}\sigma_{e2} & \sigma_{e2}^2 \end{bmatrix} \otimes \mathrm{I}, \tag{3.9}$$

where $\otimes$ is the Kronecker product operator. The three Kronecker products appearing in this equation correspond to the residual polygenic contribution, the major gene

contribution, and the random environmental contribution, respectively; ρ_g is the shared residual additive genetic influence on the traits; ρ_q is the shared major gene effects; and ρ_e is the shared environmental influence on the traits. Φ is the kinship matrix whose elements are defined as the probability that a gene selected randomly from person *i* and a gene selected randomly from the same autosomal locus of person j are identical by descent (IBD) (Bauman 2005). Π is a matrix whose elements are the estimated proportion of genes shared identical by descent at the QTL, by individuals *i* and *j* (Williams 1999).

If we replace the 2×2 covariance matrices on the left of each Kronecker product by $r \times r$ covariance matrices, this formula will be generalized to *r* multivariate traits.

3.3 Analysis of Pleiotropic Effect Using Genetic Correlation

Mahaney et al (1995) introduced a decomposition of the phenotypic correlation, which consists of the genetic components and environmental components.

$$\rho_P = \rho_g\sqrt{h_{r1}^2 h_{r2}^2} + \rho_e\sqrt{(1-h_{r1}^2)(1-h_{r2}^2)} \tag{3.10}$$

where ρ is a correlation with subscripts P, g, and e representing phenotypic, polygenic additive and environmental correlations, respectively, for trait1 and trait2; h_{r1}^2 and h_{r2}^2 are the heritabilities ($h_r^2 = \sigma_g^2/\sigma_P^2$) for trait1 and trait2; $(1-h_{r1}^2)$ and $(1-h_{r2}^2)$ are the proportions of phenotypic variance due to environmental factors ($(1-h_r^2) = \sigma_e^2/\sigma_P^2$) for trait1 and trait2 (Falconer 1981; Mahaney et al. 1995).

The polygenic correlation (ρ_g) relative to the phenotypic correlation (ρ_P) can be simply evaluated in terms of model (3.10). The significance of ρ_g can be assessed by

means of likelihood ratio tests. The likelihood of model where the correlation is estimated is compared with the likelihood of model where the correlation is constrained to zero (rejection of $\rho_g = 0$ indicates pleiotropy) or to 1 (failure to reject $|\rho_g| = 1$ indicates complete pleiotropy) (Vinson et al. 2008).

If the null hypothesis of $\rho_g = 0$ is rejected indicating that traits share the same genetic effects, a QTL that influences phenotypic variation in both traits may be detected (Vinson et al. 2008). Linkage study or association study can be employed to investigate a specific region or a specific marker, aiming to evaluate the major genetic correlation. The genetic components can be further partitioned into a major gene components and residual polygenic components

$$\rho_P = \rho_{g^*}\sqrt{h^2_{r1^*}h^2_{r2^*}} + \rho_q\sqrt{h^2_{q1}h^2_{q2}} + \rho_e\sqrt{\left(1 - h^2_{r1^*} - h^2_{q1}\right)\left(1 - h^2_{r2^*} - h^2_{q2}\right)}. \tag{3.11}$$

An additional parameter estimated in this bivariate model is ρ_q, the additive genetic correlation between the traits due to the effect of the QTL (Vinson et al., 2008). QTL correlation (ρ_q) and QTL heritabilities (h^2_{q1} and h^2_{q2}) are ascribed to QTL components. The residual genetic correlation (ρ_{g^*}) and the residual heritabilities ($h^2_{r1^*}$ and $h^2_{r2^*}$) are ascribed to the polygenic components. $\sqrt{h^2_{r1^*}h^2_{r2^*}}$ is the maximum contribution to ρ_P from the polygenic components, which is weighted by ρ_{g^*} assigning a proportion of

$\sqrt{h_{r1^*}^2 h_{r2^*}^2}$ to ρ_P. $\sqrt{h_{q1}^2 h_{q2}^2}$ is the maximum contribution of the QTL to ρ_P, which is weighted by ρ_q controlling the proportion of the maximum amount of $\sqrt{h_{q1}^2 h_{q2}^2}$.

A family-based bivariate linkage analysis allows the inference of pleiotropy between the traits from significance of the QTL correlation (Hasstedt and Thomas 2011). However, in a population-based association study with samples of families or genetically unrelated subjects, existing methods are not applicable to estimate and test the shared QTL specific correlation. Therefore, our goal in this chapter is to examine the shared genetic and environmental correlations of VC analysis in bivariate polygenic models, and assess its utility for analysis of marker specific or SNP specific pleiotropic effects by comparing genetic correlations or covariances from bivariate polygenic models with and without adjustment for the SNP.

3.3.1 Proposed Method

We propose a method of estimating ρ_q from the difference of genetic correlations between two bivariate polygenic models:

Model1(Reduced Model)

$$Y_1 = \mu_1 + g_1 + e_1{}^*, \qquad (3.12)$$

$$Y_2 = \mu_2 + g_2 + e_2{}^*.$$

Model2(Full Model)

$$Y_1 = \mu_1{}^* + g_1^* + \beta_1 SNP + e_1, \qquad (3.13)$$

$$Y_2 = \mu_2{}^* + g_2^* + \beta_2 SNP + e_2;$$

Y_1 and Y_2 are two correlated traits; μ_i and $\mu_i{}^*$ are the mean traits of all individuals; g_i is a random polygenic effect and g_i^* is a random residual polygenic effect; *SNP* is 0,1 and 2 if the genotype of the individual at the SNP is AA, Aa and aa, respectively; β_i is the size of genetic effect of the SNP of interest on the *i*th trait; e_i and $e_i{}^*$ are the non-genetic (residual) effects, $E(e_i) = 0$, $Cov(e_i, e_j) = \sigma_{ij}$, $E(e_i^*) = 0$, $Cov(e_i^*, e_j^*) = \sigma_{ij}^*$, $i, j =$ 1,2.

The reduced model (3.12) is a bivariate model containing an intercepts and random variables capturing the polygenic effect and a combination of random environment and measurement errors, respectively (Lange 2002; Bauman et al. 2005). The full model (3.13) additionally contains a SNP variable whose genotypes are fixed and observable. We implement an analysis of marker specific pleiotropic effect by comparing the polygenic correlation (ρ_g) in the reduced model to the residual polygenic correlation (ρ_{g^*}) in the full model.

The hypothesis can be expressed as

$$H_0: \rho_{g_diff} = 0 \quad vs \quad H_1: \rho_{g_diff} \neq 0,$$

where $\rho_{g_diff} = \rho_g - \rho_{g^*}$ is a difference of the polygenic correlation in the reduced model to the residual polygenic correlation in the full model. The estimate could be either positive or negative based on the SNP effects on traits that are in the same direction or opposite direction.

There are no formulas to estimate the variability of the difference of genetic correlations. Fortunately, bootstrap resampling allows us to estimate standard errors and construct confidence intervals for a wide variety of statistics and provides sufficient information to perform statistical hypothesis testing (Hesterberg et al. 2005).

In bootstrapping, we resample with replacement by family from the original sample to create a bootstrap sample of the same family sizes as the original family sample. Then we generate an empirical sampling distribution for ρ_{g_diff} with a large number of such bootstrapped samples to construct confidence intervals (CIs) and test for significance. If a SNP has a pleiotropic effect between traits, we expect to see a significant difference in the genetic correlations, with zero is not included in the CI of ρ_{g_diff}, suggesting rejection of the null hypothesis.

3.3.2 Evaluation of the Proposed Method Using Simulation Studies

We conducted simulation studies to investigate the performance of VC-based genetic correlation approach. One thousand data sets were simulated using the 'simqtl' command in the statistical software package ---- Sequential Oligogenic Linkage Analysis Routines (SOLAR) (http://www.vipbg.vcu.edu/software_docs/solar/doc/index.html) that generated the marker and the phenotypic data. Two normally distributed quantitative traits and a di-allelic SNP with 10% or 50% minor allele frequency (MAF) were generated using 1,000 uncorrelated trios (2 parents and a child). The simulation designs were based on: residual polygenic correlation (ρ_{g^*}) of 0.1, 0.6 and 0.9; environmental

correlation (ρ_e) of 0.0, 0.6 and 0.9; and residual heritability ($h^2_{r^*}$) of 0.4. Seven SNP heritabilities were considered (h^2_q =0%, 0.5%, 1%, 2%, 3%, 10% and 30% giving the size of effect for each SNP as 0, 0.1, 0.1414, 0.2000, 0.2450, 0.4472 and 0.7746 units, and the standard deviation for each SNP as 1, 0.998, 0.995, 0.990, 0.985, 0.949 and 0.837, respectively). Twenty pairs were set for the bivariate traits with different SNP heritability combinations (Table 3.1). We used SOLAR's polygenic analysis function for multivariate models to conduct our VC-based analyses.

The term, t0_t0 denotes a SNP that has no effect on either trait and t0_th^2_{q2} denotes a SNP that has 0% of effect on trait1 but h^2_{q2}% of effect on trait2, a circumstance with no pleiotropy. th^2_{q1}_th^2_{q2} denotes a SNP with h^2_{q1}% effect on trait1 and h^2_{q2}% effect on trait2, indicating the existence of a SNP-specific pleiotropic effect.

Table 3. 1 Pairs of SNP effects on bivariate traits in simulations of VC analysis

No pleiotropy (No effects on T1 and T2)	t0_t0
No pleiotropy (No effect on T1)	t0_t05, t0_t1, t0_t2, t0_t3, t0_t10, t0_t30
Pleiotropic effect on both T1 and T2	t05_t05, t05_t1, t05_t2, t05_t3, t1_t1, t1_t2, t1_t3, t2_t2, t2_t3, t3_t3, t10_t10, t10_30, t30_t30

*t$\boldsymbol{h^2_{q1}}$_t$\boldsymbol{h^2_{q2}}$ denotes SNP with $\boldsymbol{h^2_{q1}}$% effect on trait1 and $\boldsymbol{h^2_{q2}}$% effect on trait2.

Simulated data from each sample were analyzed using VC analysis by the "polygenic" command in SOLAR to obtain the genetic correlation. The difference of genetic correlations from models excluding the SNP effect (3.12) and including the SNP effect (3.13) was computed as the contribution of SNP to the polygenic pleiotropic effect.

The percent bias (%) was calculated as

$$\text{Percent Bias (\%)} = \frac{\text{Est} - \text{True}}{\text{True}} \times 100\%,$$

which was used to indicate the performance of the methods being assessed, provided the true value does not equal to zero (Burton et al. 2006). Estimates with percent bias <10% are considered as acceptable.

The whole procedure of simulation can be summarized as follows:

Simulation Steps:

Step1: Generate 1,000 uncorrelated trios (2 parents and a child);

Step2: Simulate a dataset with two normally distributed quantitative traits and a di-allelic SNP using the pedigree structure in Step1. The simulation designs include 20 SNP heritability scenarios, 3 residual polygenic correlation conditions and 3 environmental correlation conditions;

Step3: Conduct bivariate VC analysis for a polygenic model excluding the SNP effect as a predictor variable. Estimate the polygenic correlation ($\hat{\rho}_{g1}$);

Step4: Conduct bivariate VC analysis for a polygenic model including the SNP effect as a predictor variable whose genotypes were fixed and observable. Estimate the residual polygenic correlation ($\hat{\rho}_{g1^*}$);

Step5: Calculate the contribution of the SNP to the polygenic pleiotropic effect by : $\hat{\rho}_{g_diff1} = \hat{\rho}_{g1} - \hat{\rho}_{g1^*}$;

Step6: Repeat Steps 2-5 1,000 times for 1,000 replicates;

Step7: Compute the mean of $\hat{\rho}_{g_diff1}, \cdots, \hat{\rho}_{g_diff1000}$ as $\hat{\rho}_{g_diff}$.

Next, 500 resamples were randomly drawn with replacement from each replicate by sampling from 1,000 uncorrelated trios with the same bootstrap sample size. VC-based analyses were performed for these bootstrap resamples. The intervals between the 2.5% and 97.5% percentiles of the bootstrap distribution of the difference of genetic correlations were used for statistical inference by examining if zero was included in these

CIs. We evaluated type I error and power by determining the proportion of CIs containing zero in these 1,000 replicates. If the CI does not contain zero then we would reject a null hypothesis of zero at the 5% significance level, allowing we to evaluate type I error. We used Bradley's criterion for determining inflated versus conservative type I error rates (Bradley 1978). The fraction of rejections above 0.055 is termed as "inflated", whereas an empirical value below 0.045 is termed as "conservative" for a nominal $\alpha = 0.05$.

The genetic correlation approach using bootstrapping in simulation method can be implemented according to the following steps:

Bootstrap Steps:

Step1: Use the same 1,000 simulated data from the simulation procedure;

Step2: Draw 1,000 trios $x_1^*, \cdots, x_{1000}^*$ from replicate 1 with replacement;

Step3: Compute $\hat{\rho}_{g_diff(1,1)}^*$ using $x_1^*, \cdots, x_{1000}^*$;

Step4: Repeat Steps 2-3 B=500 times. With a large number of new samples, generate an empirical sampling distribution for ρ_{g_diff1}: $\{\hat{\rho}_{g_diff(1,1)}^*, \cdots, \hat{\rho}_{g_diff(1,500)}^*\}$;

Step5: Compute the mean of $\hat{\rho}_{g_diff(1,1)}^*, \cdots, \hat{\rho}_{g_diff(1,500)}^*$ as $\hat{\rho}_{g_diff1}^*$;

Step6: Construct the 95% confidence interval from the empirical sampling distribution of $\{\hat{\rho}_{g_diff(1,1)}^*, \cdots, \hat{\rho}_{g_diff(1,500)}^*\}$ as $(a, b)_1$;

Step7: Repeat Steps 2-6 1,000 times for 1,000 replicates, resulting in 1,000 confidence intervals $\{(a, b)_1, \cdots, (a, b)_{1000}\}$;

Step8: Determine if these confidence intervals cover 0 or not, and calculate the proportion of times the 95% confidence interval excludes 0;

Step9: Compute the mean of $\hat{\rho}^*_{g_diff1}, \cdots, \hat{\rho}^*_{g_diff1000}$ as $\hat{\rho}^*_{g_diff}$.

The polygenic VC analysis using SOLAR is costly and time consuming. The average time for accessing the storage and computing the analysis requires 2 minutes for one process on an Advanced Micro Devices (AMD) based compute cluster which includes one head node (two quad-core 2.7 GHz 64-bit Opteron processors and 16 GB of RAM), 32 servers (two 1.8 GHz 64-bit Opteron processors and 2 GB RAM each) and 56 blades (two dual-core 2.6 GHz 64-bit Opteron processors and 12 GB RAM each (one compute node has 32 GB of RAM)). In order to calculate $\hat{\rho}_{g_diff}$, we need 2 processes for the full model and reduced model. The total time required to perform VC analysis for bootstrapping depends on the simulation size and bootstrapping size. We performed 1,000 simulations and 500 bootstrapping in each simulation. We can run 20 processes simultaneously on many compute nodes. A formula that give an approximate estimate of the time is:

$$Number\ of\ days = \frac{2(minutes) \times 1000(simulations) \times 500(bootstrappings) \times 2(models)}{60(minutes) \times 24(hours) \times 20(nodes)}$$

It takes 69 days for each SNP heritability scenario. Thus, it is an incentive to conduct the bootstrapping only for selected scenarios. We limit our discussion to nine

scenarios with zero, low, medium, or high SNP effect on one or two traits, respectively (t0_t0, t0_t3, t0_t10, t0_t30, t1_t1, t3_t3, t10_t10, t10_t30 and t30_t30).

3.3.3 Simulation Results

The difference of polygenic genetic correlations (ρ_{g_diff}) between the full model and reduced model across SNP heritability pairs were explicitly compared by three residual polygenic correlation generating values ($\rho_{g^*} = 0.1$, 0.6 and 0.9) under 50% MAF and moderate environmental correlation ($\rho_e = 0.6$) (Figure 3.1). Each figure provides the true ρ_{g_diff} and the mean of the estimated ρ_{g_diff} from VC analysis.

If a SNP has no effect on at least one trait, we would expect the estimates to be zero, indicating non-existence of pleiotropy. When the residual polygenic correlation is low, the genetic correlation approach has very little negative bias in the effect estimates, if at all (Figure 3.1(A)). The bias increases with the residual polygenic correlation level (Figure 3.1 (B)-(C)).

On the other hand, if a SNP has effects on both traits, we would expect that the resulting difference of genetic correlations is greater than zero and increases with large variant effects. A significant difference in genetic correlations should provide an evidence for the presence of SNP-specific pleiotropic effect. We found very little bias in the pleiotropic effect estimate, which decreases as the residual polygenic correlation increased.

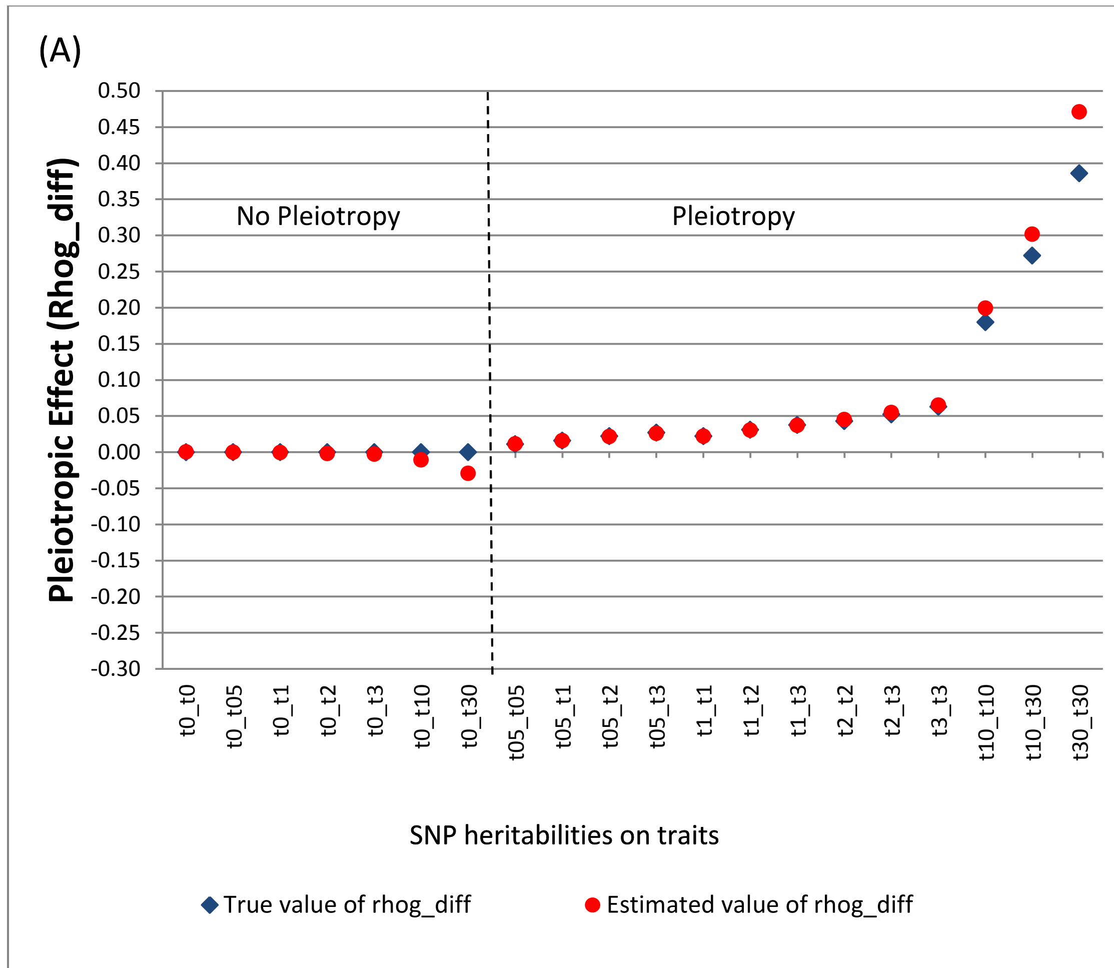

Figure 3. 1 True and estimated values of the difference between polygenic correlations from bivariate polygenic models with and without adjustment for a SNP

Three polygenic correlation generating values ($\boldsymbol{\rho_{g^*}}$): 0.1 (A), 0.6 (B) and 0.9 (C). MAF=0.5, $\boldsymbol{\rho_e = 0.6}$, $\boldsymbol{h^2_{r1^*} = h^2_{r2^*} = 0.4}$

* t$\boldsymbol{h^2_{q1}}$_t$\boldsymbol{h^2_{q2}}$ denotes SNP with $\boldsymbol{h^2_{q1}}$% effect on trait1 and $\boldsymbol{h^2_{q2}}$% effect on trait2.

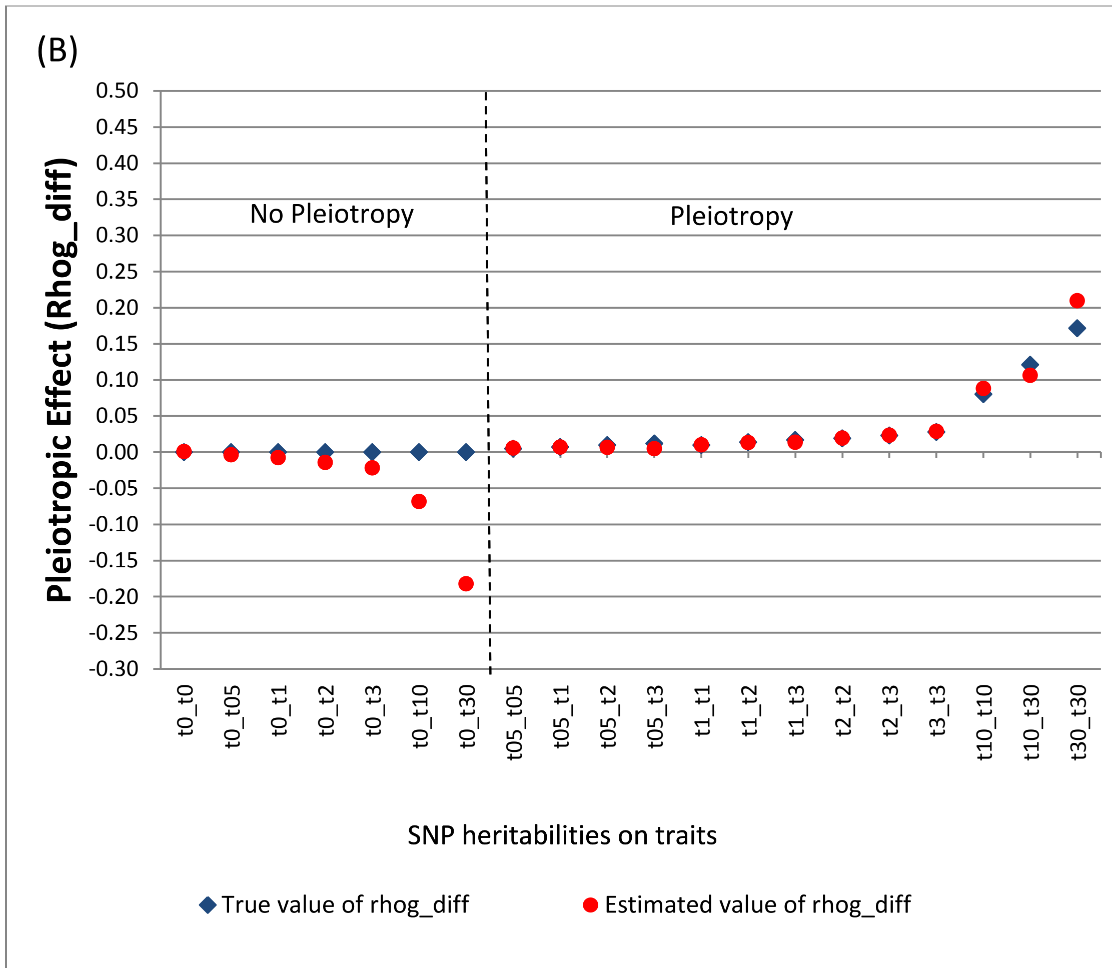

Figure 3.1 (Continued) True and estimated values of the difference between polygenic correlations from bivariate polygenic models with and without adjustment for a SNP

Three polygenic correlation generating values ($\boldsymbol{\rho_{g^*}}$): 0.1 (A), 0.6 (B) and 0.9 (C). MAF=0.5, $\boldsymbol{\rho_e = 0.6}$, $\boldsymbol{h^2_{r1^*} = h^2_{r2^*} = 0.4}$

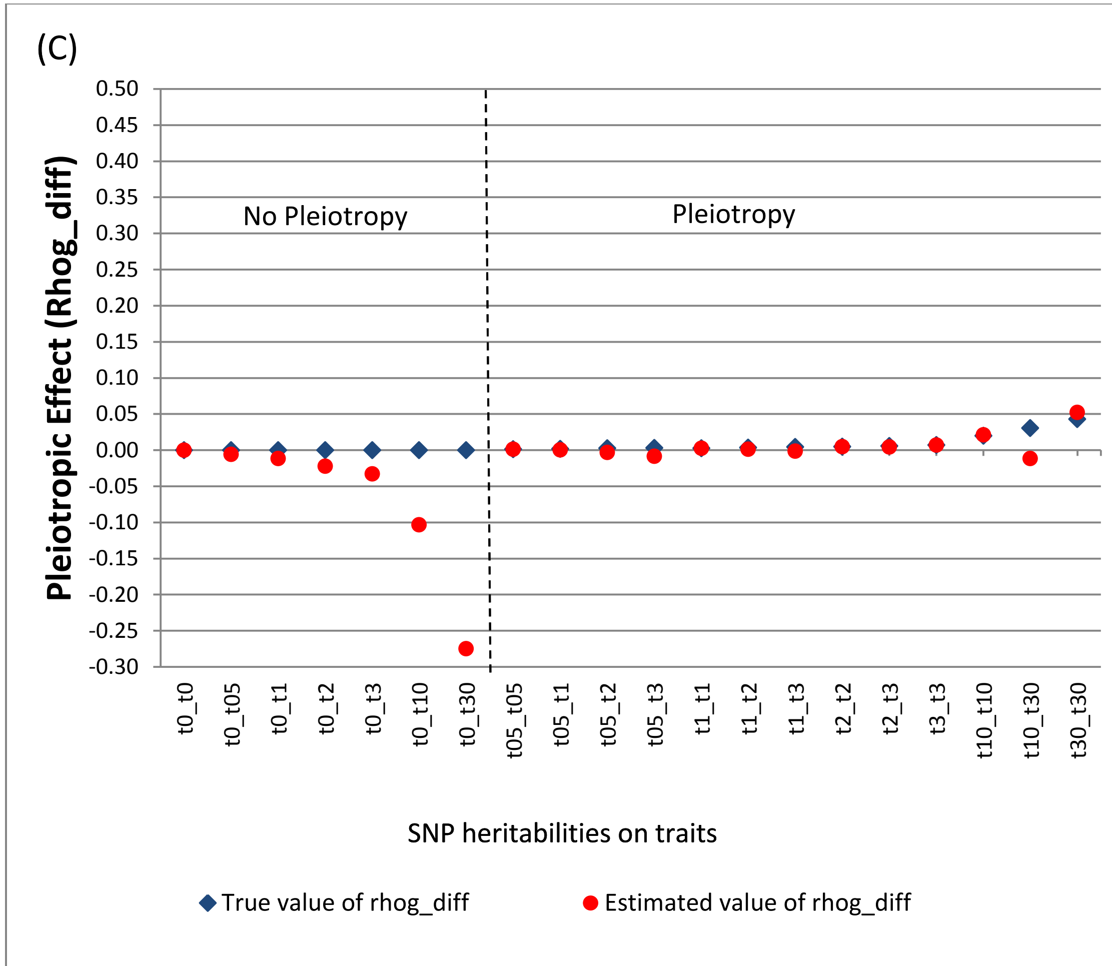

Figure 3.1 (Continued) True and estimated values of the difference between polygenic correlations from bivariate polygenic models with and without adjustment for a SNP

Three polygenic correlation generating values (ρ_{g^*}): 0.1 (A), 0.6 (B) and 0.9 (C). MAF=0.5, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$

The VC-based genetic correlation approach is affected by the residual polygenic correlation level. It is effective only when the polygenic genetic correlation is low.

Next, we explore the utility of genetic covariances instead of correlations from bivariate polygenic models in equation (3.12) and (3.13) to study SNP-specific pleiotropic effect.

3.4 Analysis of Pleiotropic Effect Using Genetic Covariance

3.4.1 Proposed Method

In (3.10), the phenotypic correlation is the sum of the genetic and environmental components. After multiplying by phenotypic variances on both sides, the phenotypic covariance can be represented as the sum of the genetic covariance and environmental covariance.

Correlation

$$\rho_P = \rho_g\sqrt{h_{r1}^2 h_{r2}^2} + \rho_e\sqrt{h_{e1}^2 h_{e2}^2}. \tag{3.14}$$

Covariance

$$\rho_P\sqrt{\sigma_{P1}^2\sigma_{P2}^2} = \rho_g\sqrt{h_{r1}^2 h_{r2}^2}\sqrt{\sigma_{P1}^2\sigma_{P2}^2} + \rho_e\sqrt{h_{e1}^2 h_{e2}^2}\sqrt{\sigma_{P1}^2\sigma_{P2}^2}. \tag{3.15}$$

Thus

$$\rho_P\sqrt{\sigma_{P1}^2\sigma_{P2}^2} = \rho_g\sqrt{\sigma_{g1}^2\sigma_{g2}^2} + \rho_e\sqrt{\sigma_{e1}^2\sigma_{e2}^2}, \tag{3.16}$$

where σ_{P1}^2 and σ_{P2}^2 denote the phenotypic variances for trait1 and trait2; σ_{g1}^2 and σ_{g2}^2 denote the polygenic variances for trait1 and trait2; σ_{e1}^2 and σ_{e2}^2 denote the environmental variances for trait1 and trait2.

The covariance model (3.16) can then be written as:

$$cov_P = cov_g + cov_e, \tag{3.17}$$

where cov is a covariance with subscripts P, g, and e representing phenotypic, polygenic and environmental covariances, respectively for trait 1 and 2. The polygenic covariance from the bivariate polygenic model (3.12) can be expressed as

$$cov_g = \rho_g \sqrt{h_{r1}^2 h_{r2}^2} \sqrt{\sigma_{P1}^2 \sigma_{P2}^2}. \tag{3.18}$$

Similar to the correlation equation (3.11), the polygenic covariance can be decomposed into a major gene covariance and residual polygenic covariance

$$cov_P = cov_{g^*} + cov_{qtl} + cov_e, \tag{3.19}$$

where cov_{qtl} denotes the QTL covariance and cov_{g^*} denotes the residual polygenic covariance. The residual polygenic covariance from bivariate polygenic model (3.13) can be expressed as

$$cov_{g^*} = \rho_{g^*} \sqrt{h_{r1^*}^2 h_{r2^*}^2} \sqrt{\sigma_{P1}^2 \sigma_{P2}^2}. \tag{3.20}$$

To detect pleiotropy, we propose estimating cov_{qtl} from the difference of the genetic covariances between bivariate polygenic models excluding and including the SNP effect (3.12-3.13).

The hypotheses can be expressed as

$$H_0: cov_{g_diff} = 0 \quad vs \quad H_1: cov_{g_diff} \neq 0,$$

where $cov_{g_diff} = cov_g - cov_{g^*}$ is a difference of the polygenic covariance in the reduced model to the residual polygenic covariance in the full model. The estimate could be either positive or negative based on the SNP effects on traits that are in the same direction or opposite direction.

As with the genetic correlation approach, bootstrap resampling is also used for the genetic covariance approach to calculated standard errors, construct confidence intervals, perform statistical hypothesis testing and compute type I error and power.

We expect that a significant difference in genetic covariances provide evidence for marker specific pleiotropic effects.

3.4.2 Evaluation of the Proposed Method Using Simulation

The same sets of simulation design and simulation data in section 3.3.2 were used here and were analyzed using VC analysis by 'polygenic' command in SOLAR to obtain the genetic correlation and heritabilities. The difference of genetic covariances from models excluding the SNP effect (3.12) and including the SNP effect (3.13) is computed as the contribution of SNP to the polygenic pleiotropic effect. The simulation procedure is summarized as follows:

Simulation Steps:

Step1: Use the same simulation design and simulation data in section 3.3.2;

Step2: Estimate the trait variances ($\hat{\sigma}_{P1}^2$ and $\hat{\sigma}_{P2}^2$);

Step3: Conduct bivariate VC analysis for a polygenic model excluding the SNP effect as a predictor variable. Estimate the polygenic correlation ($\hat{\rho}_g$) and the heritabilities ($\hat{h}_{r1}^2$ and $\hat{h}_{r2}^2$), and calculate the estimated polygenic covariance by: $\widehat{cov}_g = \hat{\rho}_g\sqrt{\hat{h}_{r1}^2\hat{h}_{r2}^2}\sqrt{\hat{\sigma}_{P1}^2\hat{\sigma}_{P2}^2}$;

Step4: Conduct bivariate VC analysis for a polygenic model including the SNP effect as a predictor variable whose genotypes are fixed and observable. Estimate the residual polygenic correlation ($\hat{\rho}_{g^*}$) and the residual heritabilities ($\hat{h}_{r1^*}^2$ and $\hat{h}_{r2^*}^2$), and calculate the estimated residual polygenic covariance by: $\widehat{cov}_{g^*} = \hat{\rho}_{g^*}\sqrt{\hat{h}_{r1^*}^2\hat{h}_{r2^*}^2}\sqrt{\hat{\sigma}_{P1}^2\hat{\sigma}_{P2}^2}$;

Step5: Compute the SNP-specific pleiotropic effect by: $\widehat{cov}_{g_diff} = \widehat{cov}_g - \widehat{cov}_{g^*}$:

Step6: Repeated Steps 2-5 1,000 times for 1,000 replicates.

The procedure of the genetic covariance approach using bootstrapping in simulations is summarized as follows:

Bootstrap Steps:

Step1: Use the same bootstrap resampling data from the genetic correlation approach in section 3.3.2;

Step2: Compute $\{\widehat{cov}^*_{g_diff(1,1)}, \cdots, \widehat{cov}^*_{g_diff(1,500)}\}$ for 500 bootstrap samples in simulation1 and generated an empirical sampling distribution;

Step3: Compute the mean of $\widehat{cov}^*_{g_diff(1,1)}, \cdots, \widehat{cov}^*_{g_diff(1,500)}$ as $\widehat{cov}^*_{g_diff1}$;

Step4: Construct the 95% confidence interval from the empirical sampling distribution of $\{\widehat{cov}^*_{g_diff(1,1)}, \cdots, \widehat{cov}^*_{g_diff(1,500)}\}$ as $(c, d)_1$;

Step5: Repeat Steps 2-4 1,000 times for 1,000 replicates, resulting in 1,000 confidence intervals $\{(c, d)_1, \cdots, (c, d)_{1000}\}$;

Step6: Determine if these confidence intervals cover 0 or not, and calculate the percentage of time the 95% confidence interval includes 0;

Step7: Compute the mean of $\widehat{cov}^*_{g_diff1}, \cdots \widehat{cov}^*_{g_diff1000}$ as $\widehat{cov}^*_{g_diff}$.

3.4.3 Simulation Results

The difference of polygenic genetic covariances between the full model and reduced model across SNP heritability pairs were explicitly compared by three residual polygenic covariance generating values ($cov_{g^*} = 0.04$, 0.24 and 0.36 corresponding to $\rho_{g^*} = 0.1$, 0.6 and 0.9, respectively) under 50% MAF and moderate environmental correlation ($\rho_e = 0.6$) (Figure 3.2). Each figure displays the true cov_{g_diff} and the mean of the estimate from VC analysis.

