Statistical Methods for Modeling Heterogeneous Effects in Genetic Association Studies

by

Jingchunzi Shi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2018

Doctoral Committee:

Associate Professor Seunggeun Shawn Lee, Chair
Professor Michael Boehnke
Professor Bhramar Mukherjee
Associate Professor Cristen Willer

Jingchunzi Shi

shijingc@umich.edu

ORCID id: 0000-0002-9671-2568

To my family for their love and support !

ACKNOWLEDGEMENTS

The pursuit of my doctorate degree has been a rewarding yet extremely humbling experience. This dissertation would not have been possible without the support and guidance from my committee, friends and family.

First, I would like to express my deepest gratitude to my committee chair, Dr. Seunggeun Shawn Lee, who introduced me to statistical genetics and persuaded me to pursue research in statistical methods development. Since then, he has provided me invaluable research opportunities and guidance to build a solid foundation in genetics and biostatistics. He has been extremely patient, accommodating and encouraging. His excellent mentorship has helped me mature as an independent scientist and as an individual alike.

I would like to extend the gratitude to Dr. Bhramar Mukherjee, who is always there to help and offer her guidance on research as well as career building. Her dedication and enthusiasm in science has profoundly inspired me to keep fighting the research challenges in those late nights in SPH. I truly look up to Dr. Mukherjee as my role model of a successful female scientist. I would also like express my heartfelt thanks to Dr. Michael Boehnke and Dr. Cristen Willer for their valuable and insightful advice for my research projects.

In addition, I would like to express my deep appreciation to Kirsten Herold, who has offered tremendous help on my scientific writing. I have benefited greatly from her instruction on becoming a better writer. I would also like to thank Alan Kwong for his generous help on Linux and Python programming. The same gratitude goes

to members of the Lee lab for their great exchange of scientific ideas. I truly treasure our scholastic discussions and friendships.

Last but not least, I would like to thank Paul Imbriano and my parents, for their continuous encouragement and support. They are always there to cheer me up, listen to my fears, and share my laugh. It is their love and understanding that help me get through hard times and keep me brave to face challenges. Without their blessings and encouragement, I would not have been able to finish my journey in graduate school.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF ABBREVIATIONS

HMP3: Phase III of the HapMap Project

GLMM: Generalized Linear Mixed Model

GWAS: Genome-wide Association Studies

LD: Linkage Disequilibrium

LDL-C: Low Density Lipoprotein Cholesterol

LMM: Linear Mixed Model

LRT: Likelihood Ratio Test

MAF: Minor Allele Frequency

MGI: Michigan Genomics Initiative

SAMM: Semi-parametric Additive Mixed Model

SNP: Single Nucleotide Polymorphism

T2D: Type-2 Diabetes

T2D-GENES: Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples

WES: Whole Exome Sequencing

WGS: Whole Genome Sequencing

ABSTRACT

Effect-size heterogeneity is a commonly observed phenomenon when aggregating studies from different ancestries to conduct trans-ethnic meta-analysis. Irrespective of the sources of heterogeneity, classical meta-analysis approaches cannot appropriately account for the expected between-study heterogeneity. Therefore, to bridge the methodological gap, in the first two projects, I develop statistical methods for modeling the heterogeneous effects in trans-ethnic meta-analysis for genome-wide association studies (GWASs). In the third project, I extend the methods in trans-ethnic GWAS meta-analysis to a general statistical framework for modeling heterogeneity in biomedical studies.

In the first project, I develop a score test for the common variant GWAS trans-ethnic meta-analysis. To account for the expected genetic effect heterogeneity across diverse populations, I adopt a modified random effects model from the kernel regression framework, and use the adaptive variance component test to achieve robust power regardless of the degree of genetic effect heterogeneity. From extensive simulation studies, I demonstrate that the proposed method has well-calibrated type I error rates at very stringent significance levels and can improve power over traditional meta-analysis methods.

In the second project, I extend the common variant meta-analysis approach to the gene-based rare variant trans-ethnic meta-analysis. I develop a unified score test which is capable of incorporating different levels of heterogeneous genetic effects across multiple ancestry groups. I employ a resampling-based copula method to estimate the asymptotic distribution of the proposed test, which enables efficient estimation of p-values. I conduct simulation studies to demonstrate that the proposed approach is well-calibrated at stringent significance levels and improves power over current approaches under the existence of genetic effect heterogeneity. As a real data application, I further apply the proposed method to the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia data to explore rare variant associations with several traits.

In the third project, I develop a supremum score test for jointly testing the fixed and random effects in a generalized linear mixed model (GLMM). The joint testing framework has many applications in biomedical studies. One example is to use such tests for ascertaining associations under the existence of heterogeneity in GWAS meta-analysis; another example is the nonparametric test of spline curves. The supremum score test first re-parameterizes the fixed effects terms as a product of a scale parameter and a vector of nuisance parameters. With such re-parameterization, the joint test is equivalent to testing whether the scale parameter is zero. Since the nuisance parameters are unidentifiable under the null hypothesis, I propose using the supremum of score test statistics over the nuisance parameters. I employ a resampling-based copula method to approximate the asymptotic null distribution of the proposed score test statistic. I first investigate the performance of the method through simulation studies. Using the Michigan Genomics Initiative (MGI) data, I then demonstrate its application by assessing whether the genetics effects to Low Density Lipoprotein Cholesterol (LDL-C) can be modified by age.

CHAPTER I

Introduction

1.1 Mapping Human Complex Traits

The development of recombinant DNA and other molecular techniques in the 1970s have profoundly altered the practice of human biology. Since then, new findings in genetics and molecular biology are emerging at an unprecedented clip. These findings provide new insights into the human genome, and are continuously shaping our understanding of the genetic basis of health and disease. As of April 2018, the genes underlying 76% of 6,727 known monogenic Mendelian disorders have been identified (Amberger et al., 2014). Despite the success in gene mapping of Mendelian disorders, identifying the genetic risks for complex diseases/traits remains a challenging task, since those disorder/phenotype susceptibilities are usually influenced by genetic variants in multiple genes, environmental/behavioral factors, as well as their possible interactions.

In 1996, Risch and Merikangas (Risch and Merikangas, 1996) predicted that association studies, which compare the frequency of alleles in a particular variants between affected and unaffected individuals, can be effective tools for studying complex traits because of their statistical power to detect genes of small effect. However, such an approach was constrained due to the limited number of polymorphisms that could be genotyped at that time. For example, the available number of markers was typically in the tens, and the sample size was usually in the hundreds. In the beginning of the 21st century, advances in the genotyping technology and the dramatically decreased genotyping cost began to facilitate the detection of a large number of polymorphisms across the entire human genome. Since then, genome-wide association studies (GWAS) have led to a plethora of discoveries for various human complex diseases/traits (MacArthur et al., 2016).

Early GWAS mainly focused on identifying common genetic variants using single nucleotide polymorphism (SNP) arrays, but the success of these GWAS has been primarily confined to European populations. The ability to aggregate cohort-specific summary statistics from multiple studies via meta-analysis techniques has further promoted more GWAS findings in populations of European descent. However, accumulating evidence has demonstrated that, for a variety of complex diseases/traits, there is substantial overlap in trait-associated loci between different ethnicities (Farrer et al., 1997; Dumitrescu et al., 2011; Carlson et al., 2013). Therefore, it is expected that the efficiency of complex-trait association studies can be further improved when populations of non-European descent are analyzed in conjunction with the European populations via trans-ethnic meta-analysis. However, the classical meta-analysis approaches – both fixed-effects and random-effects models – are not appropriate for combining data across race and ethnicity, and quite limited research exists in developing powerful multi-ethnic GWAS meta-analysis methods for common variants. Thus, trans-ethnic meta-analysis methods for common variant associations are greatly needed.