Similar to the genetic correlation approach, if a SNP has no effect on at least one trait, the estimates are expected to be zero, indicating no pleiotropy. When the residual polygenic covariance is low, the genetic covariance approach is approximately unbiased (Figure 3.1(A)). However, larger residual polygenic covariance typically yields greater bias (Figure 3.1 (B)-(C)).

If a SNP has effects on both traits, we expect the resulting difference of genetic correlations to be greater than zero and increase with large SNP effects. A substantial difference indicates a SNP-specific pleiotropic effect. Under low residual polygenic covariance level, the genetic covariance approach provides almost unbiased estimate. However, larger residual polygenic correlation generates greater bias.

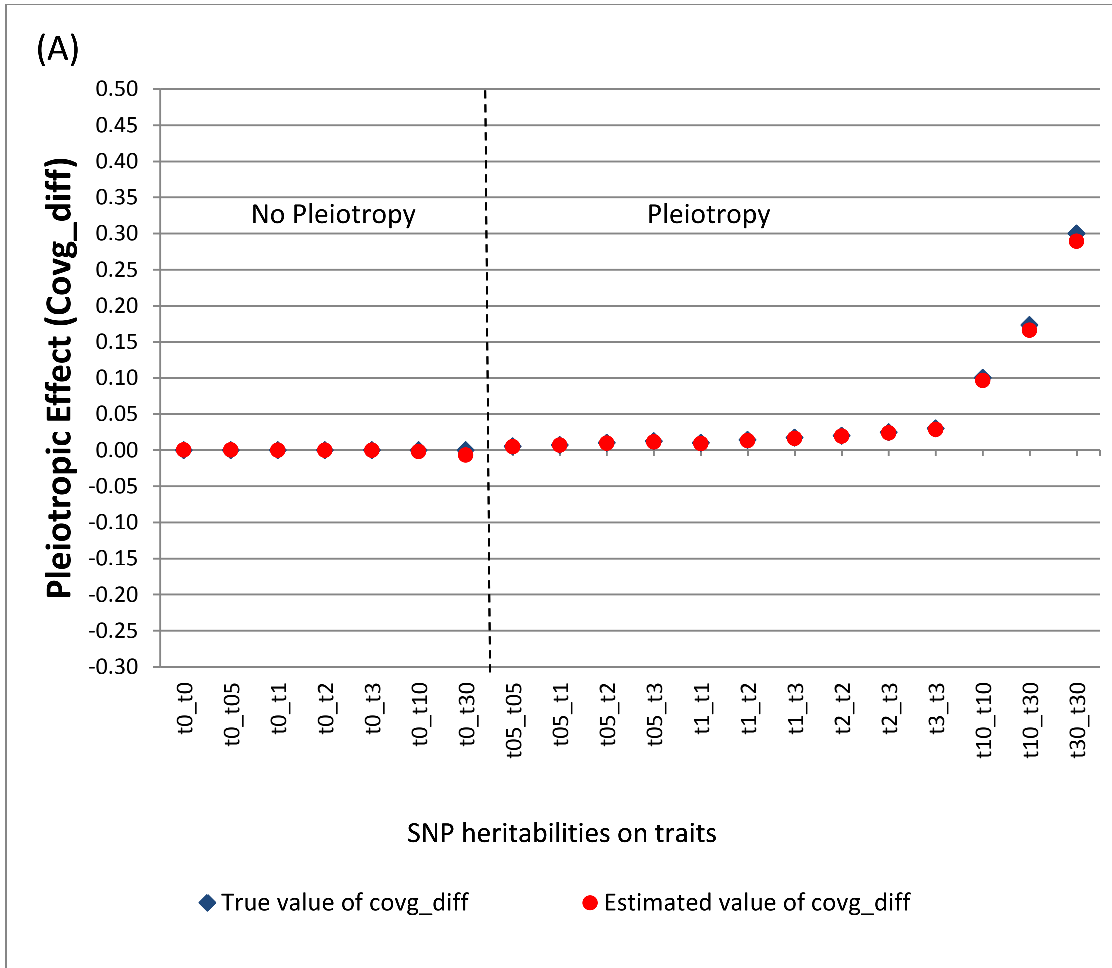

Figure 3. 2 True and estimated values of the difference between polygenic covariances from bivariate polygenic models with and without adjustment for a SNP.

Three polygenic covariance generating values ($\boldsymbol{cov_{g^*}}$): 0.04 (A), 0.24 (B) and 0.36 (C) which correspond to the polygenic correlation generating values ($\boldsymbol{\rho_{g^*}}$): 0.1, 0.6 and 0.9. MAF = 0.5, $\boldsymbol{\rho_e = 0.6}$, $\boldsymbol{h^2_{r1^*} = h^2_{r2^*} = 0.4}$

* t$\boldsymbol{h^2_{q1}}$_t$\boldsymbol{h^2_{q2}}$ denotes SNP with $\boldsymbol{h^2_{q1}}$% effect on trait1 and $\boldsymbol{h^2_{q2}}$% effect on trait2.

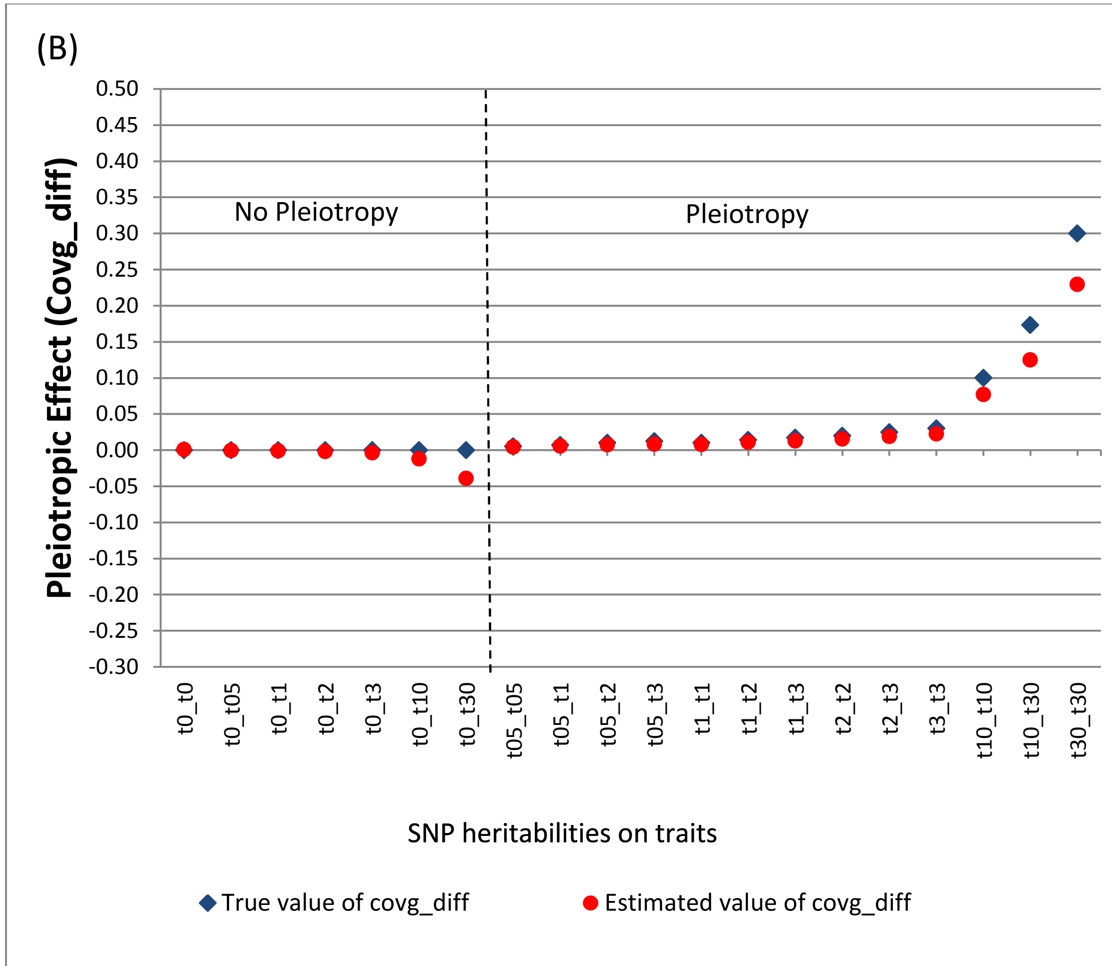

Figure 3.2 (Continued) True and estimated values of the difference between polygenic covariances from bivariate polygenic models with and without adjustment for a SNP.

Three polygenic covariance generating values ($\mathbf{cov_{g^*}}$): 0.04 (A), 0.24 (B) and 0.36 (C) which correspond to the polygenic correlation generating values ($\boldsymbol{\rho_{g^*}}$): 0.1, 0.6 and 0.9. MAF = 0.5, $\boldsymbol{\rho_e = 0.6}$, $\boldsymbol{h^2_{r1^*} = h^2_{r2^*} = 0.4}$

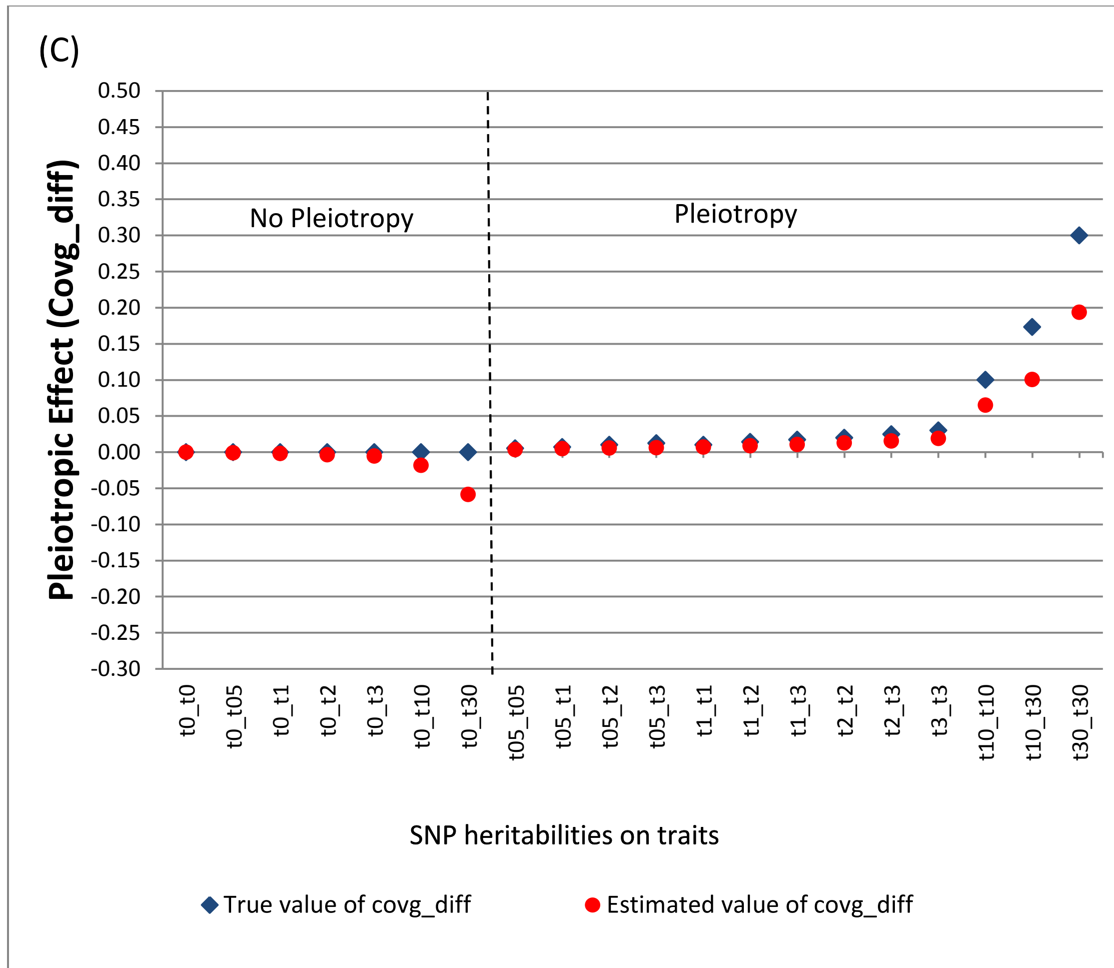

Figure 3.2 True and estimated values of the difference between polygenic covariances from bivariate polygenic models with and without adjustment for a SNP.

Three polygenic covariance generating values ($\mathbf{cov_{g^*}}$): 0.04 (A), 0.24 (B) and 0.36 (C) which correspond to the polygenic correlation generating values ($\boldsymbol{\rho_{g^*}}$): 0.1, 0.6 and 0.9. MAF = 0.5, $\boldsymbol{\rho_e = 0.6}$, $\boldsymbol{h^2_{r1^*} = h^2_{r2^*} = 0.4}$

In summary, polygenic covariance level also influences the genetic covariance approach. It produces greater bias under higher polygenic covariance level. Therefore, the VC-based genetic covariance approach is more efficient when the polygenic covariance is low.

3.5 Comparison between Polygenic Genetic Correlation and Genetic Covariance Approaches

Comparison of Genetic Correlation and Covariance Approaches in Bootstrap Empirical Distribution

In order to examine the performance of the genetic correlation and genetic covariance approaches on SNP-specific pleiotropic effect, we compared their bootstrap empirical type I error and power under low and moderate residual polygenic correlation/covariance levels (ρ_{g^*}=0.1, 0.6 and cov_{g^*}=0.04, 0.24) (Table 3.2). We did not do these for high level (ρ_{g^*}=0.9 and cov_{g^*}=0.36) because it has the similar pattern in estimate to the moderate level, but worse. So it is unnecessary to consider this condition if both approaches is demonstrated to be ineffective in moderate level. As mentioned before, because of the computation time, both approaches are applied only to selected SNP heritabilities (t0_t0, t0_t3, t0_t10, t0_t30, t1_t1, t3_t3, t10_t10, t10_t30 and t30_t30). The minor allele frequency was fixed at 50% and the environmental correlation was fixed at 0.6 whose impact of different values would be discussed later in the chapter.

The table also displays the bootstrap empirical distributions of ρ_{g_diff} and cov_{g_diff} on the descriptive statistics of mean, standard deviation and percent bias.

Under a low residual polygenic genetic correlation level, for almost all scenarios, the genetic covariance approach has smaller estimate bias than the genetic correlation approach. When a SNP has no effect on either trait, both the genetic correlation and genetic covariance approaches almost never reject the null hypothesis at a nominal $\alpha = 0.05$. When a SNP has effect on only one trait, both approaches have somewhat inflated type I error rates, which is smaller in the genetic covariance approach. When a SNP has effect on both traits, both approaches produce similar power.

Under a moderate polygenic correlation level, when a SNP has effect on one trait, we found that type I error rates of both approaches are very large. More specifically, the genetic correlation approach performs worse than the genetic covariance approach. When a SNP has effects on both traits, the absolute values of percent biases from the genetic covariance approach become larger than those from the genetic correlation approach.

Table 3. 2 Empirical bootstrap distribution comparing the genetic correlation and covariance approaches

MAF=0.5, polygenic $\rho_{g^*} = 0.1, cov_{g^*} = 0.04$, $\rho_e = 0.6,\ h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 500

Scenario	Genetic Correlation Approach (ρ_{g_diff})					Genetic Covariance Approach (cov_{g_diff})				
	True	Est	SD	Bias(%)	Type I Error /Power	True	Est	SD	Bias(%)	Type I Error /Power
t0_t0	0.000	0.000	0.006	---[a]	0.00	0.000	0.000	0.002	---[a]	0.00
t0_t3	0.000	-0.002	0.014	---[a]	0.08	0.000	0.000	0.005	---[a]	0.07
t0_t10	0.000	-0.009	0.025	---[a]	0.10	0.000	-0.001	0.010	---[a]	0.07
t0_t30	0.000	-0.024	0.043	---[a]	0.12	0.000	-0.006	0.017	---[a]	0.09
t1_t1	0.022	0.023	0.013	6.7	0.55	0.010	0.010	0.005	-1.8	0.52
t3_t3	0.063	0.065	0.022	3.6	0.98	0.030	0.029	0.009	-3.4	0.98
t10_t10	0.180	0.206	0.042	14.3	0.99	0.100	0.098	0.017	-1.9	0.99
t10_t30	0.272	0.308	0.053	13.2	1.00	0.173	0.166	0.022	-4.2	1.00
t30_t30	0.386	0.477	0.061	23.7	1.00	0.300	0.289	0.027	-3.7	1.00

*$th^2_{q1}_th^2_{q2}$ denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2.

[a] --- indicates that the percent bias is not computed.

Table 3.2 (Continued) Empirical bootstrap distribution comparing the genetic correlation and covariance approaches

MAF=0.5, polygenic $\rho_{g^*} = 0.6, cov_{g^*} = 0.24$, $\rho_e = 0.6, h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 500

Scenario	Genetic Correlation Approach (ρ_{g_diff})					Genetic Covariance Approach (cov_{g_diff})				
	True	Est	SD	Bias(%)	Type I Error /Power	True	Est	SD	Bias(%)	Type I Error /Power
t0_t0	0.000	0.000	0.003	---[a]	0.00	0.000	0.000	0.002	---[a]	0.00
t0_t3	0.000	-0.021	0.012	---[a]	***0.50***	0.000	-0.003	0.005	---[a]	***0.13***
t0_t10	0.000	-0.069	0.021	---[a]	***0.94***	0.000	-0.012	0.010	---[a]	***0.27***
t0_t30	0.000	-0.185	0.035	---[a]	***1.00***	0.000	-0.040	0.017	---[a]	***0.64***
t1_t1	0.010	0.010	0.010	5.5	0.50	0.010	0.008	0.008	-22.9	0.45
t3_t3	0.028	0.029	0.015	2.2	0.85	0.030	0.022	0.011	-25.0	0.75
t10_t10	0.080	0.086	0.025	7.5	0.88	0.100	0.074	0.019	-26.3	0.84
t10_t30	0.121	0.104	0.031	-14.2	1.00	0.173	0.121	0.021	-29.9	1.00
t30_t30	0.171	0.210	0.035	22.8	1.00	0.300	0.226	0.026	-24.6	1.00

*$th^2_{q1}_th^2_{q2}$ denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates.

[a] --- indicates that the percent bias is not computed.

Comparison of Bootstrap Empirical Distribution and Simulation Distribution for the Genetic Correlation and Genetic Covariance Approaches

Next, we compared the bootstrap empirical distribution to the simulation distribution for the genetic correlation and genetic covariance approaches under low and moderate residual polygenic correlation/covariance (ρ_{g^*}=0.1, 0.6 and cov_{g^*}=0.04, 0.24) and selected SNP heritabilities (t0_t0, t0_t3, t0_t10, t0_t30, t1_t1, t3_t3, t10_t10, t10_t30 and t30_t30). Table 3.3 and Appendix Table S1 display the descriptive statistics of mean, standard deviation, 95% confidence interval (2.5 percentile and 97.5 percentile) and percent bias of the estimated ρ_{g_diff} and cov_{g_diff} from the simulation distribution and bootstrap empirical distribution, respectively, as well as type I error and power from bootstrap empirical distribution.

The bootstrap resampling procedure produces very similar results to the original simulations on the mean values for both approaches. Because the bootstrap resampling process introduces some additional variations, it has a little larger standard deviations and wider confidence intervals.

Due to the computational time and costs, we considered to compare the genetic correlation and covariance approaches under different residual polygenic correlation and environmental correlation levels by using the CIs of the simulation distribution instead of the bootstrap empirical distribution. It is appealing and worthy since directly using the distribution of simulations significantly reduces the time needed for computation from 69 days to 3 hours in each scenario.

Table 3. 3 Simulation distribution and empirical bootstrap distribution comparison for the genetic correlation and covariance approaches

MAF=0.5, polygenic $\rho_{g^*} = 0.1, cov_{g^*} = 0.04$, $\rho_e = 0.6,\ h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 500

Scenario	True Value	Genetic Correlation Approach (ρ_{g_diff}) at $\alpha = .05$										
		Simulation Distribution					Bootstrap Distribution					
		Est	SD	2.5th	97.5th	Bias(%)	Est	SD	2.5th	97.5th	Bias(%)	Type I Error /Power
t0_t0	0.000	0.000	0.002	-0.003	0.005	---[a]	0.000	0.006	-0.012	0.014	---[a]	0.00
t0_t3	0.000	-0.003	0.014	-0.029	0.025	---[a]	-0.002	0.014	-0.028	0.027	---[a]	0.08
t0_t10	0.000	-0.010	0.025	-0.056	0.040	---[a]	-0.009	0.025	-0.056	0.042	---[a]	0.10
t0_t30	0.000	-0.028	0.043	-0.106	0.058	---[a]	-0.024	0.043	-0.103	0.067	---[a]	0.12
t1_t1	0.022	0.022	0.012	0.003	0.051	2.4	0.023	0.013	0.002	0.052	6.7	0.55
t3_t3	0.063	0.066	0.021	0.029	0.112	4.7	0.065	0.022	0.029	0.117	3.6	0.98
t10_t10	0.180	0.198	0.038	0.133	0.284	10.2	0.206	0.042	0.133	0.299	14.3	0.99
t10_t30	0.272	0.299	0.051	0.210	0.408	10.1	0.308	0.053	0.217	0.423	13.2	1.00
t30_t30	0.386	0.469	0.058	0.372	0.590	21.5	0.477	0.061	0.371	0.613	23.7	1.00

[a] --- indicates that the percent bias is not computed.

Table 3.3 (Continued) Simulation distribution and empirical bootstrap distribution comparison for the genetic correlation and covariance approaches

MAF=0.5, polygenic $\rho_{g^*} = 0.1, cov_{g^*} = 0.04$, $\rho_e = 0.6,\ h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 500

Scenario	Genetic Covariance Approach(cov_{g_diff}) at $\alpha = .05$											
	True Value	Simulation Distribution					Bootstrap Distribution					
		Est	SD	2.5th	97.5th	Bias(%)	Est	SD	2.5th	97.5th	Bias(%)	Type I Error /Power
t0_t0	0.000	0.000	0.001	-0.001	0.002	---[a]	0.000	0.002	-0.005	0.006	---[a]	0.00
t0_t3	0.000	0.000	0.005	-0.011	0.011	---[a]	0.000	0.005	-0.010	0.011	---[a]	0.07
t0_t10	0.000	-0.002	0.010	-0.021	0.018	---[a]	-0.001	0.010	-0.021	0.018	---[a]	0.07
t0_t30	0.000	-0.006	0.018	-0.040	0.029	---[a]	-0.006	0.017	-0.040	0.028	---[a]	0.09
t1_t1	0.010	0.010	0.005	0.001	0.021	-4.2	0.010	0.005	0.001	0.021	-1.8	0.52
t3_t3	0.030	0.029	0.009	0.013	0.047	-3.7	0.029	0.009	0.013	0.047	-3.4	0.98
t10_t10	0.100	0.096	0.016	0.065	0.127	-3.9	0.098	0.017	0.066	0.134	-1.9	0.99
t10_t30	0.173	0.165	0.022	0.125	0.208	-4.7	0.166	0.022	0.124	0.210	-4.2	1.00
t30_t30	0.300	0.288	0.026	0.236	0.340	-4.0	0.289	0.027	0.238	0.342	-3.7	1.00

[a] --- indicates that the percent bias is not computed.

Comparison of Genetic Correlation and Covariance Approaches under Different Polygenic Correlation Values

We compared the distribution of 1,000 simulations of ρ_{g_diff} and cov_{g_diff} on the descriptive statistics of mean, standard deviation, 95% confidence interval (2.5 percentile and 97.5 percentile) and percent bias with all SNP heritabilities and under three different residual polygenic correlation conditions ($\rho_{g^*} = 0.1, 0.6, 0.9$ and $cov_{g^*} = 0.04, 0.24, 0.36$, respectively) at MAF of 0.5 and environmental correlation of 0.6 (Table 3.4).

When the residual polygenic genetic correlation is low, the pleiotropic SNP effect is underestimated in the genetic covariance approach and somewhat overestimated in the genetic correlation approach if a SNP has effects on both traits. On average, the absolute values of percent biases are smaller for the genetic covariance approach.

When the residual polygenic genetic correlation is moderate or high, the absolute values of biases are unacceptably large in the genetic covariance approach and increase as the residual polygenic correlation increases. In the genetic correlation approach, on average, the absolute values of biases of the genetic correlation approach also increase with the residual genetic correlation, some of which are very unstable and abnormally large under high residual genetic correlation level. Larger residual polygenic correlation produces more tendency to exclude 0 in the confidence intervals if a SNP has effect on only one trait and to include 0 in the confidence intervals if a SNP has effects on both traits, which can lead to higher type I error and lower power.

Table 3. 4 Simulation distribution comparing the genetic correlation and covariance approaches under different polygenic correlation levels

MAF=0.5, polygenic ρ_{g^*}=0.1 and $cov_{g^*} = 0.04$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$, simulation =1000

	ρ_{g_diff}						cov_{g_diff}					
scenario	**True**	**Est**	**SD**	**2.5th**	**97.5th**	**Bias(%)**	**True**	**Est**	**SD**	**2.5th**	**97.5th**	**Bias(%)**
t0_t0	0.000	0.000	0.002	-0.003	0.005	---[a]	0.000	0.000	0.001	-0.001	0.002	---[a]
t0_t05	0.000	0.000	0.006	-0.012	0.012	---[a]	0.000	0.000	0.002	-0.004	0.005	---[a]
t0_t1	0.000	-0.001	0.008	-0.015	0.017	---[a]	0.000	0.000	0.003	-0.006	0.007	---[a]
t0_t2	0.000	-0.002	0.011	-0.023	0.021	---[a]	0.000	0.000	0.004	-0.008	0.009	---[a]
t0_t3	0.000	-0.003	0.014	-0.029	0.025	---[a]	0.000	0.000	0.005	-0.011	0.011	---[a]
t0_t10	0.000	-0.010	0.025	-0.056	0.040	---[a]	0.000	-0.002	0.010	-0.021	0.018	---[a]
t0_t30	0.000	-0.028	0.043	-0.106	0.058	---[a]	0.000	-0.006	0.018	-0.040	0.029	---[a]
t05_t05	0.011	0.011	0.009	-0.002	0.033	2.8	0.005	0.005	0.004	-0.001	0.013	-2.2
t05_t1	0.016	0.016	0.010	-0.001	0.038	1.4	0.007	0.007	0.004	0.000	0.016	-3.7
t05_t2	0.022	0.022	0.013	-0.001	0.051	-1.0	0.010	0.010	0.006	0.000	0.021	-4.5
t05_t3	0.027	0.026	0.015	-0.001	0.058	-3.1	0.012	0.012	0.006	***0.000***	***0.025***	-5.5
t1_t1	0.022	0.022	0.012	***0.003***	***0.051***	2.4	0.010	0.010	0.005	***0.001***	***0.021***	-4.2
t1_t2	0.031	0.031	0.014	***0.005***	***0.063***	1.7	0.014	0.014	0.006	***0.002***	***0.026***	-4.5
t1_t3	0.037	0.038	0.017	***0.010***	***0.076***	0.8	0.017	0.016	0.007	***0.004***	***0.031***	-5.1
t2_t2	0.043	0.045	0.017	***0.017***	***0.081***	4.3	0.020	0.019	0.007	***0.007***	***0.034***	-3.1
t2_t3	0.052	0.054	0.018	***0.021***	***0.094***	4.5	0.024	0.024	0.008	***0.009***	***0.040***	-3.2
t3_t3	0.063	0.066	0.021	***0.029***	***0.112***	4.7	0.030	0.029	0.009	***0.013***	***0.047***	-3.7
t10_t10	0.180	0.198	0.038	***0.133***	***0.284***	10.2	0.100	0.096	0.016	***0.065***	***0.127***	-3.9
t10_t30	0.272	0.299	0.051	***0.210***	***0.408***	10.1	0.173	0.165	0.022	***0.125***	***0.208***	-4.7
t30_t30	0.386	0.469	0.058	***0.372***	***0.590***	21.5	0.300	0.288	0.026	***0.236***	***0.340***	-4.0

* Italic bold values are the CIs excluding 0. [a] --- indicates that the percent bias is not computed.

Table 3.4 (Continued) Simulation distribution comparing the genetic correlation and covariance approaches under different polygenic correlation levels

MAF=0.5, polygenic ρ_{g^*}=0.6 and $cov_{g^*} = 0.24$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$, simulation =1000

	ρ_{g_diff}						cov_{g_diff}					
scenario	**True**	**Mean**	**Std**	**2.5th**	**97.5th**	**Bias(%)**	**True**	**Mean**	**Std**	**2.5th**	**97.5th**	**Bias(%)**
t0_t0	0.000	0.000	0.001	-0.003	0.003	---[a]	0.000	0.000	0.001	-0.001	0.002	---[a]
t0_t05	0.000	-0.004	0.005	-0.014	0.005	---[a]	0.000	-0.001	0.002	-0.004	0.005	---[a]
t0_t1	0.000	-0.008	0.007	-0.020	0.005	---[a]	0.000	-0.001	0.003	-0.006	0.006	---[a]
t0_t2	0.000	-0.015	0.009	-0.032	0.002	---[a]	0.000	-0.002	0.004	-0.010	0.007	---[a]
t0_t3	0.000	-0.022	0.011	***-0.045***	***-0.001***	---[a]	0.000	-0.004	0.005	-0.013	0.007	---[a]
t0_t10	0.000	-0.069	0.020	***-0.107***	***-0.029***	---[a]	0.000	-0.012	0.010	-0.030	0.008	---[a]
t0_t30	0.000	-0.182	0.034	***-0.248***	***-0.113***	---[a]	0.000	-0.039	0.017	***-0.070***	***-0.005***	---[a]
t05_t05	0.005	0.005	0.004	-0.002	0.015	2.7	0.005	0.004	0.004	-0.002	0.013	-21.1
t05_t1	0.007	0.007	0.006	-0.003	0.019	-4.7	0.007	0.006	0.004	-0.002	0.016	-21.9
t05_t2	0.010	0.006	0.007	-0.008	0.022	-34.8	0.010	0.007	0.005	-0.003	0.019	-28.1
t05_t3	0.012	0.005	0.009	-0.013	0.024	-60.7	0.012	0.008	0.006	-0.003	0.022	-32.9
t1_t1	0.010	0.010	0.006	***0.001***	***0.024***	3.8	0.010	0.008	0.005	-0.001	0.019	-21.4
t1_t2	0.014	0.013	0.008	0.000	0.030	-6.5	0.014	0.011	0.006	0.000	0.023	-24.3
t1_t3	0.017	0.013	0.009	-0.004	0.032	-19.9	0.017	0.013	0.007	***0.000***	***0.028***	-26.7
t2_t2	0.019	0.019	0.008	***0.006***	***0.039***	2.1	0.020	0.015	0.007	***0.003***	***0.030***	-23.7
t2_t3	0.023	0.023	0.009	***0.007***	***0.045***	-0.5	0.024	0.019	0.008	***0.005***	***0.034***	-24.4
t3_t3	0.028	0.029	0.010	***0.011***	***0.052***	2.8	0.030	0.023	0.009	***0.008***	***0.041***	-24.5
t10_t10	0.080	0.088	0.020	***0.055***	***0.136***	10.0	0.100	0.077	0.015	***0.049***	***0.107***	-23.3
t10_t30	0.121	0.106	0.030	***0.053***	***0.173***	-12.3	0.173	0.125	0.020	***0.088***	***0.166***	-27.9
t30_t30	0.171	0.210	0.034	***0.154***	***0.283***	22.2	0.300	0.229	0.026	***0.179***	***0.282***	-23.5

* Italic bold values are the CIs excluding 0. [a] --- indicates that the percent bias is not computed.

Table 3.4 (Continued) Simulation distribution comparing the genetic correlation and covariance approaches under different polygenic correlation levels

MAF=0.5, polygenic ρ_{g^*}=0.9 and $cov_{g^*} = 0.36$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$, simulation =1000

	ρ_{g_diff}						cov_{g_diff}					
scenario	**True**	**Mean**	**Std**	**2.5th**	**97.5th**	**Bias(%)**	**True**	**Mean**	**Std**	**2.5th**	**97.5th**	**Bias(%)**
t0_t0	0.000	0.000	0.001	-0.002	0.001	---[a]	0.000	0.000	0.002	-0.001	0.002	---[a]
t0_t05	0.000	-0.006	0.004	-0.015	0.001	---[a]	0.000	-0.001	0.003	-0.004	0.005	---[a]
t0_t1	0.000	-0.012	0.006	***-0.024***	***-0.001***	---[a]	0.000	-0.002	0.003	-0.007	0.005	---[a]
t0_t2	0.000	-0.023	0.008	***-0.039***	***-0.008***	---[a]	0.000	-0.004	0.004	-0.010	0.006	---[a]
t0_t3	0.000	-0.033	0.010	***-0.057***	***-0.015***	---[a]	0.000	-0.005	0.005	-0.014	0.005	---[a]
t0_t10	0.000	-0.104	0.019	***-0.144***	***-0.068***	---[a]	0.000	-0.018	0.009	-0.036	0.001	---[a]
t0_t30	0.000	-0.275	0.034	***-0.345***	***-0.210***	---[a]	0.000	-0.059	0.016	***-0.090***	***-0.025***	---[a]
t05_t05	0.001	0.001	0.001	-0.001	0.004	-16.1	0.005	0.003	0.004	-0.003	0.012	-35.3
t05_t1	0.002	0.001	0.002	-0.004	0.005	-67.6	0.007	0.004	0.004	-0.003	0.014	-36.5
t05_t2	0.002	-0.003	0.005	-0.013	0.005	-233.3	0.010	0.006	0.005	-0.004	0.017	-44.0
t05_t3	0.003	-0.009	0.006	-0.022	0.003	-387.2	0.012	0.006	0.006	-0.005	0.020	-51.1
t1_t1	0.002	0.002	0.002	-0.001	0.006	-6.0	0.010	0.007	0.005	-0.002	0.017	-33.8
t1_t2	0.003	0.001	0.003	-0.005	0.008	-59.1	0.014	0.009	0.006	-0.001	0.021	-37.6
t1_t3	0.004	-0.002	0.005	-0.012	0.007	-140.7	0.017	0.010	0.007	-0.001	0.025	-41.0
t2_t2	0.005	0.005	0.003	***0.001***	***0.011***	-4.2	0.020	0.013	0.007	***0.001***	***0.027***	-36.2
t2_t3	0.006	0.005	0.004	-0.002	0.012	-20.2	0.024	0.015	0.008	***0.002***	***0.031***	-36.8
t3_t3	0.007	0.007	0.003	***0.002***	***0.014***	-1.6	0.030	0.019	0.008	***0.004***	***0.037***	-36.4
t10_t10	0.020	0.021	0.008	***0.007***	***0.041***	6.6	0.100	0.065	0.015	***0.038***	***0.094***	-35.2
t10_t30	0.030	-0.012	0.017	-0.045	0.026	-138.5	0.173	0.100	0.020	***0.064***	***0.140***	-42.1
t30_t30	0.043	0.052	0.017	***0.021***	***0.088***	21.1	0.300	0.194	0.025	***0.147***	***0.244***	-35.5

* Italic bold values are the CIs excluding 0. [a] --- indicates that the percent bias is not computed.

Comparison of Genetic Correlation and Covariance Approaches under Different Environmental Correlation Values

To explore the effects of different environmental correlation levels on the performances of the genetic correlation and covariance approaches, we also compared the distribution of 1,000 simulations of ρ_{g_diff} and cov_{g_diff} on the descriptive statistics of mean, standard deviation, 95% confidence interval (2.5 percentile and 97.5 percentile) and percent bias under three environmental correlation conditions $\rho_e = 0.0,\ 0.6, 0.9$) at MAF of 0.5 and polygenic correlation of 0.1 (Appendix Table S2).

Consistent across the environmental correlation levels is that both approaches are effective under low polygenic correlation condition.

Environmental correlation slightly affects both approaches. The absolute values of estimated bias of the genetic covariance approach are in general larger under lower environmental correlation level, and the confidence intervals are more likely to exclude 0 when a SNP truly has a pleiotropic effect. In contrast, the estimated biases of the genetic correlation approach are smaller under lower environmental correlation level.

Comparison of Genetic Correlation and Covariance Approaches under Different Minor Allele Frequencies

We also explored 0.1 for minor allele frequency at the polygenic correlation of 0.1, and environmental correlation of 0.6 (Appendix Table S3), and compared it with minor allele frequency of 0.5 (Table 3.4).

The simulation distributions of both approaches generate very similar results at different magnitude of MAF.

3.6 Summary

We have examined two VC-based approaches of decomposing the share polygenic effects for the analysis of SNP-specific pleiotropic effects in bivariate models. One approach compares genetic correlations and the other approach compares genetic covariances from bivariate polygenic models with and without adjustment for a SNP. We compared these two approaches in the simulation study.

There was evidence that both approaches are affected by the residual polygenic genetic correlation level. Both approaches are recommended for low residual polygenic correlation condition. High polygenic correlation typically produce large bias, inflated type I error if a SNP has effect on one trait, and low power. Under low polygenic correlation condition, the genetic covariance approach has smaller estimate bias and type I error than the genetic correlation approach, especially when the SNP effects on traits are large. Both approaches produce similar powers.

We also explored other possible values for environmental correlation and minor allele frequency which do not alter the results materially.

Chapter 4: Covariance Analysis on Pleiotropic Effect

In this chapter, we propose a multivariate regression-based approach for population-based studies. First, the newly proposed approach is applied to single SNP pleiotropic effect analysis. Its performance is evaluated using simulation studies on bias, type I error and power. We also examine the influence of polygenic correlation, environmental correlation and minor allele frequency on this approach. Then we extend this approach to multiple SNPs pleiotropic effects.

4.1 Introduction

We have seen increasing application of linear modeling or path analysis to genetic problems in the literature. Studies of phenotypic covariance between quantitative traits have long suggested the presence of pleiotropy (Wright, 1977; Falconer and Mackay 1996; Flint and Mackay, 2009). Carey (1986) proposed a general multivariate approach to linear modeling in genetics:

$$C_{Y1Y2} = \sum_{i=1}^{m} \sum_{j=1}^{n} \beta_{1i} C_{AiAj} \beta_{2j}^{T} \tag{4.1}$$

where

$$Y_1 = \beta_{11}A_1 + \beta_{12}A_2 + \cdots + \beta_{1m}A_m, \tag{4.2}$$

$$Y_2 = \beta_{21}A_1 + \beta_{22}A_2 + \cdots + \beta_{2n}A_n,$$

are the multivariate linear models. Y_1 and Y_2 are the trait values; β_{1i} is the regression coefficient from the dependent variable Y_1 to the independent variable A_i; β_{2j} is the regression coefficient from the dependent variable Y_2 to the independent variable A_j;

C_{Y1Y2} is the covariance between Y_1 and Y_2; C_{AiAj} is the covariance between A_i and A_j, $i = 1, \cdots, m, j = 1, \cdots, n$.

Based on this idea, we developed a multivariate regression-based approach for pleiotropic effects analysis by computing genetic covariance from the effects of underlying loci, which we named, covariance analysis (CovA).

4.2 Single SNP Analysis

The CovA approach can link multiple markers together for two traits in population-based studies with samples of families or unrelated subjects. Here we focus our discussion on a single SNP. Similar to (4.2), we have bivariate linear models with two independent variables: *SNP* coded as 0, 1 and 2 for the number of minor alleles, and *V* representing a covariate. We assume $Cov(SNP, V) = 0$.

$$Y_1 = \mu_1 + \beta_{11}SNP + \beta_{12}V + e_1; \tag{4.4}$$

$$Y_2 = \mu_2 + \beta_{21}SNP + \beta_{22}V + e_2,$$

where μ_i is the mean trait of all individuals for the *i*th trait; e_i is a non-genetic (residual) effect, $E(e_i) = 0$, $Cov(e_i, e_j) = \sigma_{ij}$, $i, j = 1,2$.

4.2.1 Proposed Method

By using Carey's formula, we obtain the phenotypic covariance

$$C_{Y_1Y_2} = \hat{\beta}_{11}cov(SNP, SNP)\hat{\beta}_{21} + \hat{\beta}_{11}cov(SNP,\ V)\hat{\beta}_{22} + \hat{\beta}_{12}cov(SNP,\ V)\hat{\beta}_{21}$$

$$+\hat{\beta}_{12}cov(V, V)\hat{\beta}_{22} + cov(e_1, e_2). \tag{4.5}$$

Since we are only concerned with the covariance between traits attributable to the SNP effect, the pleiotropic effect is defined as

$$C_P = \hat{\beta}_{11} cov(SNP, SNP) \hat{\beta}_{21} = \hat{\beta}_{11} var(SNP) \hat{\beta}_{21}, \quad (4.6)$$

where $\hat{\beta}_{11}$ and $\hat{\beta}_{21}$ represent the estimated size of genetic effects of the SNP of interest on trait1 and trait2, respectively. $var(SNP)$ is the variance of the SNP.

The variance of this pleiotropic effect can be written as

$$var(C_P) = var\left[\hat{\beta}_{11} var(SNP) \hat{\beta}_{21}\right] = var^2(SNP) var(\hat{\beta}_{11}\hat{\beta}_{21}). \quad (4.7)$$

Goodman (1960) proposed a formula to compute the approximate variance of the product of two random variables where these variables are dependent

$$var(XY) = [E(X)]^2 var(Y) + [E(Y)]^2 var(X) + 2E(X)E(Y)cov(X, Y), \quad (4.8)$$

where $E(X)$ and $E(Y)$ denote the expected value of X and Y, and $var(X)$ and $var(Y)$ denote the variance of X and Y, respectively. $cov(X, Y)$ denotes the covariance between X and Y. Thus the variance of the product of regression coefficients is

$$var\left(\hat{\beta}_{11}\hat{\beta}_{21}\right) = \left[E\left(\hat{\beta}_{11}\right)\right]^2 var\left(\hat{\beta}_{21}\right) + \left[E\left(\hat{\beta}_{21}\right)\right]^2 var\left(\hat{\beta}_{11}\right)$$

$$+2E\left(\hat{\beta}_{11}\right)E\left(\hat{\beta}_{21}\right)cov\left(\hat{\beta}_{11}, \hat{\beta}_{21}\right). \quad (4.9)$$

Under the null hypothesis of no pleiotropy, there should be no covariance attributable to the SNP. The alternative is that the SNP takes up some of the covariation between traits. Thus, these traits are affected by the SNP. Symbolically, these hypotheses can be expressed as

$$H_0: C_P = 0 \ (\beta_{11} = 0 \ or \ \beta_{21} = 0);$$

$$H_a: C_P \neq 0 \ (\beta_{11} \neq 0 \ and \ \beta_{21} \neq 0).$$

If the distribution of test statistic under the null hypothesis approximately follows a standard normal distribution, we can use a test for the statistical testing

$$Z_{stat} = \frac{\hat{C}_P}{\sqrt{var(\hat{C}_P)}}$$

$$= \frac{\hat{\beta}_{11} var(SNP) \hat{\beta}_{21}}{\sqrt{var^2(QTL)[[E(\hat{\beta}_{11})]^2 var(\hat{\beta}_{21}) + [E(\hat{\beta}_{21})]^2 var(\hat{\beta}_{11}) + 2E(\hat{\beta}_{11})E(\hat{\beta}_{21})cov(\hat{\beta}_{11}, \hat{\beta}_{21})]}} \tag{4.10}$$

4.2.2 Evaluation of the Proposed Method Using Simulation Studies

CovA is a more general approach, which can be used for population-based studies with samples of families or unrelated subjects. We start our investigation from unrelated data. Family relationship is discussed later in this chapter. We conducted simulation studies using independent observations and a single SNP to evaluate the performance of the CovA approach.

A dataset of one thousand family trios were generated Then two normally distributed quantitative traits and a single di-allelic SNP with MAF of 10% or 50% were simulated using the “simqtl” command in SOLAR. The simulation designs set up twelve conditions in respect to residual polygenic correlation (ρ_g) of 0.0, 0.1 0.6 and 0.9, environmental correlation (ρ_e) of 0.0, 0.6 and 0.9; and residual heritability (h^2_{r1} and h^2_{r2}) of 1% for ρ_g = 0.0 and 40% for others. Six SNP heritabilities were considered (h^2_{qtl} =0%, 1%, 2%, 3%, 10% and 30% giving the size of effect for each SNP as 0,

0.1414, 0.2000, 0.2450, 0.4472 and 0.7746 units, and the standard deviation for each SNP as 1, 0.995, 0.990, 0.985, 0.949 and 0.837, respectively). Twenty-one scenarios were evaluated according to SNP heritability pairs (Table 4.1). We performed 1,000 replicates for each scenario to increase the accuracy of the estimates.

Table 4. 1 Pairs of SNP effects on bivariate traits in simulations of CovA.

No pleiotropy (No effects on T1 and T2)	t0_t0
No pleiotropy(No effect on T1)	t0_t1, t0_t2, t0_t3, t0_t10, t0_t30
Pleiotropic effect on both T1 and T2	t1_t1, t1_t2, t1_t3, t1_t10, t1_t30, t2_t2, t2_t3, t2_t10, t2_t30, t3_t3, t3_t10, t3_t30, t10_t10, t10_30, t30_t30

*t$\boldsymbol{h}^2_{q1}$_t$\boldsymbol{h}^2_{q2}$ denotes SNP with $\boldsymbol{h}^2_{q1}$% effect on trait1 and $\boldsymbol{h}^2_{q2}$% effect on trait2.

We randomly selected one subject from each family, and then had 1,000 unrelated subjects in each sample. Using the subset data, we conducted multivariate regression modeling (4.11) to obtain the regression coefficients.

$$Y_1 = \mu_1 + \beta_1 SNP + e_1, \quad (4.11)$$

$$Y_2 = \mu_2 + \beta_2 SNP + e_2.$$

The MODEL procedure (Statistical Analysis System (SAS) Institute Inc., Cary, NC, USA) is a tool to analyze the structure of simultaneous equations. It can specify and estimate the covariance and correlation structure of the parameter estimates which is an

essential element in computation of variance (4.3). We used the Full Information Maximum Likelihood (FIML) option for parameter estimation which assumes that the equation errors have a multivariate normal distribution. The pleiotropic effect is computed as the product of regression coefficients and variance of the SNP of interest by using the CovA approach. The simulation procedure is summarized as follows:

Simulation Steps:

Step1: Generate 1,000 uncorrelated trios (2 parents and a child);

Step2: Simulate a dataset with two normally distributed quantitative traits and a di-allelic SNP using the pedigree structure in Step1. The simulation designs include 21 SNP heritability scenarios, 4 residual polygenic correlation and 3 environmental correlation conditions;

Step3: Randomly select 1,000 independent subjects as a subset of the total (1 subject from each family);

Step4: Conduct the multivariate regression analysis;

Step5: Compute the covariance attributable to a SNP effect: $\hat{C}_{P1}$ and its variance $var_C(\hat{C}_{P1})$. Calculated the test statistic $Z_{C_stat1} = \hat{C}_{P1} / \sqrt{var_C(\hat{C}_{P1})}$;

Step6: Repeat Steps 2-5 1,000 times for 1,000 replicates;

Step7: Compute the mean of $\hat{C}_{P1}, \cdots, \hat{C}_{P1000}$ as $\hat{C}_P$, the mean of $var_C(\hat{C}_{P1}), \cdots, var_C(\hat{C}_{P1000})$ as $var_C(\hat{C}_P)$, and the variance of $\hat{C}_{P1}, \cdots, \hat{C}_{P1000}$ as $var(\hat{C}_P)$.

4.2.3 Simulation Results

First, we present the simulation results of moderate residual polygenic and environmental correlation levels (ρ_g=0.6 and ρ_e=0.6) at 50% MAF as an example to explicate the process to implement CovA approach. Next, we explore the impact of different residual polygenic and environmental correlation levels on the performance of the CovA approach. Twelve conditions are established for residual polygenic (ρ_{g^*}) of 0.0, 0.1, 0.6 and 0.9 and environmental correlation (ρ_e) of 0.1, 0.6 and 0.9 at MAF of 50%. Finally, we examine the effect of different MAFs on CovA by comparing the simulation results of MAF of 10% and 50%.

Bias of the Estimator

Figure 4.1 displays the estimated pleiotropic effect across SNP heritability pairs using CovA. In order to explicitly reflect the relationship between the SNP-specific pleiotropic effect and its individual effects on traits, we also presented the estimated SNP R^2 on trait1 and trait2, respectively. The estimated pleiotropic SNP effect and individual effects are unbiased. When a SNP has no effect on at least one trait, the estimated pleiotropic effect is always zero. When a SNP has effect on both traits, the value of pleiotropic effect lies between the individual SNP effects. For example, the estimated pleiotropic effect was 1.824% when a SNP had 1% effect on trait1 and 3% effect on trait2. The estimated SNP-specific pleiotropic effect strongly depends on the joint effect of the proportion of phenotypic variances attributable to the SNP.

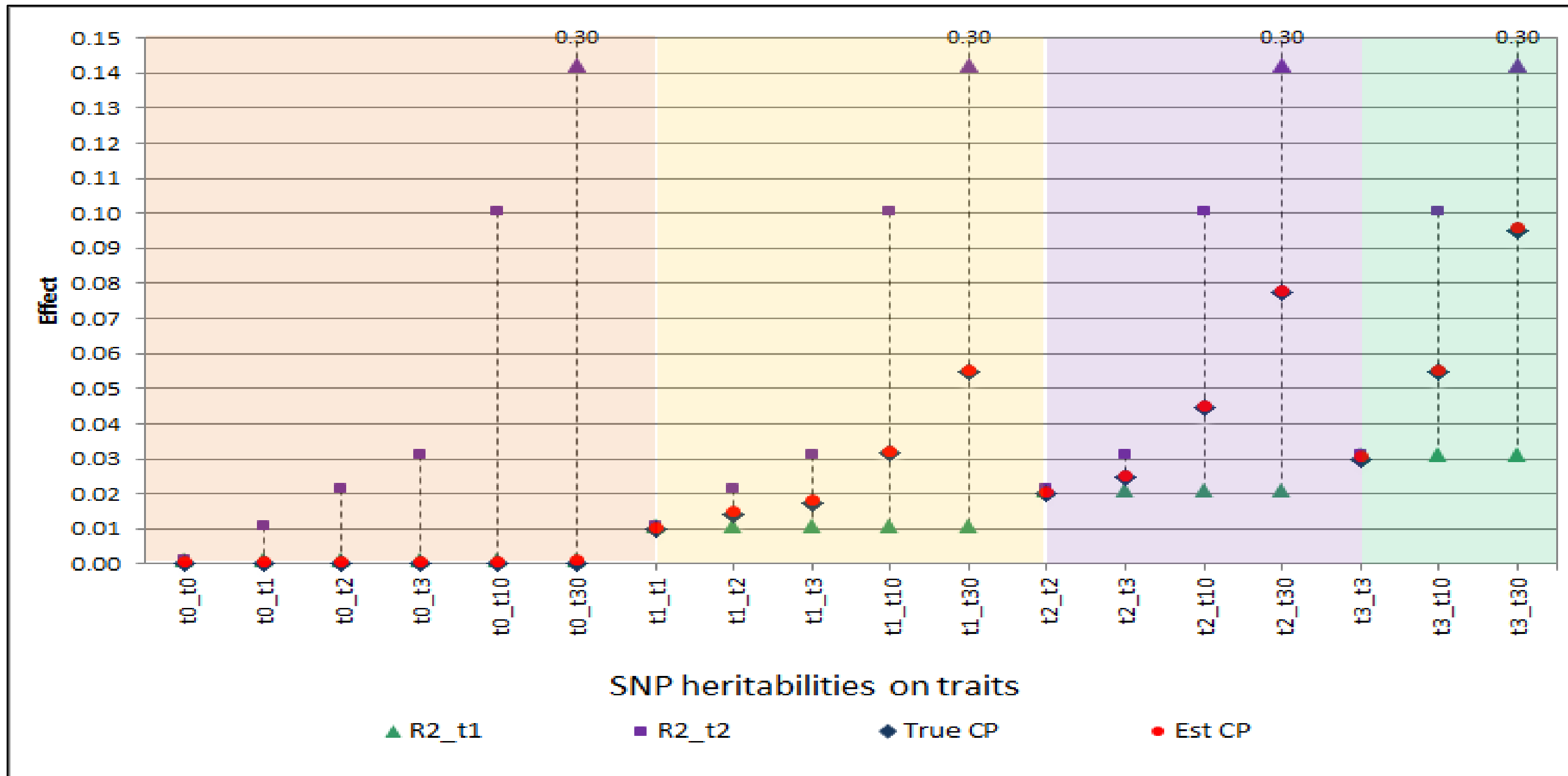

Figure 4. 1 Pleiotropic SNP effect and individual SNP effects on traits using CovA

MAF=0.5, polygenic $\boldsymbol{\rho_{g^*} = 0.6}$, $\boldsymbol{\rho_e = 0.6}$, $\boldsymbol{h^2_{r1^*} = h^2_{r2^*} = 0.4}$.

* $\boldsymbol{th^2_{q1}}$_t$\boldsymbol{h^2_{q2}}$ denotes SNP with $\boldsymbol{h^2_{q1}}$% effect on trait1 and $\boldsymbol{h^2_{q2}}$% effect on trait2. R2_t1 and R2_t2 denote the estimated SNP regression $\boldsymbol{R^2}$s on trait1 and trait2, respectively. C_P is the pleiotropic SNP effect using CovA.

Distribution of Test Statistic

Casella and Berger (2001) assert that we have to determine the sampling distribution of a test statistic under the null hypothesis in statistical hypothesis testing, allowing us to calculate a p-value. Based on the central limit theorem, if a test statistic approximately follows a normal distribution for large samples, Z-test will be a good choice for the hypothesis testing. Thus, we used histograms and Q-Q plots to examine the sampling distribution of the test statistic under the six null condition scenarios from our simulations (0%_0%, 0%_1%, 0%_2%, 0%_3%, 0%_10% and 0%_30%.).

The histogram and density curve display the variation of Z_{stat} in the simulation samples (Figure 4.2). When a SNP has no effect on either trait, the distribution possesses a stronger peak, more rapid decay, and lighter tails than the normal density (Figure 4.2 (A)). In contrast, when a SNP has effect on only one trait, the samples are approximately normally distributed (Figure 4.2 (B)-(F)). In addition, as the SNP effect on trait2 becomes larger, the sampling distributions become closer to a standard normal distribution.

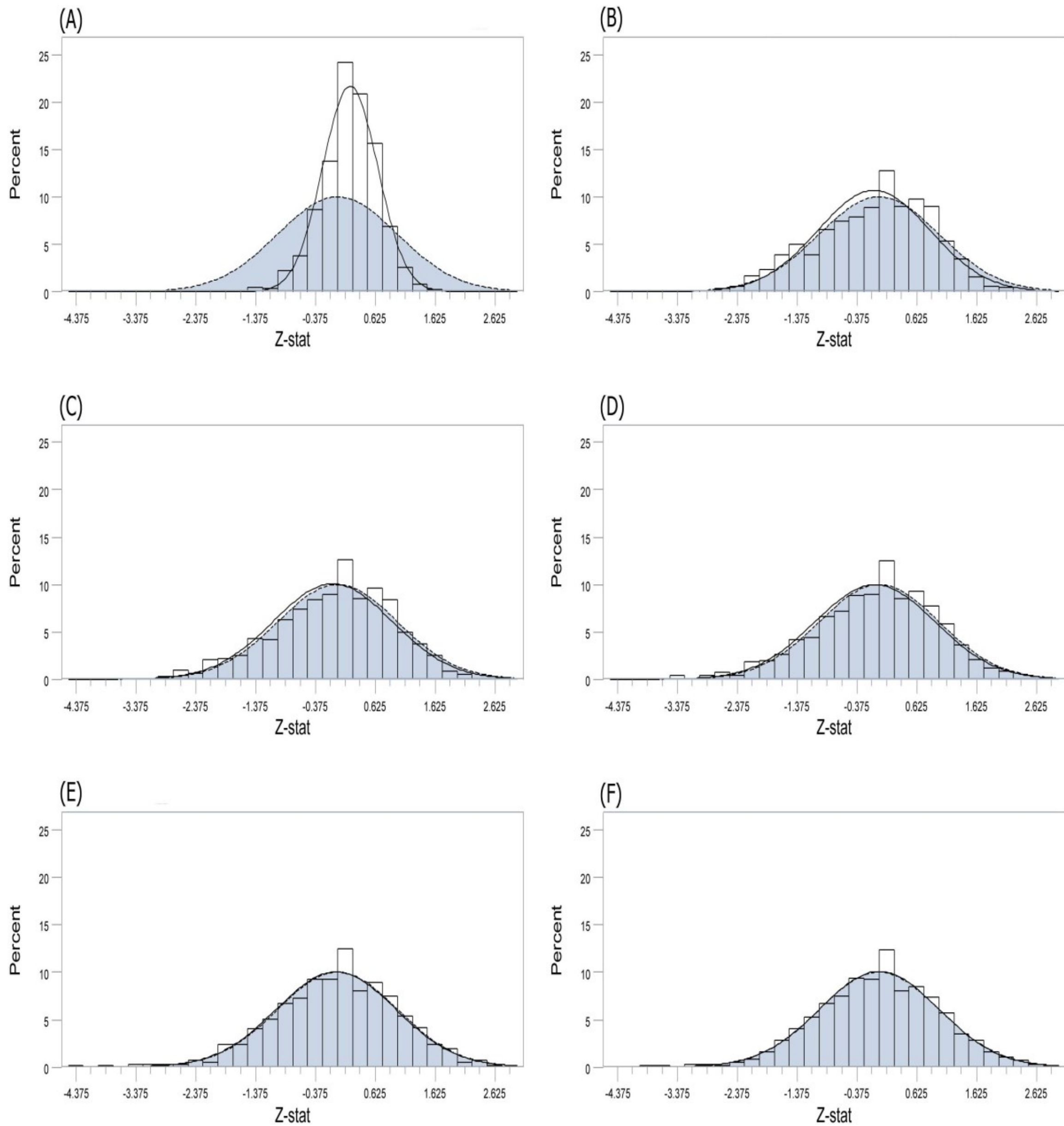

Figure 4. 2 Histogram examining the distribution of test statistic in CovA using 1000 simulation samples

Plots are shown for six null condition scenarios: (A) 0%_0%, (B) 0%_1%, (C) 0%_2%, (D) 0%_3%, (E) 0%_10% and (F) 0%_30%, where MAF=0.5, polygenic $\boldsymbol{\rho_{g^*} = 0.6}$, $\boldsymbol{\rho_e = 0.6}$, $\boldsymbol{h^2_{r1^*} = h^2_{r2^*} = 0.4}$. The line indicates the empirical probability density function of the test statistic. Shadow area indicates the theoretical cumulative distribution function of the test statistic under sampling from a fitted standard normal distribution.

Q-Q plots are an informative approach to examine deviation, comparing our simulated data to a theoretical distribution by plotting their quantiles against each other (Wilk and Gnanadesikan 1986). We employed a Q-Q plot of p-values from Z_{stat} by matching each observed p-value with an expected p-value, then the observed and expected p-values were transformed into –log10(p-value) because we were most interested in the smallest p-value. If the empirical distribution of Z_{stat} is similar to the standard normal distribution, the points in the Q-Q plot should approximately lie on the line, $y = x$. The non-linearity of the points in a plot indicates a departure from normality.

When a SNP has no effect on either trait, points are below the line, indicating short tails at both ends of the data distribution (Figure 4.3 (A)). Thus, simulated data centralizes in a small range with small variation. It is very clear that Z_{stat} does not follow a standard normal distribution and would generate a very conservative type I error rate by using a normal approximation method. When a SNP has effect on one trait, majority of points reassuringly fit the line, $y = x$, indicating an approximately standard normal distribution (Figure 4.3 (B)-(F)). When a SNP has 0% effect on trait1 and 1% effect on trait2, the rest of points are below the line showing evidence of a little conservative type I error (Figure 4.3 (B)). When the SNP effect is greater than 1% on trait2, the rest of points are above the line, resulting in an inflated type I error (Figure 4.3 (C)-(F)).

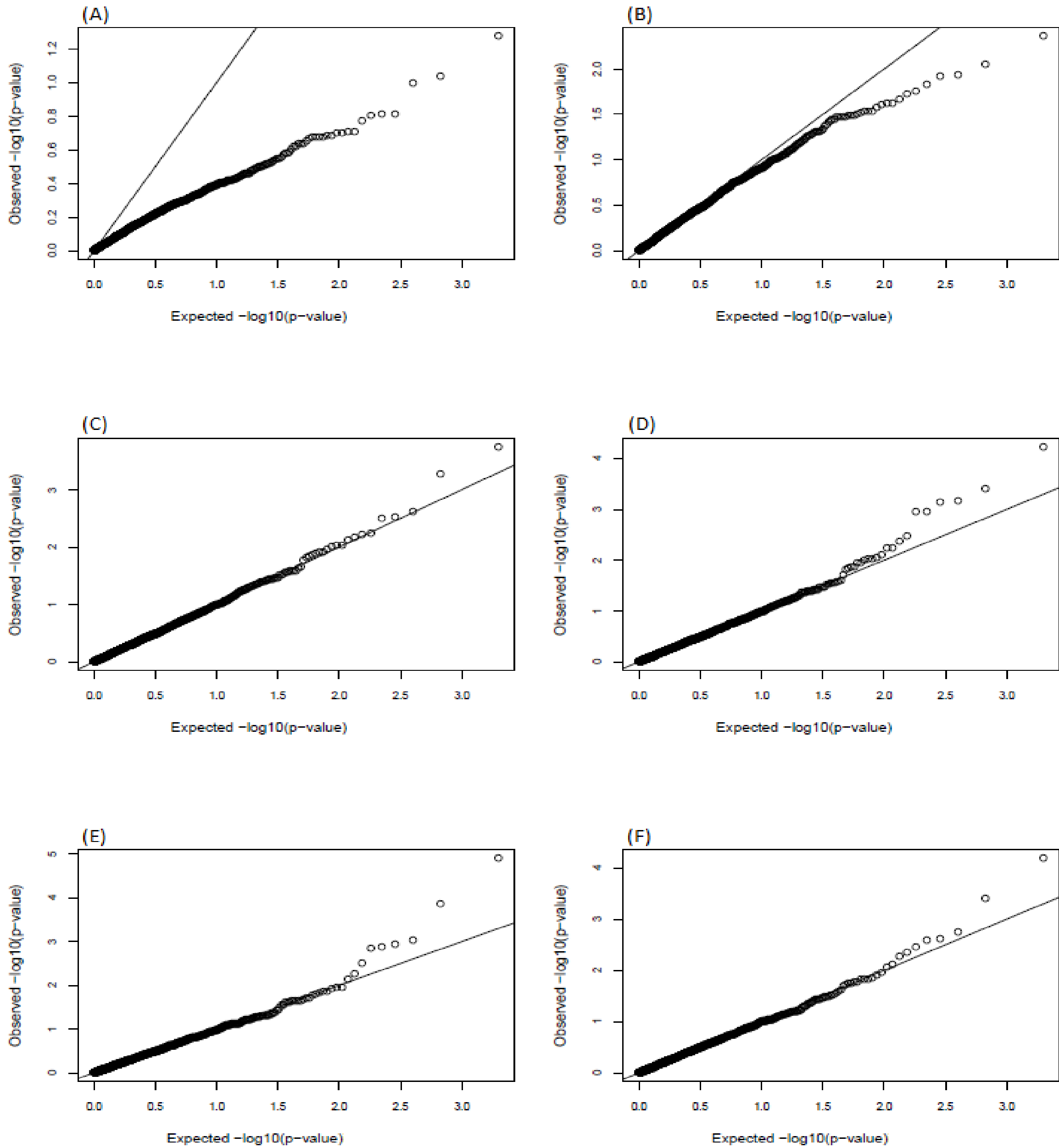

Figure 4. 3 Q-Q plot comparing the distribution of test statistic in CovA using 1000 simulation samples to a standard normal distribution

Plots are shown for six null condition scenarios: (A) 0%_0%, (B) 0%_1%, (C) 0%_2%, (D) 0%_3%, (E) 0%_10% and (F) 0%_30% where MAF=0.5, polygenic $\boldsymbol{\rho_{g^*}} = \mathbf{0.6}$, $\boldsymbol{\rho_e} = \mathbf{0.6}$, $\boldsymbol{h^2_{r1^*}} = \boldsymbol{h^2_{r2^*}} = \mathbf{0.4}$. Line indicates the expected –log10(p-values) of a standard normal distribution. Points indicate the actual –log10(p-values) of an empirical distribution of $\boldsymbol{Z_{stat}}$.

In sum, when a SNP has no effect on either trait, the asymptotic distribution violates the normality assumption. It can be dangerous to use a Z-test in such analysis, where in this situation it would result in a very conservative type I error rate. When a SNP has effect on one trait, the sampling distributions generally obey the normality assumption. Depending on how much the SNP effect on the other trait is presented, there are two cases. When a SNP has 0% effect on trait1 and 1% effect on trait2, a normality-based method generates a conservative type I error; while when a SNP has >1% effect on trait2, a normality-based method generates an inflated type I error. Above all, the distributions of test statistic under the null hypothesis are inconsistent, that is, there are different error rates under different scenarios. In reality, we cannot know the distribution in advance. Therefore, in this situation, methods based on normality are inappropriate and invalid.

Violation of Normality Assumption and Bootstrap Resampling Method

We proposed to implement a resampling method that can be safely applied when the sampling distributions are different from normal distribution under the null. Resampling methods are popular used statistical tools in analysis. These methods involve either sampling or scrambling (randomization test) the original data numerous times. The bootstrap resampling method was used in this study.

During the resampling procedure, we assume that the sub-samples come from the same distribution of population, but each sample drawn independently from other samples. It means two assumptions must be satisfied in bootstrapping method. First, the

bootstrap sample mimics the general distribution of the original population. Second, each sub-sample should be independently and identically distributed (i.i.d.) (Hesterberg et al. 2005).

The advantage of bootstrap method is its great simplicity and convenience. Usually it is difficult to derive estimates of standard errors and confidence intervals for complex parameters of the distribution, such as proportions, odds ratio, and correlation coefficients, but it is not a problem for the bootstrap. However, it is time-consuming. It is feasible in practice only with software that automates the heavy computation (Hesterberg et al. 2005).

The implementation of CovA using bootstrap resampling method proceeded according to the following steps:

Bootstrap Steps:

Step1: Use the same 1,000 simulated data from simulation procedure;

Step2: Draw sample $x_1^*, \cdots, x_{1000}^*$ from replicate 1 with replacement;

Step3: Conduct the multivariate regression analysis;

Step4: Compute the covariance attributable to the SNP effect $\hat{C}_{P(1,1)}^*$ by formula (4.6) and its variance $var_C(\hat{C}_{P(1,1)}^*)$ by formula (4.7) using $x_1^*, \cdots, x_{1000}^*$;

Step5: Repeat Steps 2-4 B=1,000 times. With a large number of new samples, generate an empirical sampling distribution for C_{P1}: $\{\hat{C}_{P(1,1)}^*, \cdots, \hat{C}_{P(1,1000)}^*\}$;

Step6: Compute the mean of $\hat{C}^*_{P(1,1)}, \cdots, \hat{C}^*_{P(1,1000)}$ as $\hat{C}^*_{P1}$, the mean of $var_C(\hat{C}^*_{P(1,1)}), \cdots, var_C(\hat{C}^*_{P(1,1000)})$ as $var_C(\hat{C}^*_{P1})$ and the variance of $\hat{C}^*_{P(1,1)}, \cdots, \hat{C}^*_{P(1,1000)}$ as $var_B(\hat{C}^*_{P1})$. Calculate the test statistic $Z_{B_stat1} = \hat{C}_{P1} / \sqrt{var_B(\hat{C}^*_{P1})}$;

Step7: Construct the 95% confidence interval from the empirical sampling distribution of $\{\hat{C}^*_{P(1,1)}, \cdots, \hat{C}^*_{P(1,1000)}\}$ as $(u, v)_1$;

Step8: Repeat Steps 2-7 1,000 times for 1,000 replicates, resulting in 1,000 confidence intervals $\{(u, v)_1, \cdots, (u, v)_{1000}\}$ and 1,000 test statistics $Z_{B_stat1}, \cdots, Z_{B_stat1000}$;

Step9: Determine if these confidence intervals cover 0 or not, and calculate the proportion of time the 95% confidence intervals excludes 0;

Step10: Compute the mean of $\hat{C}^*_{P1}, \cdots \hat{C}^*_{P1000}$ as $\hat{C}^*_{P}$, the mean of $var_B(\hat{C}^*_{P1}), \cdots, var_B(\hat{C}^*_{P1000})$ as $var_B(\hat{C}^*_{P})$, the mean of $var_C(\hat{C}^*_{P1}), \cdots, var_C(\hat{C}^*_{P1000})$ as $var_C(\hat{C}^*_{P})$, and the variance of $\hat{C}^*_{P1}, \cdots \hat{C}^*_{P1000}$ as $var(\hat{C}^*_{P})$.

The bootstrap resampling distribution of Z_{B_stat} ($Z_{B_stat1}, \cdots, Z_{B_stat1000}$) is portrayed by Q-Q plot (Figure 4.4) using the –log10(p-value) of Z_{B_stat} where the estimated variance used in test statistic is computed from bootstrapping, rather than our derived formula.

Hesterberg et al. (2005) explained that a statistic would vary from sample to sample, so this random variation needs to be considered in inference about population. The variation in a statistic due to randomly sample selections is displayed by the sampling distribution of the statistic. For example, the margin of error in a confidence interval expresses the uncertainty due to sampling variation. The empirical distribution of bootstrapping in the simulation acts as a substitute for the initial simulation distribution. This introduces two sources of random variation: choosing simulation samples at random from the population, and choosing bootstrap resamples at random from the simulation samples.

Comparing Figure 4.4 to Figure 4.3, the empirical bootstrapping resampling distributions are very similar in shape, center and spread to the original simulation sampling distributions. The shape and spread of a bootstrap distribution does not rely too much on the original sample if the sample is moderately large from the population, and the bootstrap distribution is a good mimic on the shape and spread of the sampling distribution. Therefore, a bootstrap distribution can inform us about the shape, bias, and spread of the sampling distribution (Hesterberg et al., 2005).

There are still some discrepancies between the simulation distribution and the empirical distribution of bootstrapping in each sample of simulation. Random variation of each simulation is accounted for by bootstrapping. Therefore, the statistical testing using bootstrapping in the simulation is more conservative, and tends to have smaller type I error. For example, in the scenario of t0_t1, more points were below the normality line, thus the conservative type I error became smaller in the bootstrap resampling approach

(Figure 4.4 (B)). In the scenario of t0_t2 and t0_t3, some points moved from above the line to below the line, therefore, the inflated type I error rates correspondingly changed into conservative ones (Figure 4.4 (C)- (D)). In the scenario of t0_t10 and t0_t30, points were closer to the line which leaded to less inflated type I errors (Figure 4.4 (E)-(F)).