As of April 2018, while GWASs have successfully identified 53,069 unique SNP-trait associations (MacArthur et al., 2016), most tend to have low to moderate effect and explain only a fraction of the overall heritability (Manolio et al., 2009). The fact that array-based GWAS findings have not been able to fully explain the trait variations led to the widely debated "missing heritability" question after the 1^{st} wave of GWAS. A natural conjecture for the possible explanation to this problem is that the missing disease heritability is due to rare and low-frequency variants, some with large effects, which could not be captured in the genotyping array platforms (Frazer et al., 2009; Eichler et al., 2010). While the array-based GWAS continues to unearth trait-associated variants, accessibility in terms of cost and technology of next-generation sequencing has opened up the entire spectrum of genome variations for the analysis of complex diseases/traits. In 2014, the cost for whole-genome sequencing (WGS) reached the US$1000 per genome milestone. In contrast to the array-based GWAS, which focus on SNPs that are in linkage disequilibrium (LD) with the causal variants, whole-genome sequencing shifts our research interest toward analyzing causal variants and genes directly. Recent advances in sequencing technology, availability of high-quality human reference panels, and improvements in genotyping imputation accuracy have made it possible to comprehensively catalog genetic variation in population samples. In fact, sequencing studies have successfully identified rare variants that are involved in complex traits, including prostate cancer (Gudmundsson et al., 2012), Alzheimer disease (Cruchaga et al., 2014), lipids and coronary artery disease (Peloso et al., 2014) and many others.

Despite its potential contributions to solving the "missing heritability" problem, one inevitable challenge for the design and analysis of sequencing-based GWAS is that rare variant tests are usually underpowered without an exceptionally large sample size or a sufficient number of rare alleles captured (Bansal et al., 2010). One practical strategy to improve power is to conduct trans-ethnic meta-analysis, which combines summary statistics across studies from different ethnicities to increase sample sizes. However, under the presence of inter-study genetic effect heterogeneity across ancestries, existing meta-analysis approaches may be unsatisfactory because they do not take into consideration that studies from the more closely related ancestries can be more homogeneous than those that are more distantly related. In order to take full advantage of the strengths of multi-ethnic meta-analysis, powerful trans-ethnic GWAS meta-analysis methods for rare variant associations are greatly needed.

1.2 The Need for Trans-Ethnic Meta-Analysis

Meta-Analysis is a practical tool to aggregate studies that have already been conducted. In ideal situations, meta-analysis can achieve eventually equal power as the joint analysis (Liu et al., 2014). Besides its ability of increasing sample sizes without the cost of additional genotyping, meta-analysis has several logistic and ethical advantages over the joint analysis of individual level data. First, meta-analysis only requires summary statistics from each participating study, which avoids the cumbersome integration of genotype and phenotype data from different studies, and protects the privacy of study participants. Second, different studies may require different sets of covariates, which can be difficult to accommodate in a joint analysis, but can be easily incorporated at the summary level statistic in meta-analysis.

Initially, GWAS meta-analyses were mostly European-based, and have proved to be worthwhile in identifying additional complex trait loci (Bustamante et al., 2011). Recently, GWAS have been undertaken in other ethnic groups including Africans, East and South Asians and Hispanics (Popejoy and Fullerton, 2016). With the increasing availability of GWAS from distinct ethnicities, trans-ethnic meta-analysis offers an exciting opportunity to enrich the association strengths with the further increased sample sizes and fine mapping through different LD patterns (Li and Keating, 2014). In fact, trans-ethnic meta-analyses have successfully identified novel loci that are associated with common oncologic diseases, including breast cancer (Siddiq et al., 2012) and prostate cancer (Kote-Jarai et al., 2011); metabolic and cardiovascular diseases/ traits, including high-/low- density lipoprotein (HDL/LDL) levels(Coram et al., 2013), blood pressure (Franceschini et al., 2013) and coronary artery disease (Dichgans et al., 2014); immune diseases such as rheumatoid arthritis (RA) (Okada et al., 2014) and asthma (Lasky-Su et al., 2012); and many others.