It is worthy to note that the bootstrap resampling distribution in the scenario t0_t0 was still far from normal, with a large number of values centralized in a small range (Figure 4.4(A)).

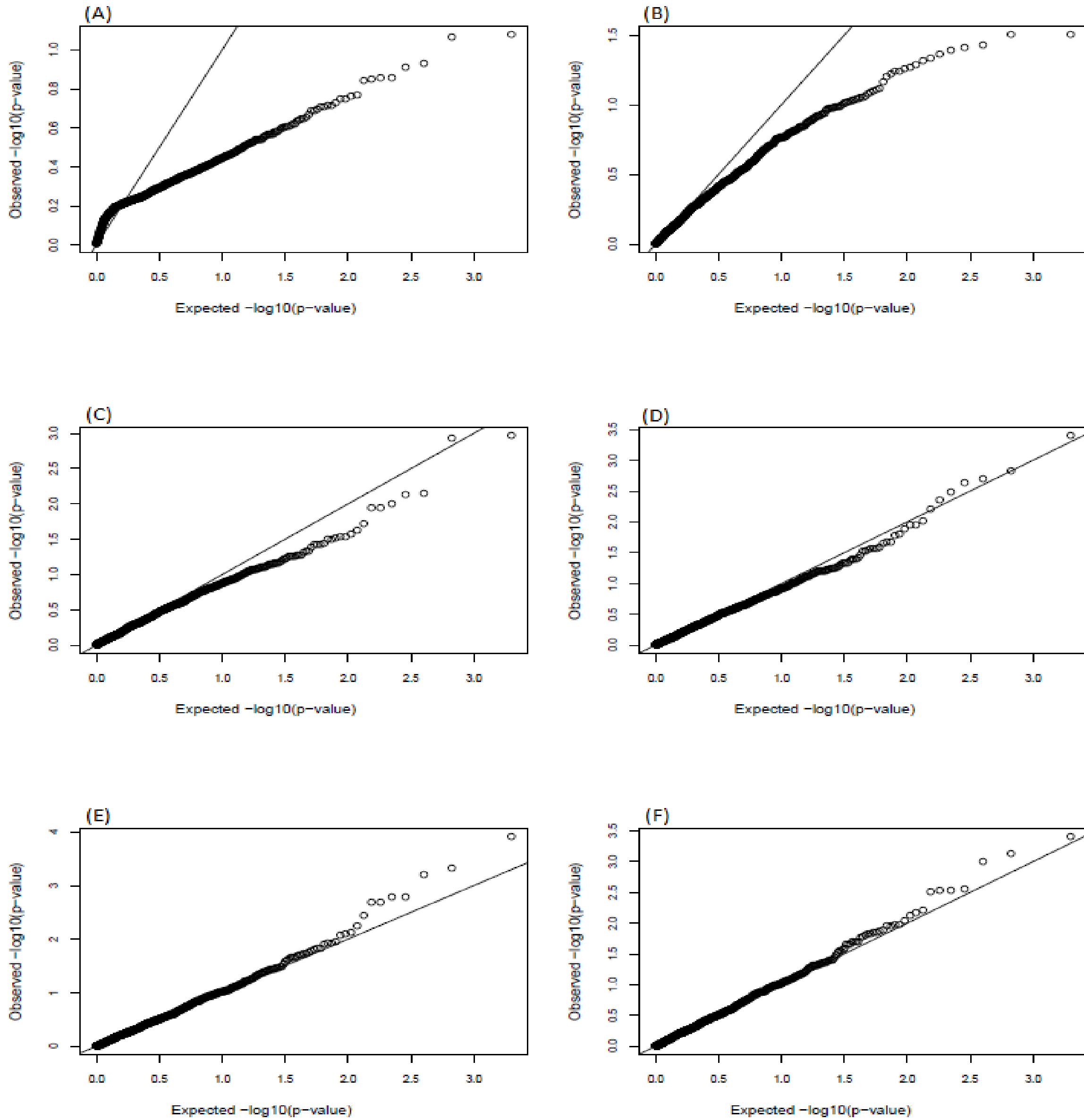

Figure 4. 4 Q-Q plot comparing the distribution of test statistic in CovA using 1000 bootstrap re-samplings method in 1000 simulation samples to a standard normal distribution

Plots are shown for six null condition scenarios: (A) 0%_0%, (B) 0%_1%, (C) 0%_2%, (D) 0%_3%, (E) 0%_10% and (F) 0%_30% where MAF=0.5, polygenic $\boldsymbol{\rho_{g^*} = 0.6}$, $\boldsymbol{\rho_e = 0.6}$, $\boldsymbol{h^2_{r1^*} = h^2_{r2^*} = 0.4}$. Line indicates the expected –log10(p-values) of a standard normal distribution. Points indicate the actual –log10(p-values) of an empirical distribution of $\boldsymbol{Z_{B_stat}}$ where the estimated variance used in $\boldsymbol{Z_{B_stat}}$ is computed from bootstrapping .

With bootstrapping there were 1,000 estimates of C_{Pi} in ith simulation, $(\hat{C}^*_{P(i,1)}, \cdots, \hat{C}^*_{P(i,1000)})$, plus the value calculated from the original sample, $\hat{C}_{Pi}$, where $i = 1, \cdots 1000$. These 1,000 bootstrapping estimates characterize properties of the true unknown estimator $\hat{C}_{Pi}$, allowing us to evaluate bias, calculate standard error, construct confidence interval directly from the empirical distribution, and compute type I error or power based on the confidence intervals of 1000 simulations.

In order to further evaluate the performance of bootstrapping in the simulation, we compared the estimate and variance of pleiotropic effect using the empirical bootstrapping distribution to the original simulation samples (Table 4.2). The variance of pleiotropic effect is computed in two ways: the estimated variance from bootstrapping and the calculated variance from our derived formula (4.7).

The estimated pleiotropic effect using bootstrapping in the simulation sample ($\hat{C}^*_P$) approximates to that using simulation sampling directly ($\hat{C}_P$). In bootstrapping approach, the estimated variance of pleiotropic effect ($var_B(\hat{C}^*_P)$) is similar to the calculated variance by using our derived formula ($var_C(\hat{C}^*_P)$). The calculated variances from the simulation samples ($var_C(\hat{C}_P)$) are a little smaller than those from bootstrap resampling in each simulation sample. The bootstrap resampling procedure introduces only a little additional variation. This fact is consistent with the notion that almost all of the variation among bootstrap distributions for a statistic comes from the selection of the original

random sample from the population (Hesterberg et al. 2005). Bootstrapping accounts for the random errors of each simulation.

Table 4. 2 Simulation distribution and empirical bootstrap distribution comparison for estimate and variance of pleiotropic effect in CovA

MAF=0.5, polygenic $\rho_{g^*} = 0.6$ and $cov_{g^*} = 0.24$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 1000

	True	Bootstrapping in the Simulation				Simulation		
scenario	C_P	$\hat{C}_P^*$	$var_B(\hat{C}_P^*)$	$var_C(\hat{C}_P^*)$	$var(\hat{C}_P^*)$	$\hat{C}_P$	$var_C(\hat{C}_P)$	$var(\hat{C}_P)$
t0_t0	0.00	0.001	0.000004	0.000006	0.000001	0.001	0.000003	0.000001
t0_t1	0.00	0.001	0.000014	0.000016	0.000011	0.001	0.000013	0.000011
t0_t2	0.00	0.001	0.000024	0.000026	0.000022	0.001	0.000023	0.000022
t0_t3	0.00	0.001	0.000034	0.000036	0.000031	0.001	0.000033	0.000032
t0_t10	0.00	0.001	0.000104	0.000105	0.000105	0.001	0.000103	0.000100
t0_t30	0.00	0.001	0.000308	0.000309	0.000310	0.001	0.000306	0.000297
t1_t1	0.01	0.011	0.000037	0.000038	0.000034	0.011	0.000035	0.000034
t1_t2	0.01	0.015	0.000051	0.000053	0.000051	0.015	0.000050	0.000051
t1_t3	0.02	0.019	0.000064	0.000067	0.000066	0.018	0.000064	0.000066
t1_t10	0.03	0.033	0.000149	0.000151	0.000153	0.032	0.000148	0.000152
t1_t30	0.05	0.056	0.000367	0.000369	0.000378	0.055	0.000366	0.000380
t2_t2	0.02	0.021	0.000067	0.000069	0.000069	0.021	0.000066	0.000069
t2_t3	0.02	0.026	0.000082	0.000084	0.000082	0.025	0.000081	0.000082
t2_t10	0.04	0.046	0.000171	0.000173	0.000175	0.045	0.000170	0.000175
t2_t30	0.08	0.078	0.000394	0.000394	0.000408	0.078	0.000391	0.000408
t3_t3	0.03	0.031	0.000099	0.000100	0.000097	0.031	0.000097	0.000097
t3_t10	0.05	0.056	0.000191	0.000192	0.000187	0.055	0.000189	0.000187
t3_t30	0.09	0.096	0.000419	0.000415	0.000416	0.096	0.000412	0.000417

Note:

a) $\hat{C}_P^*$= average of (1000 simulations of the average of (1000 bootstraps for pleiotropic effect))

b) $var_B(\hat{C}_P^*)$= average of (1000 simulations of the variance of (1000 bootstraps for pleiotropic effect))

c) $var_C(\hat{C}_P^*)$= average of (1000 simulations of the average of (1000 bootstraps for calculated variance of pleiotropic effect))

d) $var(\hat{C}_P^*)$= variance of (1000 simulations of the average of (1000 bootstraps for pleiotropic effect))

e) $\hat{C}_P$= average of (1000 simulations of pleiotropic effect)

f) $var_C(\hat{C}_P)$= average of (1000 simulations of calculated variance of pleiotropic effect)

g) $var(\hat{C}_P)$= variance of (1000 simulations of pleiotropic effect)

Type I Error and Power

To assess the empirical type I error rate and power of the evidence of pleiotropy, we adopt %BOOTCI macro from the "JACKBOOT" program. "JACKBOOT" is a collection of SAS macros developed by SAS INSTITUTE INC for data resampling analysis. The %JACK macro does jackknife analyses for simple random samples, and the %BOOT macro does elementary nonparametric bootstrap analyses for simple random samples, both methods compute approximate standard errors, bias-corrected estimates, and confidence intervals assuming a normal sampling distribution. Besides, the %BOOTCI macro computes several varieties of confidence intervals that are suitable for sampling distribution that are not normal. (http://support.sas.com/kb/24/982.html)

In this study, the %BOOTCI macro was used because our empirical distribution was non-normal. Additionally, considering the bias and skewness of the statistic, we selected three major methods of bootstrapping confidence intervals. For a statistic that is unbiased and has a symmetric sampling distribution, the percentile (PCTL) method is a

good choice which simply uses the $\alpha/2$ and 1- $\alpha/2$ percentiles of the bootstrap distribution to define the interval. The interval between the 2.5% and 97.5% percentiles of the bootstrap distribution is a 95% bootstrap confidence interval. If the statistic is biased, PCTL will amplify the bias (Efron, 1982). In this case, the Bias Corrected (BC) method is more appropriate, as it adjusts the PCTL interval for bias where the bias is not mean bias, but median bias. Further, the Bias Corrected accelerated (BCa) method adjusts the PCTL interval for both bias and skewness. However, BCa is time consuming because of extra computation on an estimate of the acceleration, which needs to be estimated by jackknifing resampling analysis. For comparison, we also computed the normal confidence intervals with and without bias correction for both bootstrapping and jackknifing approaches, which rely on the use of normal distributions for data.

We evaluated the effectiveness of different bootstrapping confidence interval methods and assessed the consequence of assuming normality in the simulated data. If the confidence interval excludes zero, the result would be concluded as significantly rejecting the null hypothesis. In the simulations using the bootstrap resampling method, type I error and power were calculated as the proportion of times the 95% confidence intervals exclude zero. In the original simulation samples, type I error and power were computed as the percentage of p-values less than 0.05 assuming a theoretical normal distribution for Z-test.

Table 4.3 illustrates the estimated type I error rate and power for different approaches. The normal approximation method using simulation samples almost never

rejects the null hypothesis when a SNP has no effect on either trait. In addition, power is poor when a SNP has small effects on both traits.

Bootstrapping PCTL method improves the power and type I error. In the scenario of no SNP effect on either trait, the estimated type I error rate of bootstrapping PCTL was 0.013, compared to 0.000 of normal approximation method at nominal $\alpha = 0.05$. In comparison with other methods assuming normal approximate distribution, the PCTL method also results in greater statistical power. Even at the scenario of both traits having only 1% of SNP effects, it achieved 81.4% of power at $\alpha = 0.05$, much better than 25.0%-32.4% of power from simulation samples and other normal-based bootstrapping approaches.

Bootstrapping BC and BCa methods have similar powers to the PCTL method, but somewhat inflated type I error rates.

Of all of these approaches, only PCTL well controls type I error and maintains good power consistently across different scenarios. The deviations from nominal α are in our acceptable range except for low effects. Therefore, the bootstrapping PCTL method is suggested for the statistical testing of SNP-specific pleiotropic effect in CovA.

Table 4. 3 Empirical type I error and power of single SNP pleiotropic effect comparing normal approximation approach and bootstrapping confidence interval approaches

MAF=0.5, polygenic ρ_{g^*}=0.6 and $cov_{g^*} = 0.24$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 1000

	Statistic Value		Type I Error or Power at α=0.05							
			Simulation	Bootstrap(Normal)		Jackknife		Bootstrap		
scenario	C_P	$\widehat{C}_P$	**Normal**	**NBC**	**NBUC**	**NBC**	**NBUC**	**PCTL**	**BC**	**BCa**
t0_t0	0.000	0.001	**0.000**	**0.001**	**0.000**	**0.082**	**0.001**	**0.013**	**0.016**	**0.016**
t0_t1	0.000	0.001	**0.037**	**0.035**	**0.014**	**0.074**	**0.036**	**0.033**	**0.079**	**0.079**
t0_t2	0.000	0.001	0.053	0.052	**0.039**	**0.064**	0.051	**0.043**	**0.074**	**0.075**
t0_t3	0.000	0.001	0.050	0.054	**0.042**	**0.062**	0.053	0.047	**0.067**	**0.068**
t0_t10	0.000	0.001	**0.043**	**0.043**	**0.041**	0.049	**0.040**	0.047	0.049	0.049
t0_t30	0.000	0.001	**0.044**	**0.044**	**0.043**	**0.043**	**0.044**	0.047	0.048	0.048
t1_t1	0.010	0.011	0.315	0.250	0.312	0.277	0.324	0.814	0.839	0.839
t1_t2	0.014	0.015	0.563	0.495	0.561	0.504	0.584	0.883	0.894	0.894
t1_t3	0.017	0.018	0.676	0.618	0.673	0.622	0.682	0.884	0.890	0.892
t1_t10	0.032	0.032	0.811	0.803	0.817	0.782	0.818	0.884	0.888	0.888
t1_t30	0.055	0.055	0.852	0.845	0.857	0.842	0.857	0.884	0.886	0.886
t2_t2	0.020	0.021	0.851	0.795	0.839	0.786	0.850	0.986	0.990	0.990
t2_t3	0.024	0.025	0.933	0.911	0.932	0.898	0.934	0.993	0.995	0.995
t2_t10	0.045	0.045	0.983	0.978	0.985	0.977	0.984	0.993	0.997	0.997
t2_t30	0.077	0.078	0.989	0.988	0.991	0.990	0.989	0.993	0.997	0.997
t3_t3	0.030	0.031	0.985	0.978	0.984	0.975	0.985	1.000	1.000	1.000
t3_t10	0.055	0.055	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
t3_t30	0.095	0.096	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

*Shaded areas signify null hypothesis conditions. Unshaded areas reflect non-null hypothesis conditions. Unbold number reflects the error rate fells within Bradley's criterion. Bold numbers signify either inflated or conservative type I errors. NBC: normal approximation method, bias corrected; NBUC: normal approximation method, bias uncorrected; PCTL: percentile method; BC: bias corrected method; BCa: bias corrected accelerated method.

Bootstrapping Size

In order to explore the impact of changing resample numbers on the performance of bootstrapping PCTL method in CovA, we increased the number of bootstrap samples from 1,000 to 10,000 in seven scenarios (Table 4.4). Empirical type I error rate and power are similar at a nominal $\alpha = 0.05$ using different numbers of bootstrap resamples. One thousand resamples instead of ten thousand are sufficient for the bootstrap resampling procedure, introducing very little additional variation from the original sample and excellently mimicking the shape, bias and spread of the sampling distribution.

Table 4. 4 Empirical type I error and power results comparing different numbers of bootstrapping samples in CovA

MAF=0.5, polygenic $\rho_{g^*} = 0.6$ and $cov_{g^*} = 0.24$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$

Scenario	N=1000	N=10000
t0_t0	0.013	0.010
t0_t1	0.033	0.032
t0_t2	0.043	0.041
t0_t3	0.047	0.046
t0_t10	0.047	0.052
t0_t30	0.047	0.052
t1_t1	0.814	0.820

Coverage Probability

The performance of bootstrapping PCTL method in CovA is also evaluated in terms of empirical coverage probability of the confidence intervals, referring to the proportion of times that the interval contains the true value of interest (Dodge 2006;

Kysely, 2010). The coverage probability of the 95% confidence interval is lower (93.6%) than the nominal value of 95% for the Jackknife method with bias-correction under the t0_t0 scenario, indicating that the confidence interval constructed using that method is narrow and undervalued the uncertainty involved in the estimates. The 95% confidence intervals of all other methods and the Jackknife method with bias-correction under other scenarios cover 100% of cases, indicating that the confidence intervals using these methods are wider compared to the real uncertainty.

Comparison of Different Polygenic Correlation and Environmental Correlation Values

In order to provide a more comprehensive investigation of the influence of polygenic and environmental factor effects on CovA, we compared twelve sets of polygenic correlations (ρ_g) of 0.0 (none), 0.1 (low), 0.6 (moderate) and 0.9 (high), and environmental correlations (ρ_e) of 0.1 (low), 0.6 (moderate) and 0.9 (high) at 50% MAF. Appendix Table S4 provides the true value of pleiotropic effect, the mean of the estimated pleiotropic effect, percent bias (%), and type I error or power in columns for each scenarios.

The results showed that CovA is consistent across different polygenic genetic correlations, environmental correlations and MAFs with unbiased estimate, well controlled type I errors except for some cases when a SNP has no effect on either trait or small effect on one trait, and good power.

Changes in polygenic correlation do not lend a significant influence in the accuracy of pleiotropic effect estimation. With a fixed environmental correlation, the estimate biases are close to each other across all polygenic correlation levels. When a SNP has no effect on either trait, increased polygenic correlation contributes to a little improved type I error except ρ_g =0.0. When a SNP has effect on only one trait, all of the conditions show a slight decrease in type I error with increasing polygenic correlation. When a SNP has effect on both traits, larger polygenic correlation leads some improvement on power.

The accuracy of an estimator is generally affected by environmental correlation. With a fixed polygenic correlation, a larger environmental correlation results in an overestimated SNP-specific pleiotropic effect. When a SNP has no effect on either trait, increasing the environmental correlation from 0.0, 0.6 to 0.9 provides a substantive improvement to the conservative type I error rate. When a SNP has effect on only one trait, increasing the environmental correlation decreases the type I error rate. When a SNP has effect on both traits, there is a corresponding increase in power, as environmental correlation increased.

Comparison of Different Minor Allele Frequencies

We performed some additional simulations to look at the estimate, type I error and power under different minor allele frequencies (MAF=0.1 and 0.5) at polygenic correlation of 0.6 and environmental correlation of 0.6 (Appendix Table S5).

CovA does not vary much under different MAF. The estimated biases from both MAF are positive and less than 10%, indicating the accuracy is acceptable. Smaller MAF yields slightly larger bias, type I error and power, but the extent of influence is small.

4.3 Multiple SNPs Analysis

4.3.1 Proposed Method

To extend the model for multiple SNPs, we consider $\boldsymbol{SNP} = (SNP_1, SNP_2, \cdots, SNP_m)$ where SNP_j is 0, 1 and 2 if the genotype of the individual at the *j*th SNP is AA, Aa and aa, respectively.

For illustration, we focus our discussion on bivariate linear models with only two SNPs and assume that there is no linkage disequilibrium (LD) between the SNPs.

$$Y_1 = \mu_1 + \beta_{11}SNP1 + \beta_{12}SNP2 + \beta_{13}V + e_1, \tag{4.12}$$

$$Y_2 = \mu_2 + \beta_{21}SNP1 + \beta_{22}SNP2 + \beta_{23}V + e_2.$$

Similar to the single SNP pleiotropy, the effect estimate of the joint pleiotropic effects from two SNPs can be written as

$$C_P = \hat{\beta}_{11}var(SNP1)\hat{\beta}_{21} + \hat{\beta}_{11}cov(SNP1, SNP2)\hat{\beta}_{22} + \hat{\beta}_{12}cov(SNP1, SNP2)\hat{\beta}_{21} + \hat{\beta}_{12}var(SNP2)\hat{\beta}_{22}, \tag{4.13}$$

where $\hat{\beta}_{11}$, $\hat{\beta}_{12}$, $\hat{\beta}_{21}$ and $\hat{\beta}_{22}$ represent the effect size of *SNP*1 and *SNP*2 on trait1 and trait2, respectively; $var(SNP1)$ and $var(SNP2)$ are their variances. $cov(SNP1, SNP2)$ is the covariance.

The derived variance of the joint pleiotropic effects is

$$
\begin{aligned}
var(C_P) = var\big[&\hat{\beta}_{11} var(SNP1)\hat{\beta}_{21} + \hat{\beta}_{11} cov(SNP1, SNP2)\hat{\beta}_{22} \\
&+ \hat{\beta}_{12} cov(SNP1, SNP2)\hat{\beta}_{21} + \hat{\beta}_{12} var(SNP2)\hat{\beta}_{22}\big] \\
&= var^2(SNP1) var\big(\hat{\beta}_{11}\hat{\beta}_{21}\big) + var^2(SNP2) var\big(\hat{\beta}_{12}\hat{\beta}_{22}\big) \\
&+ 2var(SNP1)var(SNP2)cov\big(\hat{\beta}_{11}\hat{\beta}_{21}, \hat{\beta}_{12}\hat{\beta}_{22}\big) \\
&+ cov^2(SNP1, SNPL2)\big[var\big(\hat{\beta}_{11}\hat{\beta}_{22}\big) + var\big(\hat{\beta}_{12}\hat{\beta}_{21}\big) \\
&+ 2cov\big(\hat{\beta}_{11}\hat{\beta}_{22}, \hat{\beta}_{12}\hat{\beta}_{21}\big)\big] \\
&+ 2var(SNP1)cov(SNP1, SNP2)\big[cov\big(\hat{\beta}_{11}\hat{\beta}_{21}, \hat{\beta}_{11}\hat{\beta}_{22}\big) \\
&+ cov\big(\hat{\beta}_{11}\hat{\beta}_{21}, \hat{\beta}_{12}\hat{\beta}_{21}\big)\big] \\
&+ 2var(SNP2)cov(SNP1, SNP2)\big[cov\big(\hat{\beta}_{11}\hat{\beta}_{22}, \hat{\beta}_{12}\hat{\beta}_{22}\big) \\
&+ cov\big(\hat{\beta}_{12}\hat{\beta}_{21}, \hat{\beta}_{12}\hat{\beta}_{22}\big)\big]
\end{aligned}
\tag{4.14}
$$

Since these two SNPs are independent, $r^2 = \frac{cov^2(SNP1,SNPL2)}{var(SNP1)var(SNP2)} = 0$, indicating $cov(SNP1, SNP2) = 0$. Then the variance reduces to

$$
\begin{aligned}
var(C_P) = var\big[&\hat{\beta}_{11} var(SNP1)\hat{\beta}_{21} + \hat{\beta}_{11} cov(SNP1, SNP2)\hat{\beta}_{22} \\
&+ \hat{\beta}_{12} cov(SNP1, SNP2)\hat{\beta}_{21} + \hat{\beta}_{12} var(SNP2)\hat{\beta}_{22}\big] \\
&= var^2(SNP1) var\big(\hat{\beta}_{11}\hat{\beta}_{21}\big) + var^2(SNP2) var\big(\hat{\beta}_{12}\hat{\beta}_{22}\big) \\
&+ 2var(SNP1)var(SNP2)cov\big(\hat{\beta}_{11}\hat{\beta}_{21}, \hat{\beta}_{12}\hat{\beta}_{22}\big)
\end{aligned}
\tag{4.15}
$$

Based on the variance of the product of two dependent variables (Goodman, 1960), the variance and covariance of regression coefficients are approximated

$$var(\hat{\beta}_{11}\hat{\beta}_{21}) = [E(\hat{\beta}_{11})]^2 var(\hat{\beta}_{21}) + [E(\hat{\beta}_{21})]^2 var(\hat{\beta}_{11}) + 2E(\hat{\beta}_{11})E(\hat{\beta}_{21})cov(\hat{\beta}_{11}, \hat{\beta}_{21}) \quad (4.16)$$

$$var(\hat{\beta}_{12}\hat{\beta}_{22}) = [E(\hat{\beta}_{12})]^2 var(\hat{\beta}_{22}) + [E(\hat{\beta}_{22})]^2 var(\hat{\beta}_{12}) + 2E(\hat{\beta}_{12})E(\hat{\beta}_{22})cov(\hat{\beta}_{12}, \hat{\beta}_{22}) \quad (4.17)$$

$$cov(\hat{\beta}_{11}\hat{\beta}_{21}, \hat{\beta}_{12}\hat{\beta}_{22}) = [E(\hat{\beta}_{11})][E(\hat{\beta}_{12})]cov(\hat{\beta}_{21}, \hat{\beta}_{22}) + [E(\hat{\beta}_{21})][E(\hat{\beta}_{22})]cov(\hat{\beta}_{11}, \hat{\beta}_{12}) + E(\hat{\beta}_{11})E(\hat{\beta}_{22})cov(\hat{\beta}_{21}, \hat{\beta}_{12}) + E(\hat{\beta}_{21})E(\hat{\beta}_{12})cov(\hat{\beta}_{11}, \hat{\beta}_{22}) \quad (4.18)$$

The hypotheses can be expressed as

$$H_0: C_P = 0 \, ;$$

$$H_a: C_P \neq 0 \, .$$

The statistical testing of the two SNPs pleiotropic effects is identical to the single SNP situation.

4.3.2 Evaluation of the Proposed Method Using Simulation Studies

We performed simulations to further explore some properties of the joint pleiotropic effects. Two SNPs in linkage equilibrium (LE) that had two alleles of equal frequency were generated for 1,000 uncorrelated trios in the first step. Then two correlated quantitative traits were simulated based on various SNP effects, along with a polygenic background. These two traits were normally distributed with residual

polygenic genetic correlation of 0.6, environmental correlation of 0.6 and residual heritability of 0.4. Four SNP heritabilities were considered (0%, 1%, 2%, 3% giving the size of effect for each SNP as 0, 0.1414, 0.2000 and 0.2450 units, and the standard deviation for each SNP as 1, 0.995 and 0.990, respectively). This study design incorporated twenty-two scenarios for different combination of SNP heritabilities (Table4.5). Simulation was performed on 1000 replicates.

Table 4. 5 Pairs of two SNPs effects on bivariate traits in simulations of CovA

	Trait1[a]	Trait2[b]
No pleiotropy (No effects on T1 and T2)	t0_t0	t0_t0
No pleiotropy(No effect on T1)	t0_t0	t1_t0, t1_t1 t2_t0, t2_t1, t2_t2 t3_t0, t3_t1, t3_t2, t3_t3
Pleiotropic effect on both T1 and T2	t1_t1	t1_t0, t1_t1, t1_t2, t1_t3
	t2_t2	t2_t0, t2_t1, t2_t2, t2_t3
	t3_t3	t3_t0, t3_t1, t3_t2, t3_t3

[a](h^2_{q11}%, h^2_{q21}%) denotes SNP1 with h^2_{q11}% effect on trait1 and SNP2 with h^2_{q21}% effect on trait1.

[b](h^2_{q12}%, h^2_{q22}%) denotes SNP1 with h^2_{q12}% effect on trait2 and SNP2 with h^2_{q22}% effect on trait2.

Data were generated using the ‘simqtl’ command in SOLAR. Then we selected 1,000 unrelated subjects from each sample and used multivariate regression modeling (4.19) to obtain the regression coefficients.

$$Y_1 = \mu_1 + \beta_{11} SNP1 + \beta_{12} SNP2 + e_1, \quad (4.19)$$

$$Y_2 = \mu_2 + \beta_{21} SNP1 + \beta_{22} SNP2 + e_2.$$

The procedures for evaluating pleiotropy for two SNPs were similar to single SNP analysis.

4.3.3 Simulation Results

Bias of the Estimator

Figure 4.6 displays the estimated two SNPs joint pleiotropic effect across SNPs heritabilities using CovA. In order to explicitly reflect the relationship between the single SNP pleiotropic effect and the joint SNPs pleiotropic effects, we also presented the pleiotropic effects from SNP1 and SNP2 individually and jointly.

The estimated joint SNP pleiotropic effects are close to their true values. If neither SNP has individual pleiotropic effect, the joint pleiotropic effects are also zero. If only one SNP has individual pleiotropic effect, the joint pleiotropic effects are identical to that individual pleiotropic effect. If both SNPs have substantial pleiotropic effects, the joint pleiotropic effects are approximately the sum of the individual pleiotropic effects.

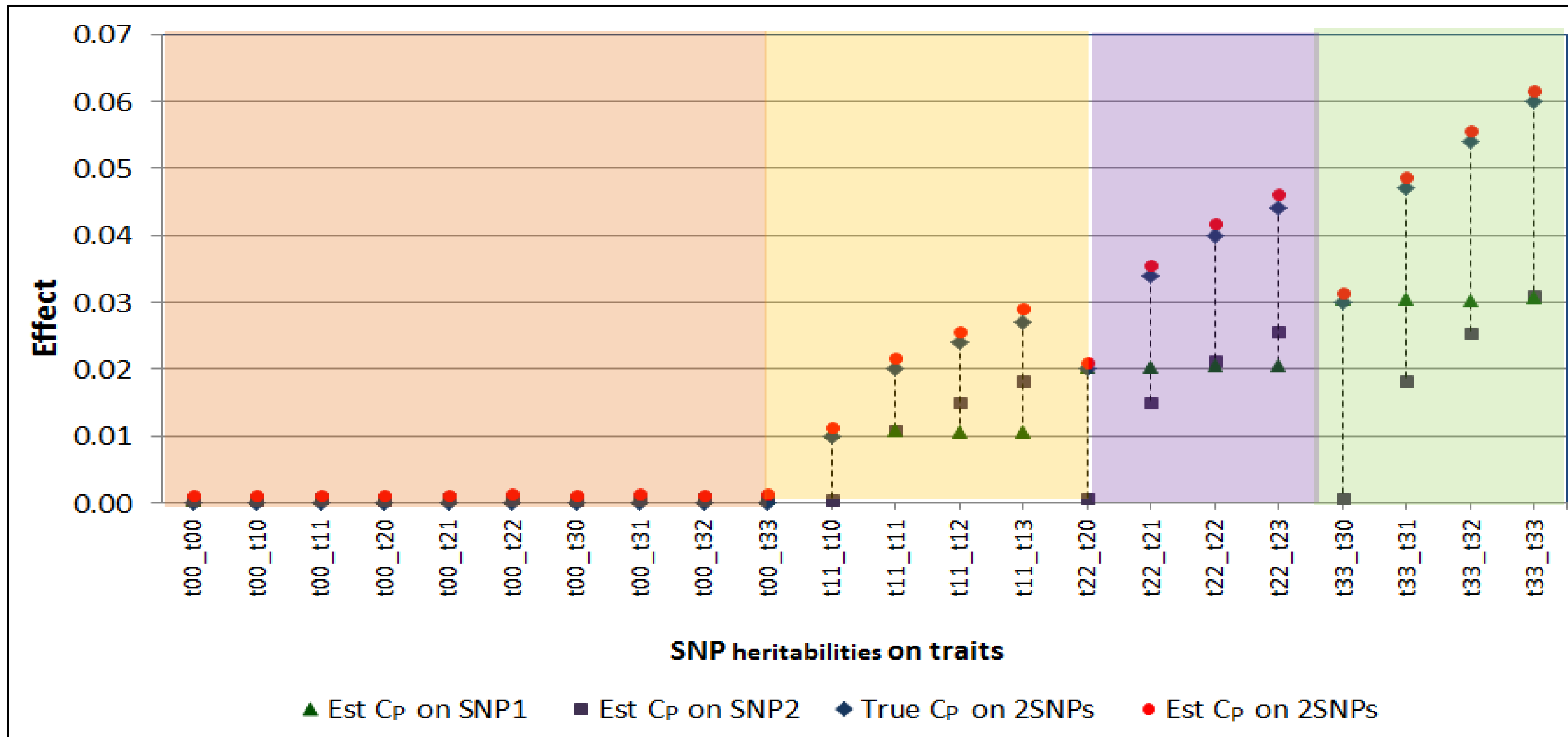

Figure 4. 5 Two SNPs joint pleiotropic effects and individual pleiotropic effects using CovA.

MAF=0.5, polygenic $\rho_{g^*} = 0.6$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$.

*t$h^2_{q11}h^2_{q12}$ _t$h^2_{q21}h^2_{q22}$ denoted SNP1 with h^2_{q11}% effect on trait1 and h^2_{q21}% effect on trait2, and SNP2 with h^2_{q12}% effect on trait1 and h^2_{q22}% effect on trait2.

Type I Error and Power

Because of the non-normal distribution of test statistic we used the bootstrap resampling method (B=1000) in statistical testing for two SNPs pleiotropic effects analysis. The empirical type I error and power of seven bootstrapping confidence interval methods were compared with the normal approximation method using simulation samples (Table 4.6). The results showed that any method based on normal approximation almost never rejects the null hypothesis when all SNPs effects are zero, and generates low power when all SNPs effects are small. This situation is greatly improved by using bootstrapping PCTL method. PCTL produces some conservative rates and some moderately liberal rates in type I error. In addition, PCTL provides the best performance in power.

Table 4. 6 Empirical type I error and power of two SNPs pleiotropic effects comparing normal approximation approach and bootstrapping confidence interval approaches

MAF=0.5, polygenic ρ_g=0.6 and $cov_g = 0.24$, $\rho_e = 0.6$, $h^2_{r1} = h^2_{r2} = 0.4$
Simulation = 1000, bootstrapping in each simulation = 1000

	Statistic Value		Type I Error or Power at α =0.05							
	True		Simulation	Bootstrap(Normal)		Jackknife		Bootstrap		
scenario	Value	$\widehat{C}_P$	Normal	NBC	NBUC	NBC	NBUC	PCTL	BC	BCa
t00_t00	0.000	0.001	**0.001**	**0.000**	**0.001**	**0.056**	**0.002**	**0.026**	**0.016**	**0.016**
t00_t10	0.000	0.001	**0.013**	**0.018**	**0.009**	**0.065**	**0.017**	**0.040**	0.055	0.055
t00_t11	0.000	0.001	**0.034**	**0.040**	**0.023**	**0.062**	**0.036**	**0.041**	**0.063**	**0.063**
t00_t20	0.000	0.001	**0.040**	**0.038**	**0.019**	**0.065**	**0.037**	**0.036**	**0.056**	**0.056**
t00_t21	0.000	0.001	**0.035**	0.046	**0.035**	**0.056**	**0.040**	**0.043**	**0.063**	**0.063**
t00_t22	0.000	0.001	**0.042**	**0.041**	**0.034**	**0.056**	**0.038**	0.046	0.053	0.055
t00_t30	0.000	0.001	**0.039**	**0.040**	**0.031**	**0.057**	0.046	**0.040**	**0.063**	**0.063**
t00_t31	0.000	0.001	**0.044**	**0.044**	**0.036**	0.055	**0.043**	**0.043**	**0.056**	**0.056**
t00_t32	0.000	0.001	**0.044**	0.047	**0.039**	**0.059**	**0.043**	0.047	0.053	0.053
t00_t33	0.000	0.001	**0.043**	0.046	**0.039**	**0.056**	**0.043**	**0.041**	0.052	0.053

*Shaded areas signify null hypothesis conditions. Unshaded areas reflect non-null hypothesis conditions. Unbold number reflects the error rate fells within Bradley's criterion. Bold numbers signify either inflated or conservative type I errors. NBC: normal approximation method, bias corrected; NBUC: normal approximation method, bias uncorrected; PCTL: percentile method; BC: bias corrected method; BCa: bias corrected accelerated method. t$h^2_{q11}h^2_{q12}$ _t$h^2_{q21}h^2_{q22}$ denoted SNP1 with h^2_{q11}% effect on trait1 and h^2_{q21}% effect on trait2, and SNP2 with h^2_{q12}% effect on trait1 and h^2_{q22}% effect on trait2.