For trans-ethnic GWAS meta-analysis, in addition to its primary objective on disease/trait locus discoveries, several other goals can be simultaneously accomplished using the features of trans-ethnic study designs. Firstly, trans-ethnic GWAS meta-analysis provides an independent replication sample set that can be used to validate single-population GWAS findings and to eliminate concerns about sub-/cryptic- population stratification in the single GWAS discoveries (Campbell et al., 2005). Those validated variants can be further used to prioritize loci for secondary replication and sequencing studies (Cantor et al., 2010). Secondly, in trans-ethnic GWAS meta-analysis, differences in LD structures across genetically diverse populations is potentially a powerful tool for fine mapping the rare or causal variants that underlie disease associations (Teo et al., 2010). Despite the promising potential, however, the between-study genetic effect heterogeneity among different ethnic groups, e.g. unequal genetic effect sizes among studies (Wang et al., 2013), presents new challenges in performing trans-ethnic meta-analysis.

In GWAS meta-analysis of common variants, several reasons can contribute to the emergence of complex between-study heterogeneity patterns. First, it is highly possible that the queried SNP is not the underlying causal SNP, but rather is correlated to the causal SNP through LD. Therefore, due to variations in the LD structures across ancestry groups, the same high-risk allele may have different patterns of association with the causal allele among different populations, leading to the observed unequal genetic effects at the marker alleles. This phenomenon can be particularly relevant under the common disease – common variants (CD-CV) model (Cargill et al., 1999; Chakravarti, 1999), in which differential recombination histories can occur due to the varying age of the mutations in the different populations (Pritchard and Przeworski, 2001). Moreover, the presence of hidden stratification in some populations may produce spurious associations or alter the patterns of true associations, and therefore further complicate the between-study heterogeneity (Morton and Collins, 1998; Pritchard and Przeworski, 2001; Thomas and Witte, 2002; Stumpf and Goldstein, 2003; Freedman et al., 2004). In addition, the genetic variant of interest may interact with other environmental, dietary and lifestyle factors. Thus, the difference in these factors among populations can generate variability in the marginal genetic effects between studies (Morris, 2011).

For rare variant GWAS meta-analysis, even if the same variant selection criteria are employed, different studies may still present different sets of rare variants, since rare variants are often population specific. Therefore, the gene-level association power will likely be unequal among studies, even when effect size across studies is the same for each variant. The possible gene-environment interaction, which contributes to heterogeneous genetic effects in common variant GWAS meta-analysis, also adds to the complex between-study heterogeneity patterns in rare variant meta-analyses. Irrespective of the sources of genetic heterogeneity, classical GWAS meta-analysis approaches cannot appropriately account for the expected between-study heterogeneity. Therefore, to bridge this methodological gap, in the first two projects of this dissertation I present two novel statistical methods for modeling the heterogeneous effects in genetic association studies – for both common variants and rare variants – to improve the power of trans-ethnic meta-analysis.

1.3 A General Statistical Framework for Modeling Heterogeneity in Biomedical Studies

The key idea in modeling the between-study heterogeneity in GWAS meta-analysis is to decompose the magnitude of the effect size into two components: a fixed constant which represents the mean effect size over all populations, and a random variable which measures deviation of the study-specific effect from the population mean.

Consequently, assessing associations in trans-ethnic meta-analysis is essentially a joint testing of fixed and random effects.

The statistical framework of jointly testing the fixed and random effects has many applications in the field of biomedical studies. One such example is the likelihood ratio test proposed by Han and Eskin (2011) for ascertaining association signals under the existence of between-study heterogeneity in GWAS meta-analysis. Another example is the SKAT-O test proposed by Lee et al. (2012), which optimally combines the burden (Madsen and Browning, 2009) and SKAT (Wu et al., 2011) tests for assessing the gene-/region-based rare variant association strengths. It can be shown that an alternative way of deriving the score statistic for SKAT-O is to jointly test the mean (i.e. the fixed effect) and variance component (i.e. the random effect) of the regression coefficient of the genetic variant. To assess the age-varying genetic effect, one can incorporate age (or any non-genetic modifier of interest) into a non-parametric functional form in the varying coefficient model and reformulate the problem of testing the varying coefficient into jointly testing the fixed and random effects in a generalized linear mixed model (GLMM) (Zhang and Lin, 2003; Wang and Chen, 2012). Last but not least, the joint testing framework can be employed in the non-parametric test of spline curves in a semi-parametric additive mixed model (SAMM).