Table 4.5 (Continued) Empirical type I error and power of two SNPs pleiotropic effects comparing normal approximation approach and bootstrapping confidence interval approaches

MAF=0.5, polygenic ρ_g=0.6 and $cov_g = 0.24$, $\rho_e = 0.6$, $h^2_{r1} = h^2_{r2} = 0.4$
Simulation = 1000, bootstrapping in each simulation = 1000

	Statistic Value		Type I Error or Power at α =0.05							
	True		Simulation	Bootstrap(Normal)		Jackknife		Bootstrap		
scenario	Value	$\widehat{C}_P$	Normal	NBC	NBUC	NBC	NBUC	PCTL	BC	BCa
t11_t10	0.010	0.011	0.280	0.173	0.254	0.188	0.276	0.566	0.515	0.517
t11_t11	0.020	0.022	0.642	0.761	0.852	0.785	0.877	0.996	0.992	0.993
t11_t12	0.024	0.026	0.872	0.870	0.921	0.881	0.937	0.995	0.993	0.993
t11_t13	0.027	0.029	0.872	0.908	0.945	0.917	0.951	0.994	0.992	0.993
t22_t20	0.020	0.021	0.962	0.559	0.626	0.566	0.645	0.830	0.795	0.795
t22_t21	0.034	0.036	0.960	0.979	0.990	0.977	0.992	0.999	0.999	0.999
t22_t22	0.040	0.042	0.992	0.998	0.999	0.997	1.000	1.000	1.000	1.000
t22_t23	0.044	0.046	1.000	0.999	1.000	0.999	1.000	1.000	1.000	1.000
t33_t30	0.030	0.031	1.000	0.832	0.866	0.831	0.867	0.941	0.927	0.928
t33_t31	0.047	0.049	0.999	0.998	1.000	0.999	1.000	1.000	1.000	1.000
t33_t32	0.054	0.056	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
t33_t33	0.060	0.062	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

*Shaded areas signify null hypothesis conditions. Unshaded areas reflect non-null hypothesis conditions. Unbold number reflects the error rate fells within Bradley's criterion. Bold numbers signify either inflated or conservative type I errors. NBC: normal approximation method, bias corrected; NBUC: normal approximation method, bias uncorrected; PCTL: percentile method; BC: bias corrected method; BCa: bias corrected accelerated method. t$h^2_{q11}h^2_{q12}$ _t$h^2_{q21}h^2_{q22}$ denoted SNP1 with h^2_{q11}% effect on trait1 and h^2_{q21}% effect on trait2, and SNP2 with h^2_{q12}% effect on trait1 and h^2_{q22}% effect on trait2.

It is important to mention that in the two SNPs pleiotropic effects, including a null effect SNP in models erode power for all eight methods of bootstrapping confidence intervals. Table 4.7 displays the simulation results of bootstrapping PCTL method for one SNP-specific pleiotropic effect and two SNPs-specific pleiotropic effect analysis. The power of a joint pleiotropic effect with a nuisance SNP in model is smaller than power of an individual pleiotropic effect. Therefore, researchers determined to evaluate multiple SNPs pleiotropic effect are advised to examine SNP effects first and avoid nuisance SNPs in models.

Table 4. 7 Power comparing single SNP pleiotropic effect and two SNPs pleiotropic effects using bootstrapping PCTL method.

MAF=0.5, polygenic ρ_g=0.6 and $cov_g = 0.24$, $\rho_e = 0.6$, $h^2_{r1} = h^2_{r2} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 1000

Power at α =0.05				Power at α =0.10			
2 SNPs Test		1 SNP Test		2 SNPs Test		1 SNP Test	
t11_t10	0.566	**t1_t1**	0.814	**t11_t10**	0.694	**t1_t1**	0.886
t22_t20	0.830	**t2_t2**	0.986	**t22_t20**	0.889	**t2_t2**	0.995
t33_t30	0.941	**t3_t3**	1.000	**t33_t30**	0.970	**t3_t3**	1.000

*t$h^2_{q11}h^2_{q12}$_t$h^2_{q21}h^2_{q22}$ denoted SNP1 with h^2_{q11}% effect on trait1 and h^2_{q21}% effect on trait2, and SNP2 with h^2_{q12}% effect on trait1 and h^2_{q22}% effect on trait2. h^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2.

4.4 Summary

We have proposed a method based on multivariate regression model by examining the contribution of specific genetic marker(s) to phenotypic covariance. This

approach can be applied for population-based studies with samples of families or unrelated subjects. Covariance analysis allows estimate and test of pleiotropic SNP effect. Because a normality assumption of the test statistic is not always warranted, a test based on a normal distribution is not valid in all cases. We compared several bootstrapping confidence interval methods to the normal approximation method in CovA. The normal approximation method almost never rejects the null hypothesis if a SNP has no effect on either trait. In addition, it yields poor power if a SNP has very small effects on both traits. Of all these approaches, only PCTL improves type I error rate and maintains good power consistently across different SNP heritabilities. Thus, we suggested the bootstrap PCTL method for the subsequence analyses. The operation is simple and it is easy to use.

Based on the simulations, CovA is very consistent across different residual polygenic genetic correlations, environmental correlations and MAFs with unbiased estimate, well controlled type I error except for some cases when a SNP has no effect on either trait or small effect on one trait, and good power. This robustness to polygenic correlation makes CovA appealing. In addition, environmental correlation and MAF do not essentially change the facts. Environmental correlation has a little influence on this approach. Increasing environmental correlation results in bigger bias, increased type I error if a SNP has no effect on either trait, decreased type I error if a SNP has effect on one trait, and increased power if a SNP has effects on both traits. We also found that CovA is slightly affected by different MAFs. Larger MAF produces larger bias, type I error and power.

Chapter 5: Compare Methods in the Context of Pleiotropic Effects

5.1 Comparison of Covariance Analysis, Multivariate Analysis Of Variance and Conditional Analysis Approaches Using Unrelated Data

CovA and MANOVA share the common bivariate regression models:

$$Y_1 = \beta_{10} + \beta_{11}SNP + e_1, \tag{4.20}$$

$$Y_2 = \beta_{20} + \beta_{21}SNP + e_2,$$

where Y_j denotes the value of the *j*th trait (*j*=1,2); SNP denotes a SNP with observed genotypes coded as 0, 1 and 2 for the number of minor alleles; β_{10}, β_{11}, β_{20} and β_{21} denote regression coefficients; e_i is a random error, $E(e_i) = 0$, $Cov(e_i, e_j) = \sigma_{ij}$, $i, j = 1,2$.

However, their hypotheses have essential distinctions:

MANOVA

$$H_0\colon \beta_{11} = \beta_{21} = 0 \;\; vs \;\; H_1\colon \beta_{11} \neq 0 \; or \; \beta_{21} \neq 0$$

CovA

$$H_0\colon \beta_{11} = 0 \; or \; \beta_{21} = 0 \;\; vs \;\; H_1\colon \beta_{11} \neq 0 \; and \; \beta_{21} \neq 0$$

In MANOVA, if a SNP has an effect on only one trait, the null hypothesis would be rejected. Undoubtedly, rejecting the null hypothesis does not agree with the hypothesis of pleiotropic SNP effect. In contrast, the null hypothesis of CovA is that a SNP has no effect on any of the traits. If a SNP has an effect on only one trait, we fail to reject the

null hypothesis and confirm non-pleiotropy. Distinguishing these two cases has important implications in pleiotropic effect analysis.

As for conditional analysis, we investigate a SNP effect on two traits by conducting a set of regression models. Let Y_1 and Y_2 denote the values of two correlated traits.

Model1: $$Y_1 = \beta_{10} + \beta_{11}SNP + \beta_{12}Y_2 + e_1, \quad (4.22)$$

Model2: $$Y_2 = \beta_{20} + \beta_{21}SNP + \beta_{22}Y_1 + e_2, \quad (4.23)$$

Model3: $$Y_1 = \beta_{30} + \beta_{31}SNP + e_3, \quad (4.24)$$

Model4: $$Y_2 = \beta_{40} + \beta_{41}SNP + e_4. \quad (4.25)$$

We alternate one trait as a covariate in the analysis of another trait and estimate the SNP effects conditional and unconditional on the covariate. Model1 (4.22) is a conditional analysis where a SNP effect on the dependent variable is smaller than the covariate. Model2 (4.23) is a conditional analysis where the SNP effect on the dependent variable is bigger than the covariate. Model3 (4.24) and model4 (4.25) are corresponding unconditional analysis.

The hypothesis testing is to examine the change of SNP effects conditioning and not conditioning one trait to the other.

Conditional Analysis

$$H_0: \beta_{31} - \beta_{11} = 0 \quad vs \quad H_1: \beta_{31} - \beta_{11} \neq 0,$$

or

$$H_0: \beta_{41} - \beta_{21} = 0 \quad vs \quad H_1: \beta_{41} - \beta_{21} \neq 0.$$

A substantive change in the effect sizes of a SNP from unconditional analysis to conditional analysis provides an evidence of a shared SNP effect between traits. Similar to CovA, the bootstrapping PCTL method is used in the conditional analyses for statistical testing. If the confidence interval of the difference between the SNP effect sizes excludes zero, the null hypothesis is rejected. Type I error and power are calculated as the proportion of times the 95% CIs exclude zero.

We compared MANOVA and conditional analysis with CovA under different residual polygenic genetic correlations and environmental correlations.

The simulation design was based on 1,000 independent subjects and 12 levels of residual polygenic and environmental correlations: residual polygenic correlation (ρ_g) of 0.0, 0.1 0.6 and 0.9; environmental correlation (ρ_e) of 0.0, 0.6 and 0.9; and residual heritability (h_{r1}^2 and h_{r2}^2) of 1% for ρ_g = 0.0 and 40% for others. SNP heritability varied from 0% to 3%. Simulated data from each sample was analyzed using each of four tests.

Table 5.1 and Appendix Table S6 display the mean of the estimated pleiotropic effect (or estimated size of the SNP effect), type I error and power of CovA, MANOVA, and conditional analysis approaches for a single SNP-specific pleiotropic effect on two traits. We also compared simulation results across the polygenic and environmental correlations from low to high.

In comparison to MANOVA and conditional analysis, only CovA can quantify pleiotropic effect. The effect estimate is unbiased and unaffected by the polygenic

correlation but slightly affected by the environmental correlation. Smaller environmental correlation contributes more accurate estimate. When a SNP has no effect on either trait, CovA demonstrates conservative type I errors across all polygenic and environmental correlation conditions. Increasing polygenic correlation or environmental correlation generally yields an improvement to type I error. When a SNP has effect on one trait, type I errors in CovA are well controlled except for some cases when SNP effect on that trait is small for all polygenic and environmental correlation conditions. Additionally, type I error rate decreases as the polygenic correlation or environmental correlation increases. When a SNP has effect on both traits, CovA presents good powers in all conditions, which are higher under larger polygenic correlation or environmental correlation level.

MANOVA does not estimate the shared SNP effect on two traits, but the SNP effects on two traits individually. When a SNP has no effect on either trait, MANOVA tended to exhibit inflated type I errors across all polygenic or environmental correlation conditions (5.5% to 6.4%). When a SNP has effect on only one trait, MANOVA does not focus on pleiotropy, whose rejection rates ranged from 81.7% to 100.0% across all conditions. When a SNP has effect on both traits, MANOVA exhibited good power (81.7% to 100.0%) which decreases with environmental correlation.

In conditional analysis method, we compare the simulation results between two conditional models: model1 using the trait with a larger SNP effect (trait2) as a covariate and model2 using the trait with a smaller SNP effect (trait1) as a covariate.

Model1 produces large type I errors when a SNP has effect on only one trait in all polygenic and environmental correlation conditions except both correlations are zeros.

Therefore, a model using the trait with a smaller SNP effect as dependent variable and the trait with a larger SNP effect as covariate is manifested to be inappropriate in conditional analysis. Researchers are advised to examine SNP effects first and determine which trait to be used as covariates appropriately (Li et al., 2006).

Model2 estimates the SNP effect on trait2 conditional on trait1. If polygenic correlation and environmental correlation are moderate to high, conditional analysis method exhibits an appearance of a great degree of effectiveness to test for pleiotropy. When a SNP has no effect on at least one trait, type I error rates are well controlled across all conditions. When a SNP has effect on both traits, conditional analysis retains good power if neither polygenic correlation nor environmental correlation is very small. Unfortunately, if both polygenic correlation and environmental correlation are zero or very small, the conditional analysis generates low power. The possible explanation can be that conditional analysis is essentially assessing any genetic or environmental effects between traits, rather than the SNP-specific pleiotropic effect.

In sum, the simulation results demonstrated that only CovA effectively and accurately quantifies and tests pleiotropic SNP effect and is consistent across all polygenic and environmental correlations.

Table 5. 1 Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.0 , \rho_e = 0.0,\ h^2_{r1^*} = h^2_{r2^*} = 0.01$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\widehat{C}_P$	$\alpha = .05$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\alpha = .05$	$\widehat{\beta}_{41} - \widehat{\beta}_{21}$	$\alpha = .05$	$\widehat{\beta}_{31} - \widehat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.000	0.002	0.002	-0.002	0.063	0.000	0.003	0.000	0.002
t0_t1	0.000	0.000	0.041	0.002	0.141	***0.817***	0.000	0.002	0.000	0.034
t0_t2	0.000	0.000	0.047	0.002	0.202	***0.989***	0.000	0.002	0.000	0.049
t0_t3	0.000	0.000	0.047	0.002	0.246	***0.997***	0.000	0.001	0.000	0.046
t1_t1	0.010	0.010	0.780	0.140	0.141	0.984	0.000	***0.047***	0.000	***0.048***
t1_t2	0.014	0.014	0.871	0.140	0.202	1.000	0.000	***0.038***	0.000	***0.051***
t1_t3	0.017	0.017	0.876	0.140	0.246	1.000	0.000	***0.027***	0.000	***0.040***
t2_t2	0.020	0.020	0.982	0.200	0.202	1.000	0.000	***0.044***	0.000	***0.044***
t2_t3	0.024	0.025	0.986	0.200	0.246	1.000	0.000	***0.048***	0.000	***0.049***
t3_t3	0.030	0.030	1.000	0.245	0.246	1.000	0.000	***0.047***	0.000	***0.047***

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table 5.1 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.0$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.01$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\hat{C}_P$	$\alpha = .05$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\alpha = .05$	$\hat{\beta}_{41} - \hat{\beta}_{21}$	$\alpha = .05$	$\hat{\beta}_{31} - \hat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.001	0.011	0.002	0.000	0.063	0.001	0.045	0.000	0.063
t0_t1	0.000	0.001	0.037	0.002	0.142	***0.948***	0.001	0.046	0.085	***0.880***
t0_t2	0.000	0.001	0.046	0.002	0.202	***0.999***	0.001	0.046	0.121	***0.993***
t0_t3	0.000	0.001	0.047	0.002	0.246	***1.000***	0.001	0.046	0.148	***0.998***
t1_t1	0.010	0.011	0.795	0.141	0.142	0.888	0.084	0.868	0.084	0.880
t1_t2	0.014	0.015	0.874	0.141	0.202	0.987	0.083	0.868	0.121	0.993
t1_t3	0.017	0.018	0.877	0.141	0.246	0.997	0.083	0.868	0.148	0.998
t2_t2	0.020	0.021	0.986	0.201	0.202	0.995	0.119	0.992	0.120	0.993
t2_t3	0.024	0.025	0.990	0.201	0.246	0.998	0.119	0.992	0.147	0.998
t3_t3	0.030	0.031	0.998	0.246	0.246	0.999	0.146	1.000	0.146	0.998

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table 5.1 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.6$, $\rho_e = 0.0$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\hat{C}_P$	$\alpha = .05$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\alpha = .05$	$\hat{\beta}_{41} - \hat{\beta}_{21}$	$\alpha = .05$	$\hat{\beta}_{31} - \hat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.000	0.002	0.001	-0.002	0.057	0.000	0.044	0.000	0.053
t0_t1	0.000	0.000	0.037	0.001	0.141	***0.848***	0.000	0.044	0.034	***0.897***
t0_t2	0.000	0.000	0.045	0.001	0.202	***0.992***	0.000	0.044	0.049	***0.995***
t0_t3	0.000	0.000	0.047	0.001	0.246	***0.999***	0.000	0.044	0.060	***1.000***
t1_t1	0.010	0.010	0.787	0.140	0.141	0.960	0.034	0.874	0.034	0.897
t1_t2	0.014	0.014	0.873	0.140	0.202	0.998	0.034	0.874	0.049	0.995
t1_t3	0.017	0.017	0.876	0.140	0.246	0.999	0.033	0.874	0.060	1.000
t2_t2	0.020	0.020	0.986	0.199	0.202	1.000	0.048	0.994	0.048	0.995
t2_t3	0.024	0.025	0.993	0.199	0.246	1.000	0.047	0.994	0.059	1.000
t3_t3	0.030	0.030	1.000	0.245	0.246	1.000	0.059	1.000	0.059	1.000

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table 5.1 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.6$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\widehat{C}_P$	$\alpha = .05$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\alpha = .05$	$\widehat{\beta}_{41} - \widehat{\beta}_{21}$	$\alpha = .05$	$\widehat{\beta}_{31} - \widehat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.001	0.013	0.001	-0.001	0.062	0.001	0.049	0.000	0.063
t0_t1	0.000	0.001	0.033	0.001	0.141	***0.956***	0.001	0.049	0.085	***0.899***
t0_t2	0.000	0.001	0.043	0.001	0.202	***1.000***	0.001	0.049	0.122	***0.994***
t0_t3	0.000	0.001	0.047	0.001	0.246	***1.000***	0.001	0.049	0.150	***1.000***
t1_t1	0.010	0.011	0.814	0.141	0.141	0.894	0.085	0.883	0.085	0.899
t1_t2	0.014	0.015	0.883	0.141	0.202	0.991	0.084	0.883	0.122	0.994
t1_t3	0.017	0.018	0.884	0.141	0.246	0.999	0.084	0.883	0.149	1.000
t2_t2	0.020	0.021	0.986	0.200	0.202	0.996	0.120	0.991	0.121	0.994
t2_t3	0.024	0.025	0.993	0.200	0.246	0.999	0.119	0.991	0.148	1.000
t3_t3	0.030	0.031	1.000	0.245	0.246	1.000	0.147	1.000	0.148	1.000

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

5.2 Comparison of Covariance Analysis, Variance Components Analysis-Based Approaches and Multivariate Analysis Of Variance using Family Data

We compared the performance of CovA, VC-based analysis and MANOVA in family data.

Likelihood ratio test (LRT) is used for the family related one-factor MANOVA analysis (Warton and Hudson, 2004). The likelihood of the polygenic model with a SNP adjustment (3.13), which is an alternative model, was compared with the likelihood of the polygenic model without a SNP adjustment (3.12), which is a null model. The test statistic of log-likelihood ratio test is

$$LRT = -2\ln\left(\frac{likelihood\ for\ null\ model}{likelihood\ for\ alternative\ model}\right) \quad (3.19)$$

$$= -2\ln(likelihood\ for\ null\ model)$$

$$+ 2\ln(likelihoods\ for\ alternative\ model)$$

It has a chi-square distribution with 2 degrees of freedom. We compared the log-likelihood ratio to a critical value to decide whether to reject the null model in favor of the alternative model. The logarithms of the likelihoods of the polygenic models were obtained from the SOLAR analysis results.

Table 5.2 displays the mean of the estimated pleiotropic effect (or the estimated size of the SNP effect in MANOVA), type I error and power of CovA, the genetic correlation and covariance approaches, and MANOVA for a single SNP-specific pleiotropic effect on two traits.

Under low residual polygenic correlation, CovA produces the smallest bias. The second smallest bias is from the genetic covariance approach. The genetic correlation approach generates the largest bias. If a SNP has no effect on either trait, type I error rates are conservative in the CovA, genetic correlation and covariance approaches, but inflated in MANOVA. If a SNP has effect on one trait, all methods have inflated type I error rates, which is largest for MANOVA, smallest from CovA, and second smallest from the genetic covariance approach. MANOVA had 100% rejection rate, indicating that MANOVA is ineffective for use under this scenarios. If there is a substantial association between a SNP and both traits, CovA and MANOVA are among the most powerful methods. The genetic correlation or covariance approaches have similar power.

Table 5. 2 Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, genetic correlation and covariance approaches and MANOVA using unrelated data

MAF=0.5, polygenic ρ_{g^*}=0.1 and $cov_{g^*} = 0.04$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 500

Scenario	CovA Analysis				VC Analysis								MANOVA(LRT)		
	C_P	$\widehat{C}_P$	Bias (%)	$\alpha = .05$	cov_{g_diff}	$\widehat{cov}_{g_diff}$	Bias (%)	$\alpha = .05$	ρ_{g_diff}	$\widehat{\rho}_{g_diff}$	Bias (%)	$\alpha = .05$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\alpha = .05$
t0_t0	0.000	0.000	---[a]	0.03	0.000	0.000	---[a]	0.00	0.000	0.000	---[a]	0.00	0.000	0.002	0.06
t0_t3	0.000	0.000	---[a]	0.07	0.000	0.000	---[a]	0.07	0.000	-0.003	---[a]	0.08	0.000	0.245	***1.00***
t0_t10	0.000	0.000	---[a]	0.06	0.000	-0.002	---[a]	0.07	0.000	-0.010	---[a]	0.10	0.000	0.447	***1.00***
t0_t30	0.000	-0.001	---[a]	0.06	0.000	-0.006	---[a]	0.09	0.000	-0.028	---[a]	0.12	0.000	0.774	***1.00***
t1_t1	0.010	0.010	3.6	1.00	0.010	0.010	-4.2	0.52	0.022	0.022	2.4	0.55	0.142	0.143	1.00
t3_t3	0.030	0.030	-0.6	1.00	0.030	0.029	-3.7	0.98	0.063	0.066	4.7	0.98	0.244	0.245	1.00
t10_t10	0.100	0.100	0.4	1.00	0.100	0.096	-3.9	0.99	0.180	0.198	10.2	0.99	0.448	0.445	1.00
t10_t30	0.173	0.173	0.0	1.00	0.173	0.165	-4.7	1.00	0.272	0.299	10.1	1.00	0.448	0.775	1.00
t30_t30	0.300	0.300	0.0	1.00	0.300	0.288	-4.0	1.00	0.386	0.469	21.5	1.00	0.774	0.775	1.00

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates.

We also explored the performance of CovA and VC-based analyses on the joint pleiotropic effects in the same and opposite directions. Simulations were conducted for four conditions: (1) SNP1 has 10% effects on trait1 and trait2 in the same direction and SNP2 has 30% effects on trait1 and trait2 in the same direction; (2) SNP1 has 10% effects on trait1 and trait2 in the same direction, and SNP2 has 30% effects on trait1 and trait2 in the opposite direction; (3) SNP1 has 30% effects on trait1 and trait2 in the same direction and SNP2 has 30% effects on trait1 and trait2 in the same direction; (4) SNP1 has 30% effects on trait1 and trait2 in the same direction and SNP2 has 30% effects on trait1 and trait2 in the opposite direction. We set a polygneic correlation (ρ_g) of 0.1, environmental correlation (ρ_e) of 0.6, and residual heritabilities (h_{r1}^2 and h_{r2}^2) of 0.4.

Table1 5.3 displays the distribution of 1,000 simulations of CovA, the genetic correlation approach and the genetic covariance approach on the mean of the estimate, percent bias, standard deviation, and 95% confidence interval (2.5 percentile and 97.5 percentile).

CovA maintains robustness across all conditions. The estimates are unbiased. The next smallest estimate biases are from the genetic covariance approach. The genetic correlation approach generates the largest estimate biases. In addition, its estimate biases are larger if SNP1 and SNP2 have the same directions of effects compared to if they have the opposite directions of effects, because the genetic correlation approach performs poor if the SNP(s) effects on traits are too big.

Table 5. 3 Simulation distribution comparing CovA, and genetic correlation and covariance approaches for two SNPs pleiotropic effects with the same or opposite direction

MAF=0.5, polygenic ρ_{g^*}=0.1 and $cov_{g^*} = 0.04$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$, Simulation=1000

	CovA (C_P)					
scenario	**True**	**Mean**	**Bias(%)**	**Std**	**2.5th**	**97.5th**
t10_10_t30_30[a]	0.400	0.400	0.109	0.020	0.363	0.440
t10_10_t30_R30[b]	-0.200	-0.200	-0.191	0.014	-0.228	-0.171
t30_30_t30_30[a]	0.600	0.601	0.103	0.026	0.547	0.651
t30_30_t30_R30[b]	0.000	0.000	---[a]	0.018	-0.036	0.036

*th^2_{q11}_h^2_{q21}_th^2_{q12}_h^2_{q22} denoted SNP1 with h^2_{q11}% effect on trait1 and h^2_{q21}% effect on trait2, and SNP2 with h^2_{q12}% effect on trait1 and h^2_{q22}% effect on trait2.

[a]SNP1 and SNP2 have the same direction.

[b]SNP1 and SNP2 have the opposite direction.

[a] --- indicates that the percent bias is not computed.

Table 5.3 (Continued) Simulation distribution comparing CovA, and genetic correlation and covariance approaches for two SNPs pleiotropic effects with the same or opposite direction

MAF=0.5, polygenic ρ_g=0.1 and $cov_g = 0.04$, $\rho_e = 0.6$, $h^2_{r1} = 0.4$, $h^2_{r2} = 0.4$

	VC Analysis (ρ_{g_diff})					
scenario	**True**	**Mean**	**Bias(%)**	**Std**	**2.5th**	**97.5th**
t10_10_t30_30[a]	0.200	0.535	167.308	0.065	0.431	0.681
t10_10_t30_R30[b]	-0.400	-0.349	-12.743	0.052	-0.444	-0.240
t30_30_t30_30[a]	0.240	0.621	158.872	0.070	0.503	0.769
t30_30_t30_R30[b]	0.000	-0.064	---[a]	0.067	-0.190	0.074

*th^2_{q11}_h^2_{q21}_th^2_{q12}_h^2_{q22} denoted SNP1 with h^2_{q11}% effect on trait1 and h^2_{q21}% effect on trait2, and SNP2 with h^2_{q12}% effect on trait1 and h^2_{q22}% effect on trait2.

[a]SNP1 and SNP2 have the same direction.

[b]SNP1 and SNP2 have the opposite direction.

[a] --- indicates that the percent bias is not computed.

Table 5.3 (Continued) Simulation distribution comparing CovA, and genetic correlation and covariance approaches for two SNPs pleiotropic effects with the same or opposite direction.

MAF=0.5, polygenic ρ_g=0.1 and $cov_g = 0.04$, $\rho_e = 0.6$, $h^2_{r1} = 0.4$, $h^2_{r2} = 0.4$

	VC Analysis (cov_{g_diff})					
scenario	**True**	**Mean**	**Bias(%)**	**Std**	**2.5th**	**97.5th**
t10_10_t30_30[a]	0.400	0.385	-3.827	0.034	0.323	0.453
t10_10_t30_R30[b]	-0.200	-0.215	7.432	0.027	-0.268	-0.163
t30_30_t30_30[a]	0.600	0.578	-3.623	0.045	0.492	0.668
t30_30_t30_R30[b]	0.000	-0.023	---[a]	0.038	-0.097	0.054

*th^2_{q11}_h^2_{q21}_th^2_{q12}_h^2_{q22} denoted SNP1 with h^2_{q11}% effect on trait1 and h^2_{q21}% effect on trait2, and SNP2 with h^2_{q12}% effect on trait1 and h^2_{q22}% effect on trait2.

[a]SNP1 and SNP2 have the same direction.

[b]SNP1 and SNP2 have the opposite direction.

[a] --- indicates that the percent bias is not computed.

5.3 Summary

We performed simulation studies to compare CovA, MANOVA, and conditional analysis approaches across different residual polygenic and environmental correlations using unrelated subjects. The simulation results show that MANOVA does not focus on pleiotropy, which generates large type I error rates if evaluated under the hypothesis that a SNP has effect only on one trait, which actually falls into its intended alternative hypothesis. Conditional analysis presents conservative type I error and good power if neither residual polygenic correlation nor environmental correlation is very small. Unfortunately, small values of both residual polygenic correlation and environmental correlation lead to conservative type I error rate and low power.

We compared CovA, the genetic correlation and covariance approaches to MANOVA under low residual polygenic correlation and moderate environmental correlation using related subjects. CovA is most efficient, in terms of bias, type I error and power. If a SNP has an effect on one trait, MANOVA has 100% rejection rate, demonstrating that MANOVA is ineffective for use under this scenarios. The genetic correlation and covariance approaches have similar power, but the genetic covariance approach has comparably smaller bias and type I error than the genetic correlation approach.

We also compared CovA and the genetic correlation and covariance approaches for the joint pleiotropic effects of the same and opposite directions under low residual polygenic correlation and moderate environmental correlation. We found that CovA

maintains robustness and it estimates are unbiased. The next smallest estimate biases are from the genetic covariance approach. The genetic correlation approach generates the largest estimate biases. The percent biases are larger if SNP1 and SNP2 have the same directions of effects compared to if they have the opposite directions of effects, because the genetic correlation approach performs poor if the SNP(s) effects on traits are too large.

Chapter 6: Application to Real Data

6.1 Introduction

It has long been recognized that genetic factors play an important role in the pathogenesis of osteoporosis (Ralston and Crombrugghe 2006). Over the decades, a large scale longitudinal study, the Framingham Osteoporosis Study (FOS) has been conducted to identify new candidate genes that are involved in the regulation of osteoporosis-related phenotypes (Karasik et al. 2004; Demissie et al. 2007; Kiel et al. 2007; Karasik et al. 2008; Karasik and Ferrari 2008; Karasik et al. 2010; Gupta et al. 2010, Hsu et al. 2010). Several studies have discovered shared effects of genes between multiple bone health related phenotypes. One of these studies suggested principal components analysis to linkage studies using the linear combination of several correlated bone phenotypes which facilitated detection of possible pleiotropy of chromosomal loci at different skeletal sites (Karasik et al. 2004). Some other studies performed GWAS of multiple phenotypes associated with bone fractures and found that the similarity of association between quantitative bone phenotypes may be attributed to pleiotropic effects of genes (Karasik et al. 2010; Gupta et al. 2010). Karasik et al (2010) conducted the first phenomic scan and found that BMD at the lumbar spine and femoral neck are indeed genetically alike. They listed thirty-eight potential candidate pleiotropic genetic variants associated at $p < .001$ with both FN and LS BMD (some SNPs in linkage disequilibrium with $r^2 \geq 0.5$). Further empirical research is needed to evaluate the pleiotropic effect of these putative

SNPs and validate Karasik's approach on selection of presumably pleiotropic candidate genes.

6.2 The Framingham Osteoporosis Study (FOS)

The FOS is an ancillary study of the Framingham Heart Study (FHS). The FHS is a population-based, multigenerational cohort study established in 1948 to examine risk factors for cardiovascular disease in a group of subjects from Framingham, MA (Dawber and Moore 1951; Dawber et al. 1963). In the original cohort, 5,209 participants (2,336 men and 2,873 women) were enrolled at the first examination and were examined every 2 years. The offspring cohort was initiated in 1971 to evaluate the role of genetic factors in the etiology of coronary artery disease, consisting of 5,124 participants (2,483 men and 5,641 women) who were the adult offspring and their spouses of the original cohort study and were followed every four years (Feinleib et al. 1975; Tucker et al. 2006; Demissie et al. 2002; Karasik et al. 2004; Cusano et al. 2012). The sample of FOS was drawn from two cohorts of the FHS. BMD measurement, a diagnostic test used to measure the amount of mineral in bone, were collected from 1992 to 1996 during the 22nd and 24th examination for the original cohort and from 1996 to 2001 during the 6th and 7th examination for the offspring (Tucker et al. 2006, Demissie et al. 2007; Kiel et al. 2007; Gupta et al. 2011). For this study, there were 3480 participants who had BMD measurements at the spine and hip, 720 from the original cohort (270 men and 450 women) and 2760 from the offspring cohort (1220 men and 1540 women). Because the

contribution of genetic factors to BMD has been show to differ by gender and bone fracture are more common in women, Karasik evaluated SNPs in women (Karasik et al. 2010). Therefore, our research also focuses on women. We use cohort-specific residual phenotypes adjusted for covariates for analysis, which have been previously described (Karasik et al. 2010).

The FOS uses the genotype data from the FHS SNP Health Association Resource (SHARe) project initiated in 2007. The SHARe genotyping was conducted using the Affymetrix 500K mapping array plus Affymetrix 50K supplemental array in over 9,300 Framingham Study participants with DNA available. The details of genotype cleaning including quality control and population substructure have been published elsewhere (Karasik et al. 2010).

In order to determine the nature of the shared SNPs, we apply (1) CovA to a sample of genetically unrelated subjects and (2) CovA, genetic correlation and genetic covariance approaches to a sample of family related subjects.

6.3 Pleiotropic Effect Using Unrelated Data

A sub-set of 879 unrelated women (237 original cohort and 642 offspring cohort) who had phenotypic measurements and 38 SNPs available for analysis was randomly selected from a sample of 8481 family related subjects. The correlation between FNBMD and LSBMD was 0.54 (p<0.0001) in this sample. Multivariate analysis was carried out using the PROC MODEL procedure in the SAS software package to estimate the

regression coefficients of FNBMD and LSBMD simultaneously. Then we implemented CovA to compute the SNP-specific pleiotropic effect. The bootstrap resampling PCTL method (1,000 repetitions) was implemented to construct 95% confidence intervals for the statistical hypothesis testing.

Figure 6.1 displays the estimated SNP-specific pleiotropic effect across SNPs using CovA in unrelated subjects. It also reports the SNP R^2s on FNBMD and LSBMD to illuminate the relationship between the shared SNP effect and individual SNP effects on traits. All of the estimated pleiotropic SNP effects are greater than 0 and lay between the individual SNP R^2 effects.

The estimated beta coefficients, R^2s, pleiotropic effects and 95% confidence intervals are displayed in Table 6.1. Note that the estimated beta coefficients of our study using unrelated subjects are different from that of Karasik's study using family subjects. We found that none of these candidate SNPs present significant pleiotropic effects at a significance level of 0.05 if we treat each participants as independent of others.

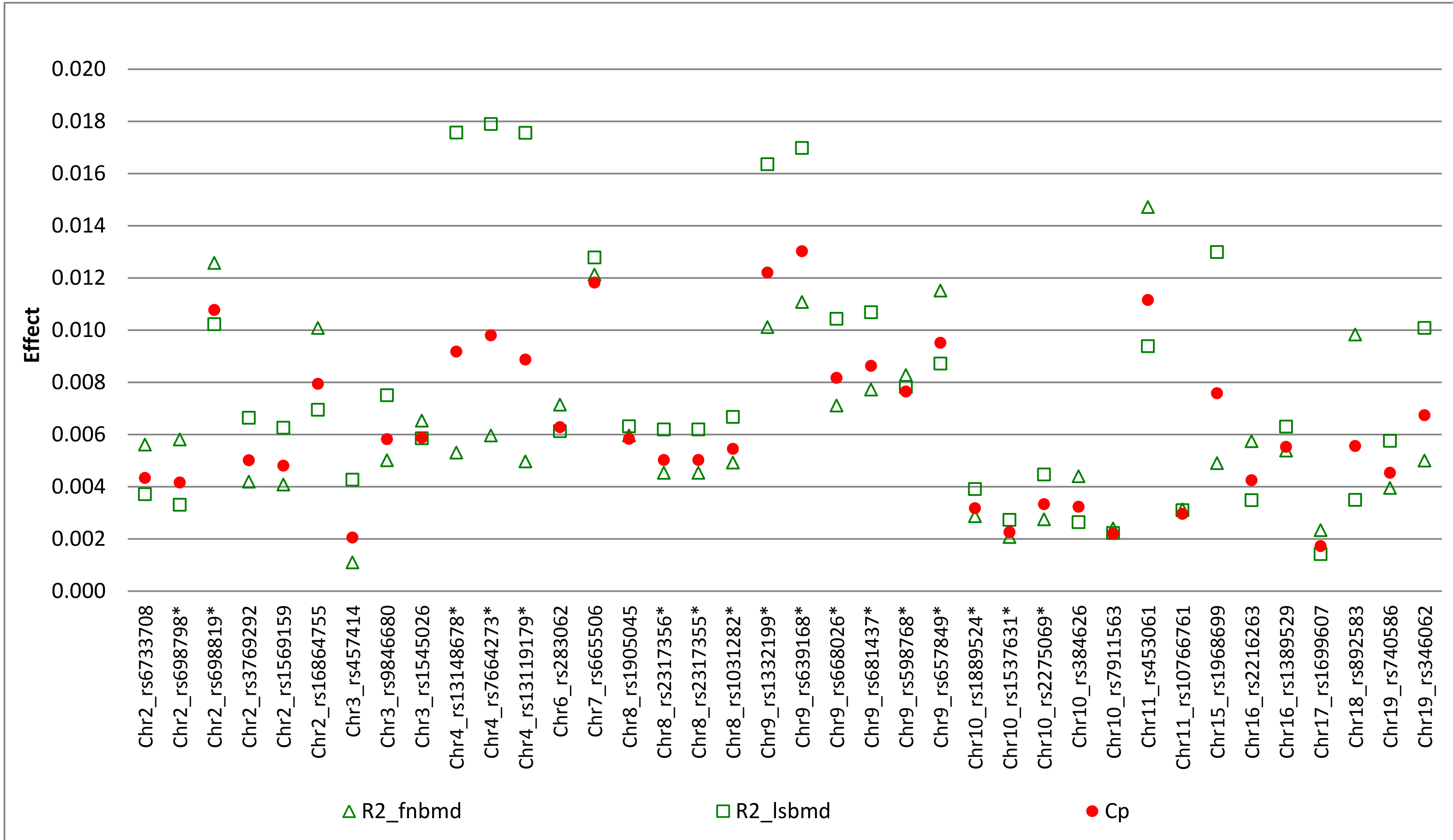

Figure 6. 1 Estimated SNP-specific pleiotropic effect, and SNP individual effects on FNBMD and LSBMD using CovA for unrelated subjects in Framingham Osteoporosis-BMDs study.

Table 6. 1 Single SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA for unrelated subjects in Framingham Osteoporosis-BMDs study.

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{C}_P$	2.5th	97.5th
rs6733708	2	37814216	0.36	-0.110	-0.087	0.006	0.004	0.004	-0.0012	0.0252
rs698798*	2	44544018	0.08	-0.196	-0.145	0.006	0.003	0.004	-0.0014	0.0226
rs698819*	2	44566433	0.08	-0.304	-0.269	0.013	0.010	0.011	-2E-05	0.0342
rs3769292	2	173434505	0.10	-0.153	-0.189	0.004	0.007	0.005	-0.0019	0.0236
rs1569159	2	181189870	0.09	0.155	0.189	0.004	0.006	0.005	-0.0009	0.0226
rs16864755	2	223982866	0.10	0.236	0.192	0.010	0.007	0.008	-0.0005	0.0372
rs457414	3	10177884	0.32	-0.049	-0.095	0.001	0.004	0.002	-0.0022	0.0167
rs9846680	3	179603561	0.08	-0.192	-0.231	0.005	0.007	0.006	-0.0009	0.0210
rs1545026	3	196061661	0.13	0.171	0.159	0.007	0.006	0.006	-0.0008	0.0265
rs13148678*	4	74077760	0.17	0.138	0.247	0.005	0.018	0.009	-0.0015	0.0340
rs7664273*	4	74226102	0.17	0.148	0.251	0.006	0.018	0.010	-0.0013	0.0345
rs13119179*	4	74262694	0.17	0.135	0.249	0.005	0.018	0.009	-0.0019	0.0323

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\hat{C}_P$ is the estimated pleiotropic effect; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. p $< .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table 6.1 (Continued) Single SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA for unrelated subjects in Framingham Osteoporosis-BMDs study.

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{C}_P$	2.5th	97.5th
rs283062	6	118728149	0.43	0.117	0.106	0.007	0.006	0.006	-0.0012	0.0248
rs665506	7	132023073	0.35	0.164	0.165	0.012	0.013	0.012	0.0000	0.0375
rs1905045	8	75062568	0.21	0.131	0.132	0.006	0.006	0.006	-0.0010	0.0270
rs2317356*	8	137106399	0.32	0.103	0.119	0.005	0.006	0.005	-0.0012	0.0243
rs2317355*	8	137106698	0.32	0.103	0.119	0.005	0.006	0.005	-0.0012	0.0243
rs1031282*	8	137123839	0.32	0.108	0.123	0.005	0.007	0.005	-0.0011	0.0236
rs1332199*	9	9493477	0.17	0.191	0.238	0.010	0.016	0.012	-0.0002	0.0387
rs639168*	9	9529574	0.12	0.227	0.275	0.011	0.017	0.013	-0.0003	0.0414
rs668026*	9	9571692	0.31	0.127	0.150	0.007	0.010	0.008	-0.0006	0.0328
rs681437*	9	9572128	0.31	0.133	0.153	0.008	0.011	0.009	-0.0005	0.0345
rs598768*	9	9579682	0.35	0.133	0.127	0.008	0.008	0.008	-0.0007	0.0319
rs657849*	9	9583722	0.36	0.157	0.134	0.012	0.009	0.010	-0.0005	0.0360

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\hat{C}_P$ is the estimated pleiotropic effect; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table 6.1 (Continued) Single SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA for unrelated subjects in Framingham Osteoporosis-BMDs study.

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{C}_P$	2.5th	97.5th
rs1889524*	10	7657473	0.46	-0.076	-0.087	0.003	0.004	0.003	-0.0017	0.0218
rs1537631*	10	7657765	0.45	-0.065	-0.073	0.002	0.003	0.002	-0.0021	0.0187
rs2275069*	10	7658692	0.46	-0.075	-0.093	0.003	0.004	0.003	-0.0018	0.0229
rs384626	10	60806221	0.49	-0.091	-0.069	0.004	0.003	0.003	-0.0018	0.0202
rs7911563	10	60878297	0.41	-0.069	-0.065	0.002	0.002	0.002	-0.0017	0.0180
rs453061	11	8227452	0.01	0.764	0.598	0.015	0.009	0.011	-0.0009	0.0376
rs10766761	11	20972609	0.28	-0.089	-0.087	0.003	0.003	0.003	-0.0015	0.0211
rs1968699	15	97529279	0.06	-0.209	-0.334	0.005	0.013	0.008	-0.0021	0.0278
rs2216263	16	47911005	0.39	0.107	0.082	0.006	0.003	0.004	-0.0013	0.0234
rs1389529	16	53286410	0.07	0.211	0.223	0.005	0.006	0.006	-0.0011	0.0266
rs1699607	17	70130820	0.49	-0.068	-0.052	0.002	0.001	0.002	-0.0020	0.0165
rs892583	18	43170372	0.27	0.161	0.094	0.010	0.003	0.006	-0.0025	0.0288
rs740586	19	48888371	0.22	0.102	0.121	0.004	0.006	0.005	-0.0012	0.0223
rs346062	19	48890417	0.43	-0.099	-0.138	0.005	0.010	0.007	-0.0013	0.0296

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\hat{C}_P$ is the estimated pleiotropic effect; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

6.4 Pleiotropic Effect Using Family Data

Among 8481 family related subject, 1544 related women (334 original cohort and 1210 offspring cohort) in 770 families had phenotypic measurements and 38 SNPs available for analysis. The association between FNBMD and LSBMD was 0.56 (p<0.0001). SOLAR'S polygenic analysis function for multivariate models was used for our VC-based analysis approaches and its estimated regression coefficients was used in CovA. The bootstrap resampling PCTL method (1,000 repetitions) was applied to construct confidence intervals for statistical hypothesis testing.

Figure 6.2 displays the estimated SNP-specific pleiotropic effects of CovA, genetic covariance and genetic correlation approaches across SNPs in family related subjects. In order to examine the relationship between shared SNP effect and individual SNP effects on traits, we also reported the SNP R^2s that show to be similar for both BMD traits. The estimated pleiotropic SNP effects of CovA are all greater than 0 and lie between the SNP R^2s. In addition, the neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$) provides similar estimated pleiotropic effect values. However, the genetic covariance and correlation approaches yield both positive and negative estimated values. Furthermore, the shared genetic effects of the neighboring SNPs show big differences in genes PTPRD on chromosome 9.

Table 6.2 compares the estimated pleiotropic effects among these three approaches. More detailed estimated beta coefficients, R^2s, pleiotropic effects and 95% confidence intervals are presented in Table S7. The estimated beta coefficients of our study based on family member are similar to that of Karasik's study where beta

coefficients were consistent in their sign and similar for both BMD traits, thus Karasik concluded that these SNPs may be pleiotropic (Karasik et al. 2010). In CovA, we found that all 38 SNPs are identified as having significant pleiotropic effect. The number of significant SNPs is 15 in the genetic covariance approach and 9 in the genetic correlation approach. Furthermore, the genetic covariance and correlation approaches yield some negative pleiotropic effects that contradict their sign of beta coefficients. The simulation results from chapter 3 have demonstrated that the genetic covariance and correlation approaches may not be warranted when the SNP effects on phenotypes are very small, which keep us from drawing firm conclusions about these methods.

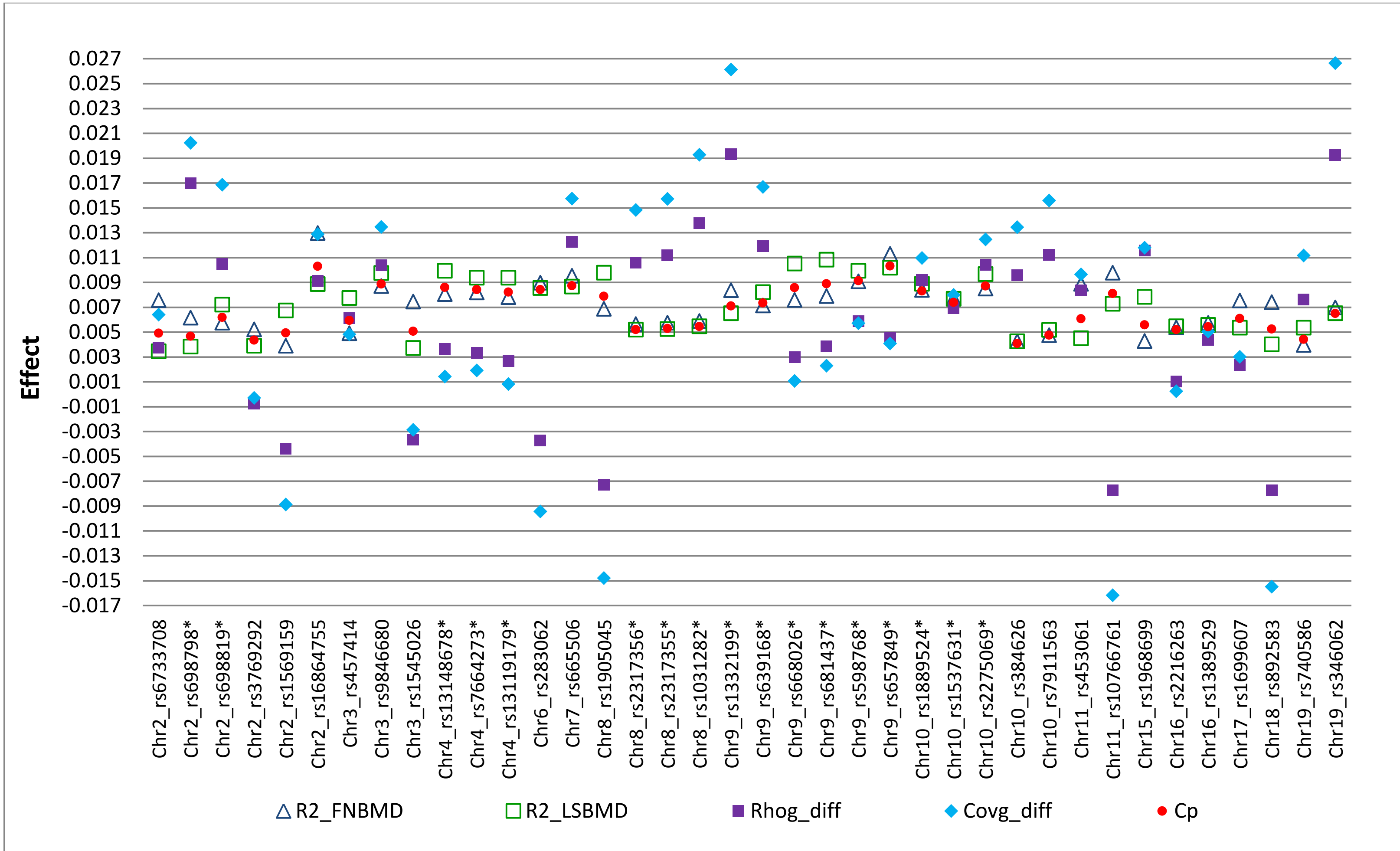

Figure 6. 2 Estimated SNP-specific pleiotropic effect, and SNP individual effects on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study.

Table 6. 2 SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

SNP	Chr.	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{\rho}_{g_diff}$	2.5th	97.5th	$\widehat{cov}_{g_diff}$	2.5th	97.5th	$\hat{C}_P$	2.5th	97.5th
rs6733708	2	0.008	0.003	0.004	-0.006	0.017	0.006	-0.005	0.020	0.005	***0.001***	***0.011***
rs698798*	2	0.006	0.004	0.017	***0.004***	***0.038***	0.020	***0.007***	***0.036***	0.005	***0.001***	***0.009***
rs698819*	2	0.006	0.007	0.010	-0.006	0.032	0.017	***0.002***	***0.036***	0.006	***0.002***	***0.013***
rs3769292	2	0.005	0.004	-0.001	-0.014	0.009	0.000	-0.010	0.010	0.004	***0.001***	***0.009***
rs1569159	2	0.004	0.007	-0.004	-0.015	0.005	-0.009	-0.021	0.005	0.005	***0.001***	***0.011***
rs16864755	2	0.013	0.009	0.009	-0.005	0.029	0.013	-0.002	0.032	0.010	***0.004***	***0.018***
rs457414	3	0.005	0.008	0.006	-0.003	0.021	0.005	-0.007	0.018	0.006	***0.002***	***0.011***
rs9846680	3	0.009	0.010	0.010	-0.002	0.032	0.013	-0.003	0.033	0.009	***0.004***	***0.016***
rs1545026	3	0.007	0.004	-0.004	-0.015	0.006	-0.003	-0.014	0.011	0.005	***0.001***	***0.011***
rs13148678*	4	0.008	0.010	0.004	-0.012	0.018	0.001	-0.014	0.019	0.009	***0.003***	***0.015***
rs7664273*	4	0.008	0.009	0.003	-0.009	0.017	0.002	-0.013	0.019	0.008	***0.003***	***0.016***
rs13119179*	4	0.008	0.009	0.003	-0.009	0.017	0.001	-0.013	0.017	0.008	***0.003***	***0.015***

Note: R^2 is the coefficient of determination; $\hat{\rho}_{g_diff}$ is the estimated pleiotropic effect in the genetic correlation approach; $\widehat{cov}_{g_diff}$ is the estimated pleiotropic effect in the genetic covariance approach; $\hat{C}_P$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table 6.2 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

SNP	Chr.	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{\rho}_{g_diff}$	2.5^{th}	97.5^{th}	$\widehat{cov}_{g_diff}$	2.5^{th}	97.5^{th}	$\hat{C}_P$	2.5^{th}	97.5^{th}
rs283062	6	0.009	0.009	-0.004	-0.016	0.008	-0.009	-0.023	0.006	0.009	***0.004***	***0.015***
rs665506	7	0.010	0.009	0.012	***5E-04***	***0.031***	0.016	***0.001***	***0.035***	0.009	***0.003***	***0.017***
rs1905045	8	0.007	0.010	-0.007	-0.020	0.003	-0.015	-0.031	0.003	0.008	***0.003***	***0.014***
rs2317356*	8	0.006	0.005	0.011	***1E-04***	***0.027***	0.015	***0.003***	***0.031***	0.005	***0.001***	***0.012***
rs2317355*	8	0.006	0.005	0.011	***4E-04***	***0.027***	0.016	***0.003***	***0.032***	0.005	***0.001***	***0.012***
rs1031282*	8	0.006	0.005	0.014	***0.003***	***0.033***	0.019	***0.005***	***0.037***	0.005	***0.001***	***0.012***
rs1332199*	9	0.008	0.007	0.019	***0.005***	***0.042***	0.026	***0.009***	***0.045***	0.007	***0.002***	***0.013***
rs639168*	9	0.007	0.008	0.012	-0.004	0.029	0.017	***0.002***	***0.035***	0.007	***0.002***	***0.014***
rs668026*	9	0.008	0.011	0.003	-0.008	0.016	0.001	-0.014	0.017	0.009	***0.004***	***0.016***
rs681437*	9	0.008	0.011	0.004	-0.007	0.016	0.002	-0.013	0.017	0.009	***0.004***	***0.016***
rs598768*	9	0.009	0.010	0.006	-0.006	0.019	0.006	-0.010	0.020	0.009	***0.004***	***0.017***
rs657849*	9	0.011	0.010	0.005	-0.009	0.018	0.004	-0.013	0.019	0.010	***0.005***	***0.018***

Note: R^2 is the coefficient of determination; $\hat{\rho}_{g_diff}$ is the estimated pleiotropic effect in the genetic correlation approach; $\widehat{cov}_{g_diff}$ is the estimated pleiotropic effect in the genetic covariance approach; $\hat{C}_P$ is the estimated pleiotropic effect in CovA; 2.5^{th} and 97.5^{th} display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table 6.2 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

SNP	Chr.	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{\rho}_{g_diff}$	2.5th	97.5th	$\widehat{cov}_{g_diff}$	2.5th	97.5th	$\hat{C}_P$	2.5th	97.5th
rs1889524*	10	0.008	0.009	0.009	-0.003	0.029	0.011	-0.004	0.029	0.008	***0.003***	***0.016***
rs1537631*	10	0.008	0.008	0.007	-0.004	0.023	0.008	-0.005	0.025	0.008	***0.003***	***0.015***
rs2275069*	10	0.009	0.010	0.010	-0.002	0.032	0.012	-0.003	0.032	0.009	***0.003***	***0.017***
rs384626	10	0.004	0.004	0.010	-0.003	0.023	0.013	***0.002***	***0.025***	0.004	***0.001***	***0.009***
rs7911563	10	0.005	0.005	0.011	-3E-04	0.029	0.016	***0.001***	***0.032***	0.005	***0.001***	***0.010***
rs453061	11	0.009	0.004	0.008	***0.002***	***0.019***	0.010	***0.002***	***0.019***	0.006	***0.001***	***0.012***
rs10766761	11	0.010	0.007	-0.008	-0.022	0.003	-0.016	***-0.031***	***-3E-04***	0.008	***0.003***	***0.015***
rs1968699	15	0.004	0.008	0.012	***0.001***	***0.027***	0.012	-0.001	0.027	0.006	***0.002***	***0.012***
rs2216263	16	0.005	0.005	0.001	-0.008	0.010	0.000	-0.012	0.013	0.005	***0.001***	***0.011***
rs1389529	16	0.006	0.006	0.004	-0.007	0.022	0.005	-0.013	0.022	0.006	***0.001***	***0.012***
rs1699607	17	0.008	0.005	0.002	-0.008	0.018	0.003	-0.009	0.020	0.006	***0.001***	***0.012***
rs892583	18	0.007	0.004	-0.008	-0.019	0.002	-0.015	***-0.028***	***-0.004***	0.005	***0.001***	***0.011***
rs740586	19	0.004	0.005	0.008	-0.004	0.019	0.011	***4E-04***	***0.023***	0.004	***0.001***	***0.010***
rs346062	19	0.007	0.007	0.019	***0.003***	***0.047***	0.027	***0.010***	***0.049***	0.007	***0.002***	***0.014***

Note: R^2 is the coefficient of determination; $\hat{\rho}_{g_diff}$ is the estimated pleiotropic effect in the genetic correlation approach; $\widehat{cov}_{g_diff}$ is the estimated pleiotropic effect in the genetic covariance approach; $\hat{C}_P$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. p $< .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

6.5 Summary

We evaluated the pleiotropic effect of 38 candidate SNPs identified by Karasik (Karasik et al. 2010) in data from the Framingham Osteoporosis Study.

With family subjects, all 38 SNPs had significantly positive pleiotropic effects on both BMD traits in CovA. In addition, the neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$) provided similar estimated pleiotropic effect values. Fifteen SNPs in the genetic covariance approach and nine SNPs in the genetic correlation approach were statistically significant at $\alpha = 0.05$. However, some estimated pleiotropic effects using the genetic covariance and correlation approaches were negative that contradicted their sign of beta coefficients. Because when SNP effects on phenotypes are very small, the genetic covariance and correlation approaches may not be warranted, it prevents us from drawing firm conclusions about these methods.

With unrelated subjects, these candidate SNPs present positive pleiotropic effects on both BMD traits in CovA, but none of them was statistically significant.

Chapter 7: Discussion and Conclusions

7.1 Discussion

Many genetic studies contain multiple phenotypes. These phenotypes are often genetically correlated by the influence of a same gene or genes. Earlier studies have put a great deal of effort into the methods of multivariate analysis of quantitative variables However, to date, no previously developed method can be directly applied to evaluate the pleiotropic effects. The purpose of this dissertation is to provide insight into the SNP-specific pleiotropic effect and develop several effective statistical approaches based on multivariate techniques.

7.1.1 The Genetic Correlation and Genetic Covariance Approaches

In Chapter 3, we have investigated variance components analysis of shared genetic and environmental correlations in bivariate polygenic models. Then we have proposed two VC-based methods of examining the contribution of specific genetic marker(s) to polygenic correlation or covariance of traits for analysis of SNP specific pleiotropic effect by comparing genetic correlations or covariances from bivariate polygenic models with and without adjustment for the SNP. The genetic correlation and genetic covariance approaches can be applied for family-based association studies including multiple markers.

Based on the simulation results, we found that the genetic correlation and covariance approaches are affected by the residual polygenic correlation level. Both

approaches are effective only under low residual polygenic correlation condition. They may generate greater bias, higher type I error if a SNP has effect on one trait, and lower power under larger residual polygenic genetic correlation condition. Under low polygenic correlation condition, the genetic covariance approach has smaller estimate bias and type I error than the genetic correlation approach, especially when the SNP effects on traits are large. Both approaches produce a similar power. Changing the magnitude of MAF or environmental correlation does not alter the results materially.

7.1.2 Covariance Analysis

In Chapter 4, we have developed a novel method referred to as covariance analysis (CovA) for SNP-specific pleiotropic effect which can incorporate multiple markers in population-based studies with samples of families or unrelated subjects. CovA allows estimating and testing pleiotropic effect of a marker(s) by examining the contribution of specific genetic marker(s) to covariance of phenotypes using multivariate linear regression models. This approach is especially desirable in that pleiotropic effect measures are easy to calculate and interpret.

Because of violation of normality, regular test based on a normal distribution is not valid. We recommended bootstrap resampling method for generating an empirical sampling distribution for the parameter of interest to construct confidence intervals and test for significance. The operation is simple and easy to use. Several approaches of bootstrapping confidence intervals were computed and compared in CovA. It showed through the simulation study that normal approximation method almost never rejects the

null hypothesis if a SNP has no effect on either trait, and generates poor power if a SNP has very small effect on both traits. The bootstrapping PCTL method improves type I error rate and maintains good power consistently across different SNP heritabilities.

It is worthy to note that CovA remains consistent regardless of the residual polygenic genetic correlation, environmental correlations and MAFs. All simulation results showed unbiased estimate, well controlled type I error and good power. We also found that CovA may be slightly influenced by environmental correlation. Larger environmental correlation yields bigger bias, increased type I error if a SNP has no effect on either trait, decreased type I error if a SNP has effect on one trait, and increased power if a SNP has effects on both traits. Additionally, CovA is also affected very little by different MAFs. Bias, type I error and power may have a corresponding increase as the MAF decrease. Based on the results from these simulations, we recommend CovA as a general method for bivariate SNP-specific pleiotropic effect analysis.

7.1.3 Comparison of Approaches

In comparison to MANOVA and Conditional analysis approaches for the pleiotropic SNP effect of bivariate trait using unrelated subjects, only CovA can be used to directly quantify pleiotropic effect. Instead of testing the pleiotropic effect, MANOVA is actually examining a SNP effect on at least one trait rather than both traits. We found that the rejections rate of MANOVA can be very large when evaluated under the hypothesis that a SNP has effect on one trait. Therefore, MANOVA does not focus on pleiotropy. We evaluated the performance of conditional analysis through different

residual polygenic and environmental correlation simulations. It presents conservative type I error and good power if neither polygenic correlation nor environmental correlation is very small. Small values of both residual polygenic correlation and environmental correlation lead to very conservative type I error rate and very low power.

We compared CovA and VC-based approaches to MANOVA for the SNP-specific pleiotropic effect using family related subjects under low polygenic correlation level. CovA is the most efficient approach, in terms of bias, type I error and power. If a SNP has an effect on one trait, MANOVA had 100% rejection rate, indicating that MANOVA cannot be used under this scenario. The genetic correlation or covariance approaches have similar power, but the genetic covariance approach has comparably smaller bias and type I error than the genetic correlation approach.

We also compared the performance of CovA and VC-based analyses on the joint pleiotropic effects in the same and opposite directions under low residual polygenic correlation level. CovA maintains robustness across all scenarios with unbiased estimates. The next smallest estimate biases are from the genetic covariance approach. The genetic correlation approach generates the largest estimate biases. Compared to the opposite directions, the genetic correlation approach yields larger bias if the SNPs effects are in the same directions, which relates well with the fact that the genetic correlation approach performs poor if the SNP(s) effects on traits are too large.

7.1.4 Application to Real Data

We have applied our newly proposed approaches to 38 candidate SNPs identified by Karasik (Karasik et al. 2010) for the pleiotropic effect analysis in FOS.

We found that all 38 SNPs have significantly positive pleiotropic effects on FN and LS BMDs using CovA with family data. Additionally, there are similar pleiotropic effects in the neighboring SNPs with LD. Only 15 SNPs in the genetic covariance approach and 9 SNPs in the genetic correlation approach are identified as significant. Because, the genetic covariance and correlation approaches may not be warranted when SNP effects on phenotypes are very small, it prevents us from drawing firm conclusions about these two approaches.

7.2 Limitation and Future Work

CovA is a multivariate regression-based approach, and the genetic correlation and covariance methods are variance components analysis-based approaches, so they all require a multivariate normal distributional assumption.

The genetic correlation and covariance approaches are limited for only family-based data whereas CovA can be applied for either family-based or population-based studies of families or genetically unrelated subjects.

Bootstrapping percentile confidence interval method is required for the statistical hypothesis testing, so the computational time is a concern for these approaches, especially in family data.

In this dissertation, we focus on two phenotypes modeling in pleiotropic effect analysis. Therefore, as future work, it would be interesting to extend CovA to more than two phenotypes data.

Another area of future research will be to explore how to estimate a pleiotropic SNP effect when the phenotypes are dichotomous, as is very common in genetic association study. For example, in osteoporosis study, we may also be interested in investigating the share effect of a gene or genes on the fractures of multiple skeletal sites.

7.3 Conclusions

In conclusion, in this dissertation we have proposed three approaches for SNP-specific pleiotropic effect of bivariate quantitative phenotypes: the genetic correlation and genetic covariance approaches (variance components analysis-based approach) and CovA (multivariate regression model-based approach). CovA outperforms the genetic correlation and covariance approaches. It has unbiased estimation, lowest type I error and highest power, which remains consistent regardless of the polygenic correlation.

APPENDIX

Table S 1 Simulation distribution and empirical bootstrap distribution comparison for the genetic correlation and covariance approaches

MAF=0.5, polygenic $\rho_{g^*} = 0.6, cov_{g^*} = 0.24$, $\rho_e = 0.6,\ h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 500

Scenario	Genetic Correlation Approach (ρ_{g_diff}) at $\alpha = .05$											
	True Value	Simulation Distribution					Bootstrap Distribution					
		Est	SD	2.5th	97.5th	Bias(%)	Est	SD	2.5th	97.5th	Bias(%)	Type I Error /Power
t0_t0	0.000	0.000	0.001	-0.003	0.003	---[a]	0.000	0.003	-0.004	0.005	---[a]	0.000
t0_t3	0.000	-0.022	0.011	-0.045	-0.001	---[a]	-0.021	0.012	-0.047	-0.001	---[a]	0.500
t0_t10	0.000	-0.069	0.020	-0.107	-0.029	---[a]	-0.069	0.021	-0.112	-0.027	---[a]	0.940
t0_t30	0.000	-0.182	0.034	-0.248	-0.113	---[a]	-0.185	0.035	-0.255	-0.117	---[a]	1.000
t1_t1	0.010	0.010	0.006	0.001	0.024	3.8	0.010	0.010	-0.001	0.035	5.5	0.50
t3_t3	0.028	0.029	0.010	0.011	0.052	2.8	0.029	0.015	0.002	0.060	2.2	0.850
t10_t10	0.080	0.088	0.020	0.055	0.136	10.0	0.086	0.025	0.042	0.142	7.5	0.880
t10_t30	0.121	0.106	0.030	0.053	0.173	-12.3	0.104	0.031	0.048	0.171	-14.2	1.000
t30_t30	0.171	0.210	0.034	0.154	0.283	22.2	0.210	0.035	0.150	0.288	22.8	1.000

[a] --- indicates that the percent bias is not computed.

Table S 1 (Continued) Simulation distribution and empirical bootstrap distribution comparison for the genetic correlation and covariance approaches

MAF=0.5, polygenic $\rho_{g^*} = 0.6, cov_{g^*} = 0.24$, $\rho_e = 0.6,\ h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 500

Scenario	Genetic Covariance Approach(cov_{g_diff}) at $\alpha = .05$											
	True Value	Simulation Distribution					Bootstrap Distribution					
		Est	SD	2.5th	97.5th	Bias(%)	Est	SD	2.5th	97.5th	Bias(%)	Type I Error /Power
t0_t0	0.000	0.000	0.001	-0.001	0.002	---[a]	0.000	0.002	-0.002	0.004	---[a]	0.001
t0_t3	0.000	-0.004	0.005	-0.013	0.007	---[a]	-0.003	0.005	-0.013	0.008	---[a]	0.130
t0_t10	0.000	-0.012	0.010	-0.030	0.008	---[a]	-0.012	0.010	-0.032	0.009	---[a]	0.270
t0_t30	0.000	-0.039	0.017	-0.070	-0.005	---[a]	-0.040	0.017	-0.072	-0.006	---[a]	0.640
t1_t1	0.010	0.008	0.005	-0.001	0.019	-21.4	0.008	0.008	-0.003	0.027	-22.9	0.45
t3_t3	0.030	0.023	0.009	0.008	0.041	-24.5	0.022	0.011	0.001	0.045	-25.0	0.750
t10_t10	0.100	0.077	0.015	0.049	0.107	-23.3	0.074	0.019	0.037	0.113	-26.3	0.840
t10_t30	0.173	0.125	0.020	0.088	0.166	-27.8	0.121	0.021	0.080	0.164	-29.9	1.000
t30_t30	0.300	0.229	0.026	0.179	0.282	-23.5	0.226	0.026	0.176	0.279	-24.6	1.000

[a] --- indicates that the percent bias is not computed.

Table S 2 Simulation distribution comparing the genetic correlation and covariance approaches under different environmental correlation levels

MAF=0.5, polygenic ρ_{g^*}=0.1 and $cov_{g^*} = 0.04$, $\boxed{\rho_e = 0.0}$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$, simulation =1000

	ρ_{g_diff}						cov_{g_diff}					
scenario	**True**	**Est**	**SD**	**2.5th**	**97.5th**	**Bias(%)**	**True**	**Est**	**SD**	**2.5th**	**97.5th**	**Bias(%)**
t0_t0	0.000	0.000	0.002	-0.005	0.004	---[a]	0.000	0.000	0.001	-0.002	0.002	---[a]
t0_t05	0.000	-0.001	0.006	-0.014	0.011	---[a]	0.000	0.000	0.002	-0.005	0.004	---[a]
t0_t1	0.000	-0.001	0.008	-0.019	0.016	---[a]	0.000	0.000	0.003	-0.007	0.007	---[a]
t0_t2	0.000	-0.003	0.012	-0.028	0.020	---[a]	0.000	-0.001	0.005	-0.010	0.009	---[a]
t0_t3	0.000	-0.004	0.014	-0.036	0.023	---[a]	0.000	-0.001	0.006	-0.013	0.010	---[a]
t0_t10	0.000	-0.012	0.026	-0.068	0.040	---[a]	0.000	-0.002	0.010	-0.024	0.018	---[a]
t0_t30	0.000	-0.032	0.043	-0.116	0.048	---[a]	0.000	-0.007	0.018	-0.042	0.026	---[a]
t05_t05	0.005	0.011	0.007	-0.002	0.027	-2.3	0.005	0.005	0.003	-0.001	0.012	-6.2
t05_t1	0.007	0.015	0.009	-0.002	0.035	-2.8	0.007	0.007	0.004	-0.001	0.014	-7.0
t05_t2	0.010	0.021	0.012	-0.001	0.048	-3.8	0.010	0.009	0.005	***0.000***	***0.020***	-6.7
t05_t3	0.012	0.025	0.014	-0.001	0.053	-6.4	0.012	0.011	0.006	***0.000***	***0.024***	-8.4
t1_t1	0.010	0.021	0.010	***0.003***	***0.045***	-4.0	0.010	0.009	0.004	***0.001***	***0.019***	-8.9
t1_t2	0.014	0.030	0.012	***0.007***	***0.053***	-2.6	0.014	0.013	0.005	***0.003***	***0.023***	-7.5
t1_t3	0.017	0.036	0.014	***0.010***	***0.065***	-4.9	0.017	0.016	0.006	***0.004***	***0.027***	-9.1
t2_t2	0.019	0.044	0.014	***0.020***	***0.076***	3.3	0.020	0.019	0.006	***0.008***	***0.032***	-3.0
t2_t3	0.023	0.053	0.016	***0.024***	***0.087***	2.5	0.024	0.024	0.007	***0.011***	***0.038***	-3.8
t3_t3	0.028	0.064	0.017	***0.035***	***0.102***	1.5	0.030	0.028	0.007	***0.015***	***0.044***	-6.1
t10_t10	0.080	0.197	0.030	***0.140***	***0.260***	9.3	0.100	0.097	0.014	***0.070***	***0.126***	-3.5
t10_t30	0.121	0.296	0.039	***0.221***	***0.378***	8.9	0.173	0.165	0.018	***0.127***	***0.203***	-4.8
t30_t30	0.171	0.466	0.046	***0.382***	***0.560***	20.9	0.300	0.288	0.023	***0.243***	***0.332***	-4.1

* Italic bold values are the CIs excluding 0. [a] --- indicates that the percent bias is not computed.