Testing the random effects involves constraints on the variance component parameters, in which classical inference with a standard null distribution no longer holds, because those parameters under the null hypothesis lie on the boundary of the maintained hypothesis (Lin, 1997; Andrews, 2001). Although the statistical literature offers an array of methods for testing the fixed and random effects jointly for Gaussian responses, corresponding methods for non-Gaussian outcomes remain limited. In response, in my last project, I propose to extend the methods in trans-ethnic GWAS meta-analysis to a general statistical framework with a score test for the joint testing problem in a GLMM.

1.4 Dissertation Outline

In Chapter II, I develop a score test for the common variant associations in trans-ethnic meta-analysis. To account for the effect-size heterogeneity across diverse populations, I adopt a modified random effects model from the kernel regression framework. Specifically, I treat the genetic effect coefficients as random variables and construct their correlation structure to reflect the level of genetic effect similarities across ancestry groups. In addition, I use an adaptive variance component test to achieve robust power regardless of the degree of genetic effect heterogeneity. Through analytical approximation of the asymptotic distribution of the proposed test, I achieve efficient computing time for genome-wide datasets, as the method requires less than 3 hours on a Linux cluster node with 2.80 GHz CPU to analyze one million variants. Using extensive simulation studies, I demonstrate that the proposed method has well-calibrated type I error rates at very stringent significance levels and improves power over the traditional meta-analysis methods. Re-analyzing a published type 2 diabetes GWAS meta-analysis (Mahajan et al., 2014), I successfully identify one additional SNP which exhibits genetic effect heterogeneity across ethnicities.

In Chapter III, I extend the score test in Chapter 2 to the gene-/region-based rare variant trans-ethnic meta-analysis in sequencing association studies. The proposed method is capable of not only accounting for the expected heterogeneous genetic effects among studies, but also flexibly modeling varying levels of heterogeneity according to the relatedness between the populations. The proposed method only requires sharing of study-specific summary statistics, such as the score statistics for each variant and the corresponding information matrices which summarize the LD structures between the variants. I employ a resampling-based copula method to estimate the asymptotic null distribution of the proposed test, which enables efficient estimation of p-values. I conduct simulation studies to demonstrate that the proposed approach is well-calibrated at stringent significance levels and improves power over current approaches under the existence of genetic effect heterogeneity. As a real data application, I further apply the proposed method to the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia data to explore rare variant associations with several traits.

In Chapter IV, I develop a supremum score test for jointly testing the fixed and random effects in a generalized linear mixed model (GLMM) for both Gaussian and non-Gaussian outcomes. The framework of jointly testing the fixed and random effects has many applications in biomedical studies. One example is to use such tests for ascertaining associations under the existence of heterogeneity in GWAS meta-analysis; another example is the nonparametric test of spline curves. Although extensive research has been conducted on testing random effect terms only, little work has been done for the joint test of fixed and random effects, especially for non-Gaussian outcomes. Hence, I propose a score test for the joint test in a GLMM to handle both the Gaussian and non-Gaussian response types, and use analytical derivation as well as numerical simulation to demonstrate that the proposed score test is asymptotically equivalent to the corresponding likelihood-ratio test (LRT). The method first re-parameterizes the fixed effects terms as a product of a scale parameter and a vector of nuisance parameters. With such re-parameterization, the joint test is equivalent to testing whether the scale parameter is zero. Since the nuisance parameters are unidentifiable under the null hypothesis, I propose using the supremum of score test statistics over the nuisance parameters. I employ a resampling-based copula method to approximate the asymptotic null distribution of the proposed score test statistic. I first investigate the performance of the method through simulation studies. Using the Michigan Genomics Initiative (MGI) data, I then demonstrate its application by assessing whether the genetics effects to Low Density Lipoprotein Cholesterol (LDL-

C) can be modified by age. Finally, in Chapter V, I discuss the implications of my work and propose potential future directions to pursue.

CHAPTER II

A Novel Random Effect Model for GWAS Meta-Analysis and its Application to Trans-Ethnic Meta-Analysis

Abstract

Meta-analysis of trans-ethnic genome-wide association studies (GWAS) has proven to be a practical and profitable approach for identifying loci that contribute to the risk of complex diseases. However, the effect-size heterogeneity cannot be easily accommodated through existing fixed-effects and random-effects methods. In response, we propose a novel random effect model for trans-ethnic meta-analysis with flexible modeling of the expected genetic effect heterogeneity across diverse populations. Specifically, we adapt a modified random effect model from the kernel regression framework, in which genetic effect coefficients are random variables whose correlation structure reflects the genetic distances across ancestry groups. In addition, we use the adaptive variance component test to achieve robust power regardless of the degree of genetic effect heterogeneity. Simulation studies show that our proposed method has well-calibrated type I error rates at very stringent significance levels and can improve power over the traditional meta-analysis methods. We re-analyze the published type 2 diabetes GWAS meta-analysis (Mahajan et al., 2014) and successfully identify one additional SNP that clearly exhibits genetic effect heterogeneity across different ancestry groups. Furthermore, our proposed method provides scalable computing time for genome-wide datasets, in which an analysis of one million SNPs would require less than 3 hours on a Linux cluster node with 2.80 GHz CPU to analyze one million variants.

Keywords: Common variants; Effect-size heterogeneity; GWAS; Kernel regression; Random effect model; Trans-ethnic meta-analysis.

2.1 Introduction

Although genome-wide association studies (GWAS) have successfully identified more than 50,000 loci that influence the severity of human health outcomes, those identified loci account for only a small fraction of the genetic contribution to most complex diseases and traits (McCarthy et al., 2008; MacArthur et al., 2016). It has been argued that numerous loci with very small effects can explain additional disease risk or trait heritability, and the challenge is to find those loci that can be identified only with very large numbers of samples (Eichler et al., 2010). Since it can be challenging to design and conduct a single study with tens or hundreds of thousands of samples, a more practical alternative is to combine studies that have already been conducted through a meta-analysis (Evangelou and Ioannidis, 2013).

A natural extension of the single-ancestry-based meta-analysis is to include samples from as many studies as possible, even if they come from genetically disparate ancestries. With the further enlarged sample size, trans-ethnic meta-analysis is expected to be more powerful at detecting novel loci without the cost of additional genotyping (Cooper et al., 2008). In fact, several trans-ethnic meta-analyses have been performed in the past few years with success in discovering risk alleles across ancestry groups. For example, five consortia (Mahajan et al., 2014) aggregated published GWAS meta-analyses of type 2 diabetes (T2D) from four ancestry groups and successfully identified seven new loci with very small effect sizes.

To take full advantage of the profitability of trans-ethnic meta-analysis, improved statistical methods are required to account for the distinctive ancestral origins among data. Existing methods for GWAS meta-analysis include the classical fixed-effects and random-effects methods, as well as the recently introduced new random-effects method by Han and Eskin (2011) and the Bayesian approach by Morris (2011). The fixed-effects method (FE) (Hedges and Vevea, 1998) is the most popular approach for synthesizing single-ancestry GWAS data. It assumes that the true effect of each risk allele is the same in each data set, and as a result, it has limited power in the presence of genetic effect heterogeneity (Evangelou and Ioannidis, 2013; Wang et al., 2013). The random-effects method (RE) was developed explicitly to model the between-study heterogeneity; however, it implicitly assumes heterogeneity under the null hypothesis, which causes it to have much lower power than FE (Han and Eskin, 2011). To relax the conservative assumption of RE, Han and Eskin (2011) developed a new random-effects model (RE-HE) which achieves higher power than RE. Morris (2011) developed a trans-ethnic meta-analysis method by means of a Bayesian partition model (MANTRA). MANTRA accounts for the relatedness of studies by grouping them into different ethnic clusters. Specifically, studies that are grouped into the same ethnic cluster share the same underlying genetic effect, while different ethnic clusters have different underlying genetic effects.