Table S 2 (Continued) Simulation distribution comparing the genetic correlation and covariance approaches under different environmental correlation levels

MAF=0.5, polygenicρ_{g^*}=0.1 and $cov_{g^*} = 0.04$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$, Simulation=1000

	ρ_{g_diff}						cov_{g_diff}					
scenario	**True**	**Est**	**SD**	**2.5th**	**97.5th**	**Bias(%)**	**True**	**Est**	**SD**	**2.5th**	**97.5th**	**Bias(%)**
t0_t0	0.000	0.000	0.002	-0.003	0.005	---[a]	0.000	0.000	0.001	-0.001	0.002	---[a]
t0_t05	0.000	0.000	0.006	-0.012	0.012	---[a]	0.000	0.000	0.002	-0.005	0.005	---[a]
t0_t1	0.000	-0.001	0.008	-0.015	0.017	---[a]	0.000	0.000	0.003	-0.006	0.007	---[a]
t0_t2	0.000	-0.002	0.011	-0.023	0.021	---[a]	0.000	0.000	0.005	-0.010	0.009	---[a]
t0_t3	0.000	-0.003	0.014	-0.029	0.025	---[a]	0.000	-0.001	0.006	-0.011	0.011	---[a]
t0_t10	0.000	-0.010	0.025	-0.056	0.040	---[a]	0.000	-0.002	0.010	-0.021	0.018	---[a]
t0_t30	0.000	-0.028	0.043	-0.106	0.058	---[a]	0.000	-0.007	0.017	-0.039	0.028	---[a]
t05_t05	0.011	0.011	0.009	-0.002	0.033	2.8	0.005	0.005	0.004	-0.001	0.013	-2.2
t05_t1	0.016	0.016	0.010	-0.001	0.038	1.4	0.007	0.007	0.004	0.000	0.016	-3.7
t05_t2	0.022	0.022	0.013	-0.001	0.051	-1.0	0.010	0.010	0.006	0.000	0.021	-4.5
t05_t3	0.027	0.026	0.015	-0.001	0.058	-3.1	0.012	0.012	0.006	***0.000***	***0.025***	-5.5
t1_t1	0.022	0.022	0.012	***0.003***	***0.051***	2.4	0.010	0.010	0.005	***0.001***	***0.021***	-4.2
t1_t2	0.031	0.031	0.014	***0.005***	***0.063***	1.7	0.014	0.014	0.006	***0.002***	***0.026***	-4.5
t1_t3	0.037	0.038	0.017	***0.010***	***0.076***	0.8	0.017	0.016	0.007	***0.004***	***0.031***	-5.1
t2_t2	0.043	0.045	0.017	***0.017***	***0.081***	4.3	0.020	0.019	0.007	***0.007***	***0.034***	-3.1
t2_t3	0.052	0.054	0.018	***0.021***	***0.094***	4.5	0.024	0.024	0.008	***0.009***	***0.040***	-3.2
t3_t3	0.063	0.066	0.021	***0.029***	***0.112***	4.7	0.030	0.029	0.009	***0.013***	***0.047***	-3.7
t10_t10	0.180	0.198	0.038	***0.133***	***0.284***	10.2	0.100	0.096	0.016	***0.065***	***0.127***	-3.9
t10_t30	0.272	0.299	0.051	***0.210***	***0.408***	10.1	0.173	0.165	0.022	***0.125***	***0.208***	-4.7
t30_t30	0.386	0.469	0.058	***0.372***	***0.590***	21.5	0.300	0.288	0.026	***0.236***	***0.340***	-4.0

* Italic bold values are the CIs excluding 0. [a] --- indicates that the percent bias is not computed.

Table S 2 (Continued) Simulation distribution comparing the genetic correlation and covariance approaches under different environmental correlation levels

MAF=0.5, polygenic ρ_{g^*}=0.1 and $cov_{g^*} = 0.04$, $\boxed{\rho_e = 0.9}$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$, simulation=1000

	ρ_{g_diff}						cov_{g_diff}					
scenario	**True**	**Mean**	**Std**	**2.5th**	**97.5th**	**Bias(%)**	**True**	**Mean**	**Std**	**2.5th**	**97.5th**	**Bias(%)**
t0_t0	0.000	0.000	0.002	-0.003	0.006	---[a]	0.000	0.000	0.001	-0.001	0.002	---[a]
t0_t05	0.000	0.000	0.008	-0.012	0.013	---[a]	0.000	0.000	0.003	-0.005	0.005	---[a]
t0_t1	0.000	-0.001	0.008	-0.016	0.019	---[a]	0.000	0.000	0.003	-0.006	0.008	---[a]
t0_t2	0.000	-0.002	0.012	-0.024	0.025	---[a]	0.000	0.000	0.005	-0.009	0.010	---[a]
t0_t3	0.000	-0.003	0.014	-0.031	0.027	---[a]	0.000	-0.001	0.006	-0.011	0.011	---[a]
t0_t10	0.000	-0.010	0.026	-0.058	0.050	---[a]	0.000	-0.002	0.010	-0.022	0.020	---[a]
t0_t30	0.000	-0.026	0.045	-0.104	0.072	---[a]	0.000	-0.006	0.019	-0.041	0.032	---[a]
t05_t05	0.001	0.012	0.010	-0.004	0.034	4.5	0.005	0.005	0.004	-0.002	0.014	-2.5
t05_t1	0.002	0.016	0.011	-0.002	0.041	2.2	0.007	0.007	0.005	-0.001	0.017	-4.8
t05_t2	0.002	0.022	0.014	-0.003	0.054	1.2	0.010	0.010	0.006	-0.001	0.023	-3.9
t05_t3	0.003	0.027	0.017	-0.002	0.063	0.7	0.012	0.012	0.007	-0.001	0.027	-4.0
t1_t1	0.002	0.022	0.013	***0.001***	***0.051***	1.0	0.010	0.009	0.005	0.000	0.021	-6.7
t1_t2	0.003	0.031	0.016	***0.005***	***0.067***	2.2	0.014	0.013	0.006	***0.002***	***0.027***	-4.8
t1_t3	0.004	0.038	0.019	***0.006***	***0.084***	0.7	0.017	0.016	0.008	***0.002***	***0.033***	-6.5
t2_t2	0.005	0.046	0.019	***0.015***	***0.085***	6.5	0.020	0.020	0.008	***0.006***	***0.036***	-2.4
t2_t3	0.006	0.055	0.022	***0.017***	***0.101***	6.1	0.024	0.024	0.009	***0.007***	***0.042***	-3.5
t3_t3	0.007	0.066	0.024	***0.025***	***0.123***	5.0	0.030	0.029	0.010	***0.010***	***0.049***	-4.0
t10_t10	0.020	0.201	0.044	***0.131***	***0.310***	11.9	0.100	0.097	0.017	***0.065***	***0.131***	-3.5
t10_t30	0.030	0.304	0.059	***0.209***	***0.423***	11.8	0.173	0.165	0.024	***0.121***	***0.210***	-4.7
t30_t30	0.043	0.473	0.066	***0.356***	***0.618***	22.7	0.300	0.288	0.028	***0.235***	***0.343***	-4.1

* Italic bold values are the CIs excluding 0. [a] --- indicates that the percent bias is not computed.

Table S 3 Simulation distribution comparing the genetic correlation and covariance approaches under different MAFs

MAF=0.1, polygenic ρ_{g^*}=0.1 and $cov_{g^*} = 0.04$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$, simulation =1000

	ρ_{g_diff}						cov_{g_diff}					
scenario	**True**	**Est**	**SD**	**2.5th**	**97.5th**	**Bias(%)**	**True**	**Est**	**SD**	**2.5th**	**97.5th**	**Bias(%)**
t0_t0	0.000	0.000	0.002	-0.003	0.005	---[a]	0.000	0.000	0.001	-0.001	0.002	---[a]
t0_t05	0.000	0.000	0.006	-0.010	0.013	---[a]	0.000	0.000	0.002	-0.004	0.005	---[a]
t0_t1	0.000	-0.001	0.008	-0.015	0.017	---[a]	0.000	0.000	0.003	-0.006	0.007	---[a]
t0_t2	0.000	-0.002	0.011	-0.022	0.021	---[a]	0.000	0.000	0.004	-0.008	0.009	---[a]
t0_t3	0.000	-0.003	0.013	-0.027	0.025	---[a]	0.000	0.000	0.005	-0.010	0.011	---[a]
t0_t10	0.000	-0.010	0.025	-0.055	0.041	---[a]	0.000	-0.002	0.010	-0.020	0.019	---[a]
t0_t30	0.000	-0.028	0.043	-0.104	0.062	---[a]	0.000	-0.006	0.018	-0.041	0.029	---[a]
t05_t05	0.005	0.012	0.009	-0.003	0.033	4.8	0.005	0.005	0.004	-0.001	0.013	-0.3
t05_t1	0.007	0.016	0.011	-0.001	0.041	1.7	0.007	0.007	0.005	-0.001	0.017	-2.9
t05_t2	0.010	0.022	0.013	-0.001	0.052	-0.1	0.010	0.010	0.006	0.000	0.022	-3.7
t05_t3	0.012	0.026	0.016	-0.001	0.062	-1.6	0.012	0.012	0.007	0.000	0.026	-4.1
t1_t1	0.010	0.023	0.013	***0.002***	***0.050***	3.0	0.010	0.010	0.005	***0.001***	***0.021***	-3.4
t1_t2	0.014	0.032	0.015	***0.005***	***0.066***	2.9	0.014	0.014	0.006	***0.002***	***0.027***	-3.3
t1_t3	0.017	0.038	0.018	***0.007***	***0.079***	2.1	0.017	0.017	0.007	***0.003***	***0.033***	-3.7
t2_t2	0.019	0.044	0.017	***0.016***	***0.080***	3.7	0.020	0.019	0.007	***0.006***	***0.034***	-3.6
t2_t3	0.023	0.054	0.019	***0.022***	***0.094***	4.0	0.024	0.024	0.008	***0.009***	***0.040***	-3.6
t3_t3	0.028	0.066	0.021	***0.030***	***0.116***	4.8	0.030	0.029	0.009	***0.013***	***0.047***	-3.5
t10_t10	0.080	0.199	0.039	***0.132***	***0.283***	10.3	0.100	0.096	0.017	***0.065***	***0.131***	-3.7
t10_t30	0.121	0.300	0.054	***0.204***	***0.418***	10.3	0.173	0.165	0.024	***0.119***	***0.213***	-4.5
t30_t30	0.171	0.468	0.059	***0.367***	***0.588***	21.4	0.300	0.288	0.032	***0.230***	***0.350***	-3.9

* Italic bold values are the CIs excluding 0. [a] --- indicates that the percent bias is not computed.

Table S 4 Estimate, percent bias, empirical type I error and power comparing different polygenic and environmental correlations for a SNP Pleiotropic Effect in CovA

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

		$\rho_{g^*} = 0.0$, $\rho_e = 0.0$, $h^2_{r1^*} = h^2_{r2^*} = 0.01$			$\rho_{g^*} = 0.1$, $\rho_e = 0.0$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$			$\rho_{g^*} = 0.6$, $\rho_e = 0.0$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$			$\rho_{g^*} = 0.9$, $\rho_e = 0.0$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$		
Scenario	C_P	$\widehat{C}_P$	Bias(%)	α =.05	$\widehat{C}_P$	Bias(%)	α =.05	$\widehat{C}_P$	Bias(%)	α =.05	$\widehat{C}_P$	Bias(%)	α =.05
t0_t0	0.000	0.000	---[a]	0.002	0.000	---[a]	0.001	0.000	---[a]	0.002	0.000	---[a]	0.006
t0_t1	0.000	0.000	---[a]	0.041	0.000	---[a]	0.037	0.000	---[a]	0.037	0.000	---[a]	0.037
t0_t2	0.000	0.000	---[a]	0.047	0.000	---[a]	0.045	0.000	---[a]	0.045	0.000	---[a]	0.045
t0_t3	0.000	0.000	---[a]	0.047	0.000	---[a]	0.047	0.000	---[a]	0.047	0.001	---[a]	0.047
t0_t10	0.000	0.000	---[a]	0.047	0.000	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047
t0_t30	0.000	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047
t1_t1	0.010	0.010	-0.8	0.780	0.010	-0.1	0.782	0.010	1.0	0.787	0.010	1.7	0.790
t1_t2	0.014	0.014	-0.2	0.871	0.014	1.0	0.872	0.014	1.6	0.873	0.014	1.9	0.874
t1_t3	0.017	0.017	-0.3	0.876	0.017	0.8	0.876	0.017	0.8	0.876	0.017	0.8	0.876
t1_t10	0.032	0.031	-1.4	0.877	0.031	-0.8	0.878	0.031	-0.6	0.878	0.031	-0.5	0.878
t1_t30	0.055	0.055	-0.2	0.877	0.055	-0.2	0.878	0.055	-0.3	0.878	0.055	-0.5	0.878
t2_t2	0.020	0.020	1.2	0.982	0.020	1.2	0.985	0.020	1.8	0.986	0.020	2.2	0.986
t2_t3	0.024	0.025	0.3	0.986	0.025	0.4	0.990	0.025	0.9	0.993	0.025	1.1	0.996
t2_t10	0.045	0.045	-0.1	0.986	0.045	-0.5	0.991	0.045	-0.1	0.993	0.045	0.1	0.996
t2_t30	0.077	0.077	-1.1	0.986	0.077	0.0	0.991	0.078	0.1	0.993	0.078	0.2	0.996
t3_t3	0.030	0.030	0.2	1.000	0.030	0.8	0.999	0.030	1.0	1.000	0.030	0.9	1.000
t3_t10	0.055	0.055	-0.3	1.000	0.055	-0.1	1.000	0.055	0.0	1.000	0.055	0.1	1.000
t3_t30	0.095	0.095	0.1	1.000	0.095	0.4	1.000	0.095	0.3	1.000	0.095	0.2	1.000

*$\text{t}h^2_{q1}_\text{t}h^2_{q2}$ denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2.

[a] --- indicates that the percent bias is not computed.

Table S 4 (Continued) Estimate, percent bias, empirical type I error and power comparing different polygenic and environmental correlations for a SNP Pleiotropic Effect in CovA

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

Scenario	C_P	$\rho_{g^*} = 0.0, \rho_e = 0.6, h^2_{r1^*} = h^2_{r2^*} = 0.01$			$\rho_{g^*} = 0.1, \rho_e = 0.6, h^2_{r1^*} = h^2_{r2^*} = 0.4$			$\rho_{g^*} = 0.6, \rho_e = 0.6, h^2_{r1^*} = h^2_{r2^*} = 0.4$			$\rho_{g^*} = 0.9, \rho_e = 0.6, h^2_{r1^*} = h^2_{r2^*} = 0.4$		
		$\hat{C}_P$	Bias(%)	α =.05	$\hat{C}_P$	Bias(%)	α =.05	$\hat{C}_P$	Bias(%)	α =.05	$\hat{C}_P$	Bias(%)	α =.05
t0_t0	0.000	0.001	---[a]	0.011	0.000	---[a]	0.005	0.001	---[a]	0.013	0.001	---[a]	0.016
t0_t1	0.000	0.001	---[a]	0.037	0.000	---[a]	0.036	0.001	---[a]	0.033	0.001	---[a]	0.032
t0_t2	0.000	0.001	---[a]	0.046	0.001	---[a]	0.045	0.001	---[a]	0.043	0.001	---[a]	0.042
t0_t3	0.000	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047
t0_t10	0.000	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047
t0_t30	0.000	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047
t1_t1	0.010	0.011	6.6	0.795	0.011	5.6	0.796	0.011	5.8	0.814	0.011	6.5	0.817
t1_t2	0.014	0.015	5.3	0.874	0.015	4.4	0.878	0.015	5.1	0.883	0.015	5.3	0.885
t1_t3	0.017	0.018	4.3	0.877	0.018	3.8	0.879	0.018	4.1	0.884	0.018	4.1	0.886
t1_t10	0.032	0.032	1.7	0.878	0.032	1.3	0.879	0.032	1.5	0.884	0.032	1.6	0.887
t1_t30	0.055	0.055	0.9	0.878	0.055	1.1	0.879	0.055	1.0	0.884	0.055	0.8	0.887
t2_t2	0.020	0.021	4.4	0.986	0.021	3.2	0.986	0.021	3.8	0.986	0.021	4.1	0.989
t2_t3	0.024	0.025	3.3	0.990	0.025	2.4	0.992	0.025	2.9	0.993	0.025	3.0	0.996
t2_t10	0.045	0.045	1.6	0.990	0.045	0.7	0.992	0.045	1.0	0.993	0.045	1.3	0.996
t2_t30	0.077	0.078	1.1	0.990	0.078	0.6	0.992	0.078	0.7	0.993	0.078	0.8	0.996
t3_t3	0.030	0.031	2.8	0.998	0.031	2.6	1.000	0.031	2.7	1.000	0.031	2.6	1.000
t3_t10	0.055	0.055	1.3	1.000	0.055	0.9	1.000	0.055	1.1	1.000	0.055	1.1	1.000
t3_t30	0.095	0.096	1.0	1.000	0.096	1.0	1.000	0.096	0.9	1.000	0.096	0.8	1.000

*$\text{t}h^2_{q1}_th^2_{q2}$ denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2.

[a] --- indicates that the percent bias is not computed.

Table S 4 (Continued) Estimate, percent bias, empirical type I error and power comparing different polygenic and environmental correlations for a SNP Pleiotropic Effect in CovA

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

Scenario	C_P	$\boldsymbol{\rho_{g^*} = 0.0, \rho_e = 0.9, h^2_{r1^*} = h^2_{r2^*} = 0.01}$			$\boldsymbol{\rho_{g^*} = 0.1, \rho_e = 0.9, h^2_{r1^*} = h^2_{r2^*} = 0.4}$			$\boldsymbol{\rho_{g^*} = 0.6, \rho_e = 0.9, h^2_{r1^*} = h^2_{r2^*} = 0.4}$			$\boldsymbol{\rho_{g^*} = 0.9, \rho_e = 0.9, h^2_{r1^*} = h^2_{r2^*} = 0.4}$		
		$\hat{C}_P$	Bias(%)	α =.05	$\hat{C}_P$	Bias(%)	α =.05	$\hat{C}_P$	Bias(%)	α =.05	$\hat{C}_P$	Bias(%)	α =.05
t0_t0	0.000	0.001	---[a]	0.037	0.001	---[a]	0.011	0.001	---[a]	0.020	0.001	---[a]	0.038
t0_t1	0.000	0.001	---[a]	0.036	0.001	---[a]	0.036	0.001	---[a]	0.029	0.001	---[a]	0.029
t0_t2	0.000	0.001	---[a]	0.042	0.001	---[a]	0.042	0.001	---[a]	0.040	0.001	---[a]	0.038
t0_t3	0.000	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.046
t0_t10	0.000	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047
t0_t30	0.000	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047	0.001	---[a]	0.047
t1_t1	0.010	0.011	9.9	0.818	0.011	7.6	0.821	0.011	9.0	0.826	0.011	9.7	0.830
t1_t2	0.014	0.015	8.5	0.886	0.015	6.6	0.891	0.015	7.3	0.893	0.015	7.5	0.895
t1_t3	0.017	0.018	1.3	0.887	0.018	5.7	0.893	0.018	6.1	0.893	0.018	6.1	0.895
t1_t10	0.032	0.033	3.8	0.889	0.033	2.9	0.893	0.033	3.1	0.895	0.033	3.2	0.897
t1_t30	0.055	0.056	2.5	0.889	0.056	2.2	0.893	0.056	2.1	0.895	0.056	1.9	0.897
t2_t2	0.020	0.021	6.5	0.986	0.021	4.6	0.986	0.021	5.2	0.990	0.021	5.5	0.994
t2_t3	0.024	0.025	1.2	0.990	0.025	3.8	0.992	0.026	4.3	0.993	0.026	4.4	0.997
t2_t10	0.045	0.045	0.7	0.990	0.045	1.7	0.992	0.046	2.0	0.993	0.046	2.2	0.997
t2_t30	0.077	0.078	0.7	0.990	0.078	1.3	0.992	0.079	1.3	0.993	0.079	1.4	0.997
t3_t3	0.030	0.031	4.5	1.000	0.031	3.8	1.000	0.031	3.9	1.000	0.031	3.8	1.000
t3_t10	0.055	0.056	2.5	1.000	0.056	1.8	1.000	0.056	1.9	1.000	0.056	2.0	1.000
t3_t30	0.095	0.096	1.7	1.000	0.096	1.5	1.000	0.096	1.4	1.000	0.096	1.3	1.000

*th^2_{q1}_th^2_{q2} denotes SNP with h^2_{q1}% effect on trait1 and h^2_{q2}% effect on trait2.

[a] --- indicates that the percent bias is not computed.

Table S 5 Estimates, percent biased, empirical type I error and power comparing different MAF for a SNP Pleiotropic Effect in CovA

Polygenic ρ_{g^*}=0.6 and $cov_{g^*} = 0.24$, ρ_e= 0.6, $h^2_{r1^*} = h^2_{r2^*} = 0.4$

Simulation = 1000, bootstrapping in each simulation = 1000

Scenario	C_P	MAF=0.1			MAF=0.5		
		$\widehat{C}_P$	Bias(%)	α =.05	$\widehat{C}_P$	Bias(%)	α =.05
t0_t0	0.000	0.001	---[a]	0.014	0.001	---[a]	0.013
t0_t1	0.000	0.001	---[a]	0.035	0.001	---[a]	0.033
t0_t2	0.000	0.001	---[a]	0.047	0.001	---[a]	0.043
t0_t3	0.000	0.001	---[a]	0.048	0.001	---[a]	0.047
t0_t10	0.000	0.001	---[a]	0.048	0.001	---[a]	0.047
t0_t30	0.000	0.001	---[a]	0.048	0.001	---[a]	0.047
t1_t1	0.010	0.011	8.2	0.830	0.011	5.8	0.814
t1_t2	0.014	0.015	7.5	0.900	0.015	5.1	0.883
t1_t3	0.017	0.018	5.1	0.905	0.018	4.1	0.884
t1_t10	0.032	0.033	3.2	0.907	0.032	1.5	0.884
t1_t30	0.055	0.056	2.6	0.907	0.055	1.0	0.884
t2_t2	0.020	0.021	6.4	0.992	0.021	3.8	0.986
t2_t3	0.024	0.026	4.2	0.993	0.025	2.9	0.993
t2_t10	0.045	0.046	2.7	0.993	0.045	1.0	0.993
t2_t30	0.077	0.079	2.3	0.993	0.078	0.7	0.993
t3_t3	0.030	0.031	3.5	1.000	0.031	2.7	1.000
t3_t10	0.055	0.056	2.3	1.000	0.055	1.1	1.000
t3_t30	0.095	0.097	2.0	1.000	0.096	0.9	1.000

Table S 6 Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.0$, $\rho_e = 0.9$, $h^2_{r1^*} = h^2_{r2^*} = 0.01$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\widehat{C}_P$	$\alpha = .05$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\alpha = .05$	$\widehat{\beta}_{41} - \widehat{\beta}_{21}$	$\alpha = .05$	$\widehat{\beta}_{31} - \widehat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.001	0.037	0.002	0.001	0.061	0.002	0.044	0.001	0.059
t0_t1	0.000	0.001	0.036	0.002	0.143	***1.000***	0.002	0.044	0.128	***0.883***
t0_t2	0.000	0.001	0.042	0.002	0.202	***1.000***	0.002	0.044	0.182	***0.997***
t0_t3	0.000	0.001	0.047	0.002	0.247	***1.000***	0.002	0.044	0.223	***0.999***
t1_t1	0.010	0.011	0.818	0.143	0.143	0.850	0.127	0.887	0.127	0.883
t1_t2	0.014	0.015	0.886	0.143	0.202	0.998	0.126	0.887	0.181	0.997
t1_t3	0.017	0.018	0.887	0.143	0.247	1.000	0.126	0.887	0.222	0.999
t2_t2	0.020	0.021	0.986	0.202	0.202	0.991	0.180	0.994	0.180	0.997
t2_t3	0.024	0.025	0.990	0.202	0.247	1.000	0.179	0.994	0.221	0.999
t3_t3	0.030	0.031	1.000	0.247	0.247	0.999	0.220	0.999	0.220	0.999

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table S 6 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.1, \rho_e = 0.0, h^2_{r1^*} = h^2_{r2^*} = 0.4$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\hat{C}_P$	$\alpha = .05$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\alpha = .05$	$\hat{\beta}_{41} - \hat{\beta}_{21}$	$\alpha = .05$	$\hat{\beta}_{31} - \hat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.000	0.001	0.001	-0.002	0.064	0.000	0.010	0.000	0.008
t0_t1	0.000	0.000	0.037	0.001	0.141	***0.819***	0.000	0.008	0.006	***0.200***
t0_t2	0.000	0.000	0.045	0.001	0.202	***0.985***	0.000	0.005	0.008	***0.224***
t0_t3	0.000	0.000	0.047	0.001	0.247	***0.997***	0.000	0.006	0.010	***0.261***
t1_t1	0.010	0.010	0.782	0.141	0.141	0.982	0.006	***0.205***	0.006	***0.205***
t1_t2	0.014	0.014	0.872	0.141	0.202	1.000	0.006	***0.215***	0.008	***0.253***
t1_t3	0.017	0.017	0.876	0.141	0.247	0.999	0.006	***0.197***	0.010	***0.247***
t2_t2	0.020	0.020	0.985	0.199	0.202	1.000	0.008	***0.251***	0.008	***0.249***
t2_t3	0.024	0.025	0.990	0.199	0.247	1.000	0.008	***0.243***	0.010	***0.248***
t3_t3	0.030	0.030	0.999	0.245	0.247	1.000	0.010	***0.250***	0.010	***0.248***

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table S 6 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.1$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\hat{C}_P$	$\alpha = .05$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\alpha = .05$	$\hat{\beta}_{41} - \hat{\beta}_{21}$	$\alpha = .05$	$\hat{\beta}_{31} - \hat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.000	0.005	0.001	-0.001	0.064	0.001	0.047	0.000	0.060
t0_t1	0.000	0.000	0.036	0.001	0.142	***0.880***	0.001	0.046	0.057	***0.883***
t0_t2	0.000	0.001	0.045	0.001	0.202	***0.997***	0.001	0.046	0.082	***0.998***
t0_t3	0.000	0.001	0.047	0.001	0.247	***0.998***	0.001	0.047	0.100	***0.998***
t1_t1	0.010	0.011	0.796	0.142	0.142	0.936	0.057	0.888	0.057	0.883
t1_t2	0.014	0.015	0.878	0.142	0.202	0.992	0.056	0.888	0.082	0.998
t1_t3	0.017	0.018	0.879	0.142	0.247	0.997	0.056	0.888	0.100	0.998
t2_t2	0.020	0.021	0.986	0.200	0.202	0.999	0.080	0.989	0.081	0.998
t2_t3	0.024	0.025	0.992	0.200	0.247	1.000	0.079	0.989	0.099	0.998
t3_t3	0.030	0.031	1.000	0.246	0.247	1.000	0.098	1.000	0.099	0.998

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table S 6 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.1$, $\rho_e = 0.9$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\hat{C}_P$	$\alpha = .05$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\alpha = .05$	$\hat{\beta}_{41} - \hat{\beta}_{21}$	$\alpha = .05$	$\hat{\beta}_{31} - \hat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.001	0.011	0.001	0.001	0.058	-0.001	0.047	0.000	0.068
t0_t1	0.000	0.001	0.036	0.001	0.142	***0.944***	0.000	0.047	0.083	***0.887***
t0_t2	0.000	0.001	0.042	0.001	0.203	***1.000***	-0.001	0.047	0.119	***0.992***
t0_t3	0.000	0.001	0.047	0.001	0.248	***1.000***	0.000	0.047	0.146	***1.000***
t1_t1	0.010	0.011	0.821	0.143	0.142	0.904	0.083	0.896	0.083	0.887
t1_t2	0.014	0.015	0.891	0.143	0.203	0.988	0.083	0.896	0.118	0.992
t1_t3	0.017	0.018	0.893	0.143	0.248	0.998	0.082	0.896	0.145	1.000
t2_t2	0.020	0.021	0.986	0.200	0.203	0.996	0.116	0.991	0.118	0.992
t2_t3	0.024	0.025	0.992	0.200	0.248	0.999	0.115	0.991	0.144	1.000
t3_t3	0.030	0.031	1.000	0.246	0.248	1.000	0.143	1.000	0.144	1.000

**th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table S 6 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.6$, $\rho_e = 0.9$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$										
Scenario	C_P	CovA		MANOVA			Conditional Analysis			
		$\hat{C}_P$	$\alpha = .05$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\alpha = .05$	$\hat{\beta}_{41} - \hat{\beta}_{21}$	$\alpha = .05$	$\hat{\beta}_{31} - \hat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.001	0.020	0.001	0.000	0.055	0.001	0.050	0.000	0.054
t0_t1	0.000	0.001	0.029	0.001	0.142	***1.000***	0.001	0.050	0.112	***0.887***
t0_t2	0.000	0.001	0.040	0.001	0.202	***1.000***	0.001	0.050	0.159	***0.995***
t0_t3	0.000	0.001	0.047	0.001	0.247	***1.000***	0.001	0.050	0.196	***1.000***
t1_t1	0.010	0.011	0.826	0.142	0.142	0.858	0.111	0.895	0.111	0.887
t1_t2	0.014	0.015	0.893	0.142	0.202	0.987	0.110	0.895	0.158	0.995
t1_t3	0.017	0.018	0.893	0.142	0.247	0.999	0.110	0.895	0.195	1.000
t2_t2	0.020	0.021	0.990	0.200	0.202	0.992	0.156	0.993	0.158	0.995
t2_t3	0.024	0.026	0.993	0.200	0.247	0.999	0.155	0.993	0.194	1.000
t3_t3	0.030	0.031	1.000	0.246	0.247	1.000	0.192	1.000	0.193	1.000

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table S 6 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.9$, $\rho_e = 0.0$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\widehat{C}_P$	$\alpha = .05$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\alpha = .05$	$\widehat{\beta}_{41} - \widehat{\beta}_{21}$	$\alpha = .05$	$\widehat{\beta}_{31} - \widehat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.000	0.006	0.001	-0.002	0.057	0.001	0.045	-0.001	0.065
t0_t1	0.000	0.000	0.037	0.001	0.141	***0.876***	0.001	0.045	0.051	***0.909***
t0_t2	0.000	0.000	0.045	0.001	0.201	***0.996***	0.001	0.045	0.073	***0.991***
t0_t3	0.000	0.001	0.047	0.001	0.245	***1.000***	0.000	0.045	0.090	***1.000***
t1_t1	0.010	0.010	0.790	0.140	0.141	0.942	0.050	0.870	0.051	0.909
t1_t2	0.014	0.014	0.874	0.140	0.201	0.992	0.050	0.870	0.073	0.991
t1_t3	0.017	0.017	0.876	0.140	0.245	0.999	0.050	0.870	0.089	1.000
t2_t2	0.020	0.020	0.986	0.199	0.201	1.000	0.072	0.994	0.072	0.991
t2_t3	0.024	0.025	0.996	0.199	0.245	1.000	0.071	0.994	0.089	1.000
t3_t3	0.030	0.030	1.000	0.245	0.245	1.000	0.088	1.000	0.088	1.000

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table S 6 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.9$, $\rho_e = 0.6$, $h^2_{r1^*} = h^2_{r2^*} = 0.4$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\hat{C}_P$	$\alpha = .05$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\alpha = .05$	$\hat{\beta}_{41} - \hat{\beta}_{21}$	$\alpha = .05$	$\hat{\beta}_{31} - \hat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.001	0.016	0.001	-0.001	0.059	0.001	0.045	0.000	0.059
t0_t1	0.000	0.001	0.032	0.001	0.141	***0.991***	0.001	0.045	0.102	***0.893***
t0_t2	0.000	0.001	0.042	0.001	0.201	***1.000***	0.001	0.045	0.146	***0.994***
t0_t3	0.000	0.001	0.047	0.001	0.246	***1.000***	0.001	0.045	0.180	***1.000***
t1_t1	0.010	0.011	0.817	0.141	0.141	0.878	0.101	0.879	0.102	0.893
t1_t2	0.014	0.015	0.885	0.141	0.201	0.987	0.101	0.879	0.146	0.994
t1_t3	0.017	0.018	0.886	0.141	0.246	0.999	0.100	0.879	0.179	1.000
t2_t2	0.020	0.021	0.989	0.200	0.201	0.995	0.144	0.994	0.145	0.994
t2_t3	0.024	0.025	0.996	0.200	0.246	0.999	0.143	0.994	0.178	1.000
t3_t3	0.030	0.031	1.000	0.245	0.246	1.000	0.177	1.000	0.177	1.000

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table S 6 (Continued) Estimate, type I error and power of a pleiotropic SNP Effect comparing CovA, MANOVA, and conditional analysis using unrelated data

MAF=0.5, simulation = 1000, bootstrapping in each simulation = 1000

$\rho_{g^*} = 0.9, \rho_e = 0.9, h^2_{r1^*} = h^2_{r2^*} = 0.4$										
Scenario	C_P	**CovA**		**MANOVA**			**Conditional Analysis**			
		$\hat{C}_P$	$\alpha = .05$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\alpha = .05$	$\hat{\beta}_{41} - \hat{\beta}_{21}$	$\alpha = .05$	$\hat{\beta}_{31} - \hat{\beta}_{11}$	$\alpha = .05$
t0_t0	0.000	0.001	0.038	0.001	0.000	0.062	0.001	0.047	0.000	0.051
t0_t1	0.000	0.001	0.029	0.001	0.142	***1.000***	0.002	0.047	0.129	***0.899***
t0_t2	0.000	0.001	0.038	0.001	0.202	***1.000***	0.001	0.047	0.183	***0.996***
t0_t3	0.000	0.001	0.046	0.001	0.246	***1.000***	0.001	0.047	0.225	***0.999***
t1_t1	0.010	0.011	0.830	0.142	0.142	0.849	0.128	0.888	0.128	0.899
t1_t2	0.014	0.015	0.895	0.142	0.202	0.984	0.127	0.888	0.182	0.996
t1_t3	0.017	0.018	0.895	0.142	0.246	0.999	0.127	0.888	0.224	0.999
t2_t2	0.020	0.021	0.994	0.201	0.202	0.992	0.181	0.996	0.181	0.996
t2_t3	0.024	0.026	0.997	0.201	0.246	0.999	0.180	0.996	0.223	0.999
t3_t3	0.030	0.031	1.000	0.246	0.246	1.000	0.221	1.000	0.222	0.999

*th^2_{q1}_th^2_{q2} denotes SNP with $h^2_{q1}\%$ effect on trait1 and $h^2_{q2}\%$ effect on trait2. Italic bold values indicate abnormal type I error rates or powers.