The aforementioned T2D trans-ethnic meta-analysis (Mahajan et al., 2014) was carried out using the FE method. In addition to identifying novel T2D susceptibility loci, they analyzed 69 established T2D susceptibility loci using Cochran Q test (Cochran, 1954) to evaluate their genetic effect heterogeneity. Among the 69 loci, 3 had very strong evidence of the heterogeneity (Cochran Q p-value $< 10^{-3}$), and 12 had some evidence of the heterogeneity ($10^{-3} \leq$ Cochran Q p-value < 0.05). For those 15 loci, FE may not be sufficiently powerful to detect the association signals. To improve power, we develop a new trans-ethnic meta-analysis approach, referred to as TransMeta, and use it to reanalyze the T2D trans-ethnic meta-data.

As mentioned above, one challenge in trans-ethnic GWAS meta-analysis is to appropriately account for the effect-size heterogeneity. There can be several reasons for the heterogeneous effect sizes. First, it is highly possible that the queried SNP is not the underlying causal SNP, but rather is correlated to the causal SNP through linkage disequilibrium (LD). Variations in the LD structures across ancestry groups can lead to the observed genetic effect heterogeneity. Second, the environmental risk factors may differ between ancestry groups. With the possibility of interaction between the causal variants and the environmental factors, marginal genetic effects may vary between populations (Morris, 2011). To address the heterogeneity issue, we consider a modified random effect model based on a kernel machine framework (Liu et al., 2007). Specifically, we treat the genetic effect coefficients as random variables, with their correlation structure across ancestry groups reflecting the expected heterogeneity (or homogeneity) among ancestry groups. To test for associations, we derive a data-adaptive variance component test with adaptive selection of the degree of heterogeneity. This adaptive test combines models of homogeneous and heterogeneous genetic effects, and provides robust power regardless of the genetic effect distribution. We provide details of our proposed method in Section 2.2.

The rest of this chapter is organized as follows: In Section 2.3, we first perform simulation studies to compare the performance of TransMeta with FE, RE, RE-HE and MANTRA for meta-analyzing GWAS across genetically diverse populations. We then illustrate application of TransMeta by reanalyzing the T2D GWAS in Mahajan et al. (2014). We conclude this chapter with a discussion in Section 2.4. Supplementary texts, tables and figures are presented in Section 2.5.

2.2 Methods

2.2.1 Statistical Models for GWAS Meta-Analysis

In this section, we first introduce statistical models of the existing GWAS meta-analysis methods. Let $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_n)^T$ be the effect-size estimates, such as the log odds ratios or regression coefficients, in n independent studies. If the sample sizes in each study are sufficiently large, then

$$\widehat{\boldsymbol{\beta}}|\boldsymbol{\beta} \sim MVN(\boldsymbol{\beta}, \Sigma), \tag{2.2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^T$, with β_i being the true effect size in the ith study; and $\Sigma = diag(\sigma_1^2, \ldots, \sigma_n^2)$, with σ_i^2 being the variance of $\widehat{\beta}_i$.

FE assumes that all the studies share a common effect-size μ (i.e. $\beta_1 = \cdots = \beta_n = \mu$). FE is powerful at detecting genetic effects that are present in most, if not all, of the studies with homogeneous effect sizes. The RE model assumes that the true effect size β_i for the ith study is generated from a normal distribution with mean μ and variance τ_1,

$$\beta_i = \mu + \eta_i, \quad \eta_i \sim N(0, \tau_1). \tag{2.2.2}$$

RE typically assumes that even under the null hypothesis of no association, β_is can be different across studies, since τ_1 is not assumed to be zero under the null hypothesis. Due to this conservative assumption, RE tends to be less powerful at detecting association signals than FE, although it is proposed to account for the expected heterogeneity. Han and Eskin (2011) developed a new RE approach (RE-HE) that assumes no genetic effect heterogeneity under the null hypothesis. Specifically, they assumed that β_is are zero among all the studies under the null hypothesis (i.e. $\mu = 0$ and $\tau_1 = 0$), and they allowed varying effect sizes among studies under the alternative hypothesis. The likelihood ratio test was employed to evaluate the null hypothesis of $\mu = 0$ and $\tau_1 = 0$. Since asymptotic p-values of RE-HE are only accurate when the number of studies (n) is very large, they provided tabulated p-values precomputed with an assumption of equal sample sizes across studies. In the presence of between-study effect-size heterogeneity, RE-HE yields higher power than FE.