Table S 7 SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

C_P

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{C}_P$	2.5th	97.5th
rs6733708	2	37814216	0.36	-0.129	-0.086	0.008	0.003	0.005	***0.001***	***0.011***
rs698798*	2	44544018	0.08	-0.198	-0.154	0.006	0.004	0.005	***0.001***	***0.009***
rs698819*	2	44566433	0.08	-0.202	-0.223	0.006	0.007	0.006	***0.002***	***0.013***
rs3769292	2	173434505	0.10	-0.171	-0.145	0.005	0.004	0.004	***0.001***	***0.009***
rs1569159	2	181189870	0.09	0.151	0.195	0.004	0.007	0.005	***0.001***	***0.011***
rs16864755	2	223982866	0.10	0.258	0.210	0.013	0.009	0.010	***0.004***	***0.018***
rs457414	3	10177884	0.32	-0.103	-0.128	0.005	0.008	0.006	***0.002***	***0.011***
rs9846680	3	179603561	0.08	-0.244	-0.254	0.009	0.010	0.009	***0.004***	***0.016***
rs1545026	3	196061661	0.13	0.180	0.125	0.007	0.004	0.005	***0.001***	***0.011***
rs13148678*	4	74077760	0.17	0.170	0.186	0.008	0.010	0.009	***0.003***	***0.015***
rs7664273*	4	74226102	0.17	0.172	0.181	0.008	0.009	0.008	***0.003***	***0.016***
rs13119179*	4	74262694	0.17	0.168	0.181	0.008	0.009	0.008	***0.003***	***0.015***

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\hat{C}_P$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table S 7 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

C_P

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{C}_P$	2.5th	97.5th
rs283062	6	118728149	0.43	0.134	0.128	0.009	0.009	0.009	***0.004***	***0.015***
rs665506	7	132023073	0.35	0.146	0.137	0.010	0.009	0.009	***0.003***	***0.017***
rs1905045	8	75062568	0.21	0.138	0.162	0.007	0.010	0.008	***0.003***	***0.014***
rs2317356*	8	137106399	0.32	0.116	0.110	0.006	0.005	0.005	***0.001***	***0.012***
rs2317355*	8	137106698	0.32	0.117	0.110	0.006	0.005	0.005	***0.001***	***0.012***
rs1031282*	8	137123839	0.32	0.119	0.113	0.006	0.005	0.005	***0.001***	***0.012***
rs1332199*	9	9493477	0.17	0.170	0.148	0.008	0.007	0.007	***0.002***	***0.013***
rs639168*	9	9529574	0.12	0.177	0.187	0.007	0.008	0.007	***0.002***	***0.014***
rs668026*	9	9571692	0.31	0.129	0.150	0.008	0.011	0.009	***0.004***	***0.016***
rs681437*	9	9572128	0.31	0.133	0.153	0.008	0.011	0.009	***0.004***	***0.016***
rs598768*	9	9579682	0.35	0.137	0.142	0.009	0.010	0.009	***0.004***	***0.017***
rs657849*	9	9583722	0.36	0.153	0.143	0.011	0.010	0.010	***0.005***	***0.018***

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\hat{C}_P$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table S 7 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

C_P

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{C}_P$	2.5th	97.5th
rs1889524*	10	7657473	0.46	-0.133	-0.134	0.008	0.009	0.008	***0.003***	***0.016***
rs1537631*	10	7657765	0.45	-0.128	-0.125	0.008	0.008	0.008	***0.003***	***0.015***
rs2275069*	10	7658692	0.46	-0.134	-0.140	0.009	0.010	0.009	***0.003***	***0.017***
rs384626	10	60806221	0.49	-0.091	-0.089	0.004	0.004	0.004	***0.001***	***0.009***
rs7911563	10	60878297	0.41	-0.098	-0.101	0.005	0.005	0.005	***0.001***	***0.010***
rs453061	11	8227452	0.01	0.592	0.415	0.009	0.004	0.006	***0.001***	***0.012***
rs10766761	11	20972609	0.28	-0.154	-0.131	0.010	0.007	0.008	***0.003***	***0.015***
rs1968699	15	97529279	0.06	-0.195	-0.260	0.004	0.008	0.006	***0.002***	***0.012***
rs2216263	16	47911005	0.39	0.105	0.104	0.005	0.005	0.005	***0.001***	***0.011***
rs1389529	16	53286410	0.07	0.213	0.207	0.006	0.006	0.005	***0.001***	***0.012***
rs1699607	17	70130820	0.49	-0.121	-0.100	0.008	0.005	0.006	***0.001***	***0.012***
rs892583	18	43170372	0.27	0.139	0.101	0.007	0.004	0.005	***0.001***	***0.011***
rs740586	19	48888371	0.22	0.104	0.119	0.004	0.005	0.004	***0.001***	***0.010***
rs346062	19	48890417	0.43	-0.116	-0.111	0.007	0.007	0.007	***0.002***	***0.014***

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\hat{C}_P$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table S 7 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

cov_{g_diff}

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\widehat{cov}_{g_diff}$	2.5th	97.5th
rs6733708	2	37814216	0.36	-0.129	-0.086	0.008	0.003	0.006	-0.005	0.020
rs698798*	2	44544018	0.08	-0.198	-0.154	0.006	0.004	0.020	***0.007***	***0.036***
rs698819*	2	44566433	0.08	-0.202	-0.223	0.006	0.007	0.017	***0.002***	***0.036***
rs3769292	2	173434505	0.10	-0.171	-0.145	0.005	0.004	0.000	-0.010	0.010
rs1569159	2	181189870	0.09	0.151	0.195	0.004	0.007	-0.009	-0.021	0.005
rs16864755	2	223982866	0.10	0.258	0.210	0.013	0.009	0.013	-0.002	0.032
rs457414	3	10177884	0.32	-0.103	-0.128	0.005	0.008	0.005	-0.007	0.018
rs9846680	3	179603561	0.08	-0.244	-0.254	0.009	0.010	0.013	-0.003	0.033
rs1545026	3	196061661	0.13	0.180	0.125	0.007	0.004	-0.003	-0.014	0.011
rs13148678*	4	74077760	0.17	0.170	0.186	0.008	0.010	0.001	-0.014	0.019
rs7664273*	4	74226102	0.17	0.172	0.181	0.008	0.009	0.002	-0.013	0.019
rs13119179*	4	74262694	0.17	0.168	0.181	0.008	0.009	0.001	-0.013	0.017

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\widehat{cov}_{g_diff}$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table S 7 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

$\boldsymbol{cov_{g_diff}}$

SNP	Chr.	Position	MAF	$\widehat{\beta}$ (FNBMD)	$\widehat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\widehat{cov}_{g_diff}$	2.5th	97.5th
rs283062	6	118728149	0.43	0.134	0.128	0.009	0.009	-0.009	-0.023	0.006
rs665506	7	132023073	0.35	0.146	0.137	0.010	0.009	0.016	***0.001***	***0.035***
rs1905045	8	75062568	0.21	0.138	0.162	0.007	0.010	-0.015	-0.031	0.003
rs2317356*	8	137106399	0.32	0.116	0.110	0.006	0.005	0.015	***0.003***	***0.031***
rs2317355*	8	137106698	0.32	0.117	0.110	0.006	0.005	0.016	***0.003***	***0.032***
rs1031282*	8	137123839	0.32	0.119	0.113	0.006	0.005	0.019	***0.005***	***0.037***
rs1332199*	9	9493477	0.17	0.170	0.148	0.008	0.007	0.026	***0.009***	***0.045***
rs639168*	9	9529574	0.12	0.177	0.187	0.007	0.008	0.017	***0.002***	***0.035***
rs668026*	9	9571692	0.31	0.129	0.150	0.008	0.011	0.001	-0.014	0.017
rs681437*	9	9572128	0.31	0.133	0.153	0.008	0.011	0.002	-0.013	0.017
rs598768*	9	9579682	0.35	0.137	0.142	0.009	0.010	0.006	-0.010	0.020
rs657849*	9	9583722	0.36	0.153	0.143	0.011	0.010	0.004	-0.013	0.019

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\widehat{cov}_{g_diff}$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table S 7 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

cov_{g_diff}

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\widehat{cov}_{g_diff}$	2.5th	97.5th
rs1889524*	10	7657473	0.46	-0.133	-0.134	0.008	0.009	0.011	-0.004	0.029
rs1537631*	10	7657765	0.45	-0.128	-0.125	0.008	0.008	0.008	-0.005	0.025
rs2275069*	10	7658692	0.46	-0.134	-0.140	0.009	0.010	0.012	-0.003	0.032
rs384626	10	60806221	0.49	-0.091	-0.089	0.004	0.004	0.013	***0.002***	***0.025***
rs7911563	10	60878297	0.41	-0.098	-0.101	0.005	0.005	0.016	***0.001***	***0.032***
rs453061	11	8227452	0.01	0.592	0.415	0.009	0.004	0.010	***0.002***	***0.019***
rs10766761	11	20972609	0.28	-0.154	-0.131	0.010	0.007	-0.016	***-0.031***	***-3E-04***
rs1968699	15	97529279	0.06	-0.195	-0.260	0.004	0.008	0.012	-0.001	0.027
rs2216263	16	47911005	0.39	0.105	0.104	0.005	0.005	0.000	-0.012	0.013
rs1389529	16	53286410	0.07	0.213	0.207	0.006	0.006	0.005	-0.013	0.022
rs1699607	17	70130820	0.49	-0.121	-0.100	0.008	0.005	0.003	-0.009	0.020
rs892583	18	43170372	0.27	0.139	0.101	0.007	0.004	-0.015	***-0.028***	***-0.004***
rs740586	19	48888371	0.22	0.104	0.119	0.004	0.005	0.011	***4E-04***	***0.023***
rs346062	19	48890417	0.43	-0.116	-0.111	0.007	0.007	0.027	***0.010***	***0.049***

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\widehat{cov}_{g_diff}$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table S 7 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

$\boldsymbol{\rho_{g_diff}}$

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{\rho}_{g_diff}$	2.5th	97.5th
rs6733708	2	37814216	0.36	-0.129	-0.086	0.008	0.003	0.004	-0.006	0.017
rs698798*	2	44544018	0.08	-0.198	-0.154	0.006	0.004	0.017	***0.004***	***0.038***
rs698819*	2	44566433	0.08	-0.202	-0.223	0.006	0.007	0.010	-0.006	0.032
rs3769292	2	173434505	0.10	-0.171	-0.145	0.005	0.004	-0.001	-0.014	0.009
rs1569159	2	181189870	0.09	0.151	0.195	0.004	0.007	-0.004	-0.015	0.005
rs16864755	2	223982866	0.10	0.258	0.210	0.013	0.009	0.009	-0.005	0.029
rs457414	3	10177884	0.32	-0.103	-0.128	0.005	0.008	0.006	-0.003	0.021
rs9846680	3	179603561	0.08	-0.244	-0.254	0.009	0.010	0.010	-0.002	0.032
rs1545026	3	196061661	0.13	0.180	0.125	0.007	0.004	-0.004	-0.015	0.006
rs13148678*	4	74077760	0.17	0.170	0.186	0.008	0.010	0.004	-0.012	0.018
rs7664273*	4	74226102	0.17	0.172	0.181	0.008	0.009	0.003	-0.009	0.017
rs13119179*	4	74262694	0.17	0.168	0.181	0.008	0.009	0.003	-0.009	0.017

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\hat{\rho}_{g_diff}$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table S 7 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

$\boldsymbol{\rho_{g_diff}}$

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{\rho}_{g_diff}$	2.5th	97.5th
rs283062	6	118728149	0.43	0.134	0.128	0.009	0.009	-0.004	-0.016	0.008
rs665506	7	132023073	0.35	0.146	0.137	0.010	0.009	0.012	***5E-04***	***0.031***
rs1905045	8	75062568	0.21	0.138	0.162	0.007	0.010	-0.007	-0.020	0.003
rs2317356*	8	137106399	0.32	0.116	0.110	0.006	0.005	0.011	***1E-04***	***0.027***
rs2317355*	8	137106698	0.32	0.117	0.110	0.006	0.005	0.011	***4E-04***	***0.027***
rs1031282*	8	137123839	0.32	0.119	0.113	0.006	0.005	0.014	***0.003***	***0.033***
rs1332199*	9	9493477	0.17	0.170	0.148	0.008	0.007	0.019	***0.005***	***0.042***
rs639168*	9	9529574	0.12	0.177	0.187	0.007	0.008	0.012	-0.004	0.029
rs668026*	9	9571692	0.31	0.129	0.150	0.008	0.011	0.003	-0.008	0.016
rs681437*	9	9572128	0.31	0.133	0.153	0.008	0.011	0.004	-0.007	0.016
rs598768*	9	9579682	0.35	0.137	0.142	0.009	0.010	0.006	-0.006	0.019
rs657849*	9	9583722	0.36	0.153	0.143	0.011	0.010	0.005	-0.009	0.018

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\hat{\rho}_{g_diff}$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

Table S 7 (Continued) SNP-specific pleiotropic effect analysis on FNBMD and LSBMD using CovA, genetic covariance and genetic correlation approaches for family related subjects in Framingham Osteoporosis-BMDs study

$\boldsymbol{\rho_{g_diff}}$

SNP	Chr.	Position	MAF	$\hat{\beta}$ (FNBMD)	$\hat{\beta}$ (LSBMD)	R^2 (FNBMD)	R^2 (LSBMD)	$\hat{\rho}_{g_diff}$	2.5th	97.5th
rs1889524*	10	7657473	0.46	-0.133	-0.134	0.008	0.009	0.009	-0.003	0.029
rs1537631*	10	7657765	0.45	-0.128	-0.125	0.008	0.008	0.007	-0.004	0.023
rs2275069*	10	7658692	0.46	-0.134	-0.140	0.009	0.010	0.010	-0.002	0.032
rs384626	10	60806221	0.49	-0.091	-0.089	0.004	0.004	0.010	-0.003	0.023
rs7911563	10	60878297	0.41	-0.098	-0.101	0.005	0.005	0.011	-3E-04	0.029
rs453061	11	8227452	0.01	0.592	0.415	0.009	0.004	0.008	***0.002***	***0.019***
rs10766761	11	20972609	0.28	-0.154	-0.131	0.010	0.007	-0.008	-0.022	0.003
rs1968699	15	97529279	0.06	-0.195	-0.260	0.004	0.008	0.012	***0.001***	***0.027***
rs2216263	16	47911005	0.39	0.105	0.104	0.005	0.005	0.001	-0.008	0.010
rs1389529	16	53286410	0.07	0.213	0.207	0.006	0.006	0.004	-0.007	0.022
rs1699607	17	70130820	0.49	-0.121	-0.100	0.008	0.005	0.002	-0.008	0.018
rs892583	18	43170372	0.27	0.139	0.101	0.007	0.004	-0.008	-0.019	0.002
rs740586	19	48888371	0.22	0.104	0.119	0.004	0.005	0.008	-0.004	0.019
rs346062	19	48890417	0.43	-0.116	-0.111	0.007	0.007	0.019	***0.003***	***0.047***

Note: $\hat{\beta}$ is the estimated regression coefficient; R^2 is the coefficient of determination; $\hat{\rho}_{g_diff}$ is the estimated pleiotropic effect in CovA; 2.5th and 97.5th display the 95% confidence interval. Italic bold values are the CIs excluding 0. $p < .001$ on both FN and LS BMD.

*Neighboring SNPs in linkage disequilibrium ($r^2 \geq 0.5$).

REFERENCES

Allison, D. B., & Beasley, M. (1998). Method and computer program for controlling the family-wise alpha rate in gene association studies involving multiple phenotypes. *Genetic epidemiology*, *15*(1), 87-101.

Allison, D. B., Neale, M. C., Zannolli, R., Schork, N. J., Amos, C. I., & Blangero, J. (1999). Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci–mapping procedure. *The American Journal of Human Genetics*, *65*(2), 531-544.

Almasy, L., Dyer, T. D., & Blangero, J. (1997). Bivariate quantitative trait linkage analysis: Pleiotropy versus co-incident linkages. *Genetic epidemiology*, *14*(6), 953-958.

Almasy, L., & Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *American journal of human genetics*, *62*(5), 1198.

Amos, C. I. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *American journal of human genetics*, *54*(3), 535.

Atchley, W. R., & Zhu, J. (1997). Developmental quantitative genetics, conditional epigenetic variability and growth in mice. *Genetics*, *147*(2), 765-776.

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational research methods*, *7*(2), 127-150.

Bauman, L. E., Almasy, L., Blangero, J., Duggirala, R., Sinsheimer, J. S., & Lange, K. (2005). Fishing for pleiotropic QTLs in a polygenic sea. *Annals of human genetics*, *69*(5), 590-611.

Blake, G. M., & Fogelman, I. (2007). Role of dual-energy X-ray absorptiometry in the diagnosis and treatment of osteoporosis. *Journal of Clinical Densitometry*, *10*(1), 102-110.

Blangero, J. (1995). Multivariate oligogenic linkage analysis of quantitative traits in general pedigrees. *Am J Hum Genet*, *57*, A11.

Blangero, J. (2005). Genetic analysis of a common oligogenic trait with quantitative correlates: summary of GAW9 results. *Genetic epidemiology*, *12*(6), 689-706.

Boehnke, M., Moll, P. P., Lange, K., Weidman, W. H., & Kottke, B. A. (1986). Univariate and bivariate analyses of cholesterol and triglyceride levels in pedigrees. *American journal of medical genetics*, *23*(3), 775-792.

Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144-152.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in medicine*, *25*(24), 4279-4292.

Carey G. (1986). A general multivariate approach to linear modeling in human genetics. *Am.J.Hum.Genet*., *39*, 775-786.

Carey, G. (1998). Multivariate analysis of variance (MANOVA): I. *Theory. Retrieved March*, *6*, 2007.

Casella G., &Berger R.L. (2001). *Statistical Inference*, Duxbury Press, Second Edition, p.374.

Cheverud, J. M. (1988). A comparison of genetic and phenotypic correlations. *Evolution*, 958-968.

Cordell, H. J., & Clayton, D. G. (2005). Genetic association studies. *The Lancet*, *366*(9491), 1121-1131.

Cusano, N. E., Kiel, D. P., Demissie, S., Karasik, D., Cupples, L.A., Corella, D., Gao Q. Richardson K., Yiannakouris N., & Ordovas, J. M. (2012). A Polymorphism in a Gene Encoding Perilipin 4 Is Associated with Height but not with Bone Measures in Individuals from the Framingham Osteoporosis Study. *Calcified tissue international*, 1-12.

Davies, G.E., Soundy, T.J. (2009). The genetics of smoking and nicotine addiction. *S D Med Spec No*: 43–49.

Davignon, J. (2004). Beneficial cardiovascular pleiotropic effects of statins. *Circulation*, *109*(23 suppl 1), III-39.

Dawber, T. R., Meadors, G. F., & Moore Jr, F. E. (1951). Epidemiological Approaches to Heart Disease: The Framingham Study. *American Journal of Public Health and the Nations Health*, *41*(3), 279-286.

Dawber, T. R., Kannel, W. B., & Lyell, L. P. (2006). An approach to longitudinal studies in a community: the Framingham Study. *Annals of the New York Academy of sciences*, *107*(2), 539-556.

Demissie, S., Cupples, L. A., Myers, R., Aliabadi, P., Levy, D., & Felson, D. T. (2002). Genome scan for quantity of hand osteoarthritis: the Framingham Study. *Arthritis &*

Rheumatism, *46*(4), 946-952.

Demissie, S., Dupuis, J., Cupples, L. A., Beck, T. J., Kiel, D. P., & Karasik, D. (2007). Proximal hip geometry is linked to several chromosomal regions: genome-wide linkage results from the Framingham Osteoporosis Study. *Bone*, *40*(3), 743-750.

Deng, H. W., Stegman, M. R., Davies, K. M., Conway, T., & Recker, R. R. (1999). Genetic determination of variation and covariation of peak bone mass at the hip and spine. *Journal of Clinical Densitometry*, *2*(3), 251-263.

Deng, H. W., Livshits, G., Yakovenko, K., Xu, F. H., Conway, T., Davies, K. M., ... & Recker, R. R. (2002). Evidence for a major gene for bone mineral density/content in human pedigrees identified via probands with extreme bone mineral density. *Annals of human genetics*, *66*(1), 61-74.

Dodge, Y., Cox, D., Commenges, D., Davison, A., Solomon, P., & Wilson, S. (Eds.). (2006). *The Oxford dictionary of statistical terms*. Oxford University Press, USA.

Eisman, J. A. (1999). Genetics of osteoporosis. *Endocr Rev*, *20*(6), 788-804.

Falconer, D. S. (1981). *Introduction to quantitative genetics* (No. Ed. 2). Longman..

Falconer, D.S., & Mackay, T.F.C. (1996) *Introduction to quantitative genetics* (No. Ed. 4., Longman, Harlow, UK.

Feingold, E. (2001). Methods for linkage analysis of quantitative trait loci in humans. *Theoretical Population Biology*, *60*(3), 167-180.

Feinleib, M., Kannel, W. B., Garrison, R. J., McNamara, P. M., & Castelli, W. P. (1975). The Framingham offspring study. Design and preliminary data. *Preventive medicine*, *4*(4), 518-525.
.
Flint J., & Mackay T.F.C. (2009). Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res.*, *19*,723-733.

French, A., Poulsen, J., & Yu, A. (2002). Multivariate analysis of variance (MANOVA). *Retrieved March*, *31*, 2005.

Gardner, K. M., & Latta, R. G. (2007). Shared quantitative trait loci underlying the genetic correlation between continuous traits. *Molecular ecology*, *16*(20), 4195-4209.

Gerbens, F., Verburg, F. J., Van Moerkerk, H. T., Engel, B., Buist, W., Veerkamp, J. H., & Te Pas, M. F. (2001). Associations of heart and adipocyte fatty acid-binding protein gene expression with intramuscular fat content in pigs. *Journal of Animal science*, *79*(2), 347-354.

Gilbert, H., & Le Roy, P. (2003). Comparison of three multitrait methods for QTL detection. *Genetics Selection Evolution*, *35*(3), 281-304.

Gilbert, H., & Le Roy, P. (2004). Power of three multitrait methods for QTL detection in crossbred populations. *Genetics Selection Evolution*, *36*(3), 347-361.

Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis for the analysis of data. *Biometrika*, *55*(1), 1-17.

Goodman A.L. (1960). On the exact variance of product. *Journal of the American Statistical Association*, *55*(292), 708-713.

Grignola, F.E., Hoeschele, I., & Tier, B. (1996). Mapping quantitative trait loci via residual maximum likelihood: I. Methodology. *Genet. Sel. Evol. 28*, 479-490.

Grignola, F.E., Hoeschele, I. Zhang, Q., & Tier, B. (1996). Mapping quantitative trait loci via residual maximum likelihood: II. A simulation study. *Genet. Sel. Evol. 28*, 491-504.

Grindflek, E., Szyda, J., Liu, Z., & Lien, S. (2001). Detection of quantitative trait loci for meat quality in a commercial slaughter pig cross. *Mammalian Genome*, *12*(4), 299-304.

Gupta, M., Cheung, C. L., Hsu, Y. H., Demissie, S., Cupples, L. A., Kiel, D. P., & Karasik, D. (2011). Identification of homogeneous genetic architecture of multiple genetically correlated traits by block clustering of genome-wide associations. *Journal of Bone and Mineral Research*, *26*(6), 1261-1271.

Haines, J. L., & Pericak-Vance, M. A. (Eds.). (1998). *Approaches to gene mapping in complex human diseases*. Wiley-Liss.

Hanis, C. L., Sing, C. F., Clarke, W. R., & Schrott, H. G. (1983). Multivariate models for human genetic analysis: aggregation, coaggregation, and tracking of systolic blood pressure and weight. *American journal of human genetics*, *35*(6), 1196-1210.

Hasstedt, S. J., & Thomas, A. (2011). Detecting pleiotropy and epistasis using variance components linkage analysis in jPAP. *Human heredity*, *72*(4), 258-263.

Hesterberg, T., Moore, D. S., Monaghan, S., Clipson, A., & Epstein, R. (2005). Bootstrap methods and permutation tests. *Introduction to the Practice of Statistics*, *5*, 1-70.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, *24*(6), 417.

Hsu, Y. H., Zillikens, M. C., Wilson, S. G., Farber, C. R., Demissie, S., Soranzo, N., ... & Kiel, D. P. (2010). An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility Loci for osteoporosis-related traits. *PLoS genetics*, *6*(6), e1000977.

Jiang, C., & Zeng, Z. B. (1995). Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics*, *140*(3), 1111-1127.

Kanis, J. A. (2002). Diagnosis of osteoporosis and assessment of fracture risk. *Lancet (London, England)*, *359*(9321), 1929-1936.

Kanis, J. A., Johnell, O., Odén, A., Johansson, H., Eisman, J. A., Fujiwara, S., ... & Tenenhouse, A. (2006). The use of multiple sites for the diagnosis of osteoporosis. *Osteoporosis international*, *17*(4), 527-534.

Karasik, D., Hannan, M. T., Kiel, D. P., Bouxsein, M. L., & Cupples, L. A. (2002). Is a factor score combining dual X-ray absorptiometry and ultrasound a better phenotype for genetic studies. *J Bone Miner Res*, *17*(suppl 1), S236.

Karasik, D., Cupples, L. A., Hannan, M. T., & Kiel, D. P. (2004). Genome screen for a combined bone phenotype using principal component analysis: the Framingham study. *Bone*, *34*(3), 547-556.

Karasik, D., & Ferrari, S. L. (2008). Contribution of Gender-Specific Genetic Factors to Osteoporosis Risk. *Annals of human genetics*, *72*(5), 696-714.

Karasik, D., Zhou, Y., Cupples, L. A., Hannan, M. T., Kiel, D. P., & Demissie, S. (2008). Bivariate Genome-Wide Linkage Analysis of Femoral Bone Traits and Leg Lean Mass: Framingham Study. *Journal of Bone and Mineral Research*, *24*(4), 710-718.

Karasik, D., Hsu, Y. H., Zhou, Y., Cupples, L. A., Kiel, D. P., & Demissie, S. (2010). Genome-wide pleiotropy of osteoporosis-related phenotypes: The Framingham Study. *Journal of Bone and Mineral Research*, *25*(7), 1555-1563.

Kiel, D. P., Demissie, S., Dupuis, J., Lunetta, K. L., Murabito, J. M., & Karasik, D. (2007). Genome-wide association with bone mass and geometry in the Framingham Heart Study. *BMC medical genetics*, *8*(Suppl 1), S14.

Knott, S. A., & Haley, C. S. (2000). Multitrait least squares for quantitative trait loci detection. *Genetics*, *156*(2), 899-911.

Korol, A. B., Ronin, Y. I., & Kirzhner, V. M. (1995). Interval mapping of quantitative trait loci employing correlated trait complexes. *Genetics*, *140*(3), 1137-1147.

Korol, A. B., Ronin, Y. I., Itskovich, A. M., Peng, J., & Nevo, E. (2001). Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. *Genetics*, *157*(4), 1789-1803.

Koshkoih, A. E. (2007). Multiple trait analysis for genetic mapping of quantitative trait loci for carcass and beef quality.

Kyselý, J. (2010). Coverage probability of bootstrap confidence intervals in heavy-tailed frequency models, with application to precipitation data. *Theoretical and applied climatology*, *101*(3), 345-361.

Lange, K., Westlake, J., & Spence, M. (1976). Extensions to pedigree analysis III. Variance components by the scoring method. *Annals of Human Genetics*, *39*(4), 485-491.

Lange, K., Boehnke, M., & Opitz, J. M. (1983). Extensions to pedigree analysis. IV. Covariance components models for multivariate traits. *American journal of medical genetics*, *14*(3), 513-524.

Lange, K. (2002). *Mathematical and statistical methods for genetic analysis* (No. Ed. 2). Springer-Verlag, New York.

Lattin, J. M., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Pacific Grove, CA: Thomson Brooks/Cole.

Lewiecki, E. M., Kendler, D. L., Kiebzak, G. M., Schmeer, P., Prince, R. L., El-Hajj Fuleihan, G., & Hans, D. (2004). Special report on the official positions of the International Society for Clinical Densitometry. *Osteoporosis international*, *15*(10), 779-784.

Lewiecki, E. M., Watts, N. B., McClung, M. R., Petak, S. M., Bachrach, L. K., Shepherd, J. A., & Downs, R. W. (2004). Official positions of the international society for clinical densitometry. *Journal of Clinical Endocrinology & Metabolism*, *89*(8), 3651-3655.

Li, R., Tsaih, S. W., Shockley, K., Stylianou, I. M., Wergedal, J., Paigen, B., & Churchill, G. A. (2006). Structural model analysis of multiple quantitative traits. *PLoS genetics*, *2*(7), e114.

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13-22.

Liu, G. F., Yang, J., Xu, H. M., Hayat, Y., & Zhu, J. (2008). Genetic analysis of grain yield conditioned on its component traits in rice (Oryza sativa L.). *Crop and Pasture Science*, *59*(2), 189-195.

Livshits, G., Karasik, D., Pavlovsky, O., & Kobyliansky, E. (1999). Segregation analysis reveals a major gene effect in compact and cancellous bone mineral density in 2 populations. *Human biology*, *71*(2), 155.

Lunetta, K. L. (2008). Genetic association studies. *Circulation*, *118*(1), 96-101.

Lynch M., & Walsh B. (1998) *Genetics and analysis of quantitative traits*. Sinauer Associates, Inc, Sunderland.

Mahaney, M. C., Blangero, J., Comuzzie, A. G., VandeBerg, J. L., Stern, M. P., & MacCluer, J. W. (1995). Plasma HDL cholesterol, triglycerides, and adiposity: a quantitative genetic test of the conjoint trait hypothesis in the San Antonio Family Heart Study. *Circulation*, *92*(11), 3240-3248.

Mangin, B., Thoquet, P., & Grimsley, N. (1998). Pleiotropic QTL analysis. *Biometrics*, 88-99.

Nakai, M., & Ke, W. (2009). Statistical Models for Longitudinal Data Analysis. *Applied Mathematical Sciences*, *3*(40), 1979-1989.

Neale, M. C. C. L., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families* (No. 67). Springer.

North, K. E., Howard, B. V., Welty, T. K., Best, L. G., Lee, E. T., Yeh, J. L., ... & MacCluer, J. W. (2003). Genetic and Environmental Contributions to Cardiovascular Disease Risk in American Indians The Strong Heart Family Study. *American journal of epidemiology*, *157*(4), 303-314.

Ovilo, C., Clop, A., Noguera, J. L., Oliver, M. A., Barragan, C., Rodriguez, C., ... & Perez-Enciso, M. (2002). Quantitative trait locus mapping for meat quality traits in an Iberian× Landrace F2 pig population. *Journal of Animal Science*, *80*(11), 2801-2808.

O'Brien, P. C. (1984). Procedures for comparing samples with multiple endpoints. *Biometrics*, 1079-1087.

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559-572.

Ralston, S. H. (2002). Genetic control of susceptibility to osteoporosis. *Journal of Clinical Endocrinology & Metabolism*, *87*(6), 2460-2466.

Ralston, S. H., & De Crombrugghe, B. (2006). Genetic regulation of bone mass and susceptibility to osteoporosis. *Genes & development*, *20*(18), 2492-2506.

Ricki L. (2003) Multifactorial Traits. McGraw-Hill Higher Education.

Rosa, G. J., Valente, B. D., de los Campos, G., Wu, X. L., Gianola, D., & Silva, M. A. (2011). Inferring causal phenotype networks using structural equation models. *Genetics Selection Evolution*, *43*(1), 1-13.

Schork, N. J. (1993). Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *American journal of human genetics*, *53*(6), 1306–1319.

Schork, N. J., Weder, A. B., Trevisan, M., & Laurenzi, M. (1994). The contribution of pleiotropy to blood pressure and body-mass index variation: the Gubbio Study. *American journal of human genetics*, *54*(2), 361.

Searle, S. R. (1961). Phenotypic, genetic and environmental correlations. *Biometrics*, 474-480.

Shi C.H., Zhu Q.Q., Wang K.M., Ge G.K., Wu J.G., and Xu Z.G. (2010) Detecting relationships between amylose content and amino acid contents of indica rice with conditional approach. Journal of Genetics, 89(1): 1-8.

Shipley, B. (2000). *Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference*. Cambridge University Press.

Stearns, T. M., Beever, J. E., Southey, B. R., Ellis, M., McKeith, F. K., & Rodriguez-Zas, S. L. (2005). Evaluation of approaches to detect quantitative trait loci for growth, carcass, and meat quality on swine chromosomes 2, 6, 13, and 18. I. Univariate outbred F2 and sib-pair analyses. *Journal of animal science*, *83*(7), 1481-1493.

Stein, C. M., Schick, J. H., Taylor, H. G., Shriberg, L. D., Millard, C., Kundtz-Kluge, A., ... & Iyengar, S. K. (2004). Pleiotropic effects of a chromosome 3 locus on speech-sound disorder and reading. *American journal of human genetics*, *74*(2), 283.

Stoner, J. A., & Leroux, B. G. (2002). Analysis of clustered data: A combined estimating equations approach. *Biometrika*, *89*(3), 567-578.

Tucker, K. L., Morita, K., Qiao, N., Hannan, M. T., Cupples, L. A., & Kiel, D. P. (2006). Colas, but not other carbonated beverages, are associated with low bone mineral density in older women: The Framingham Osteoporosis Study. *The American journal of clinical nutrition*, *84*(4), 936-942.

Vinson, A., Mahaney, M. C., Cox, L. A., Rogers, J., VandeBerg, J. L., & Rainwater, D. L. (2008). A pleiotropic QTL on 2p influences serum Lp-PLA< sub> 2</sub> activity and LDL cholesterol concentration in a baboon model for the genetics of atherosclerosis risk factors. *Atherosclerosis*, *196*(2), 667-673.

Wei, L. J., & Johnson, W. E. (1985). Combining dependent tests with incomplete repeated measurements. *Biometrika*, *72*(2), 359-364.

Weller, J. I., Wiggans, G. R., VanRaden, P. M., & Ron, M. (1996). Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multi-trait experiment. *TAG Theoretical and Applied Genetics*, *92*(8), 998-1002.

Williams, J. T., Van Eerdewegh, P., Almasy, L., & Blangero, J. (1999). Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. I. Likelihood formulation and simulation results. *American journal of human genetics*, *65*(4), 1134.

Williams, J. T., Begleiter, H., Porjesz, B., Edenberg, H. J., Foroud, T., Reich, T., ... & Blangero, J. (1999). Joint multipoint linkage analysis of multivariate qualitative and quantitative traits. II. Alcoholism and event-related potentials. *American journal of human genetics*, *65*(4), 1148.

Wright, S. (1921). Correlation and causation. *Journal of agricultural research*, *20*(7), 557-585.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, 161-215.

Wright, S. (1977). *Evolution and the genetics for populations, Vol III: Experimental results and evolutionary deductions.* Chicago University Press.

Wu. H. (2009). *Methods for genetic association studies using longitudinal and multivariate phenotypes in families*. (Doctoral dissertation).

Yang, Q., Wu, H., Guo, C. Y., & Fox, C. S. (2010). Analyze multivariate phenotypes in genetic association studies by combining univariate association tests. *Genetic epidemiology*, *34*(5), 444-454.

Zeger, S. L., & Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 121-130.

Zhao, J., Xiao, P., Guo, Y., Liu, Y. J., Pei, Y. F., Yang, T. L., ... & Deng, H. W. (2008). Bivariate genome linkage analysis suggests pleiotropic effects on chromosomes 20p and 3p for body fat mass and lean mass. *Genetics research*, *90*(03), 259-268.

Zhu, J. (1995). Analysis of conditional genetic effects and variance components in developmental genetics. *Genetics*, *141*(4), 1633-1639.

Zorn, C. J. (2001). Generalized estimating equation models for correlated data: A review with applications. *American Journal of Political Science*, 470-490.

CURRICULUM VITAE