The aforementioned three frequentist meta-analysis methods can all be summarized under model (2.2.2) with certain assumptions on τ_1. With $\tau_1 = 0$ under both the null and the alternative hypotheses, model (2.2.2) is exactly the same as FE. RE assumes that τ_1 is non-zero under both the null and the alternative hypotheses, and tests whether $\mu = 0$ or not, while accounting for the between-study variance τ_1. RE-HE assumes that $\tau_1 = 0$ under the null hypothesis, and tests whether both μ and τ_1 are zero under the alternative hypothesis.

Unlike the frequentist approaches, the Bayesian meta-analysis approach, MANTRA, assigns studies into ethnic clusters under model (2.2.1). It assumes that studies that are grouped into the same ethnic cluster share the same underlying genetic effect. If we fix the number of clusters as one, all the studies are grouped into one ethnic cluster with homogeneous genetic effects; in this case, MANTRA can be viewed as a Bayesian implementation of the fixed-effects method. If the number of cluster is fixed to be the same as the number of studies (n), each study is assigned to be its own cluster; in this case, MANTRA can be viewed as a Bayesian implementation of the random-effects method. MANTRA uses the Bayesian partition model to adaptively determine the number of ethnic clusters and the cluster membership and assesses the association evidence by means of the Bayes factor.

2.2.2 New Model Framework for GWAS Meta-Analysis

The existing frequentist meta-analysis methods based on (2.2.2) are not optimal when the effect sizes exhibit certain structures across studies. In multi-ethnic meta-analysis, for example, the studies can be grouped by their ethnicities. Genetically similar groups may have more homogeneous genetic effects compared to genetically diverse groups. In response, we propose a statistical framework that can accommodate prior assumptions on genetic effect distributions. Specifically, we adapt the kernel machine framework (Liu et al., 2007) to flexibly model the genetic effect distributions. Instead of assuming η_is are i.i.d normal samples, we assume that η_is jointly follow a mean zero Gaussian process with kernel function $\tau_1 K(\cdot,\cdot)$, where $K(\cdot,\cdot)$ is a bivariate function to represent genetic similarity between two groups. This kernel regression framework has been successfully applied in many areas of genetic studies, including rare variant association analysis (Wu et al., 2011) and pathway analysis (Liu et al., 2007). In Section 2.2.3, we will discuss choices of kernels for trans-ethnic meta-analysis.

We first propose to extend (2.2.2) to a hierarchical model by modeling μ as a random variable with distribution $N(0,\tau_2)$. From this extension, our proposed model framework can be summarized as

$$\begin{aligned} \widehat{\boldsymbol{\beta}}|\boldsymbol{\beta} &\sim MVN(\boldsymbol{\beta},\Sigma) \\ \boldsymbol{\beta}|\tau_1,\tau_2 &\sim MVN(\mathbf{0},\tau_1\mathbf{K}+\tau_2\mathbf{1}\mathbf{1}^T), \end{aligned} \tag{2.2.3}$$

where $\mathbf{K}$ is an $n\times n$ kernel matrix and $\mathbf{1}=(1,\cdots,1)^T$. We then apply a re-parameterization $\tau_1=\tau(1-\rho)$ and $\tau_2=\tau\rho$, where ρ reflects whether genetic effects are homogeneous ($\rho=1$) or heterogeneous ($\rho=0$) across ancestry groups, and τ represents the size of the regression coefficients $\boldsymbol{\beta}$. From this re-parameterization, testing for both μ and τ_1 being zero becomes testing for the common variance component τ being zero. Our final model framework is

$$\begin{aligned} \widehat{\boldsymbol{\beta}}|\boldsymbol{\beta} &\sim MVN(\boldsymbol{\beta},\Sigma) \\ \boldsymbol{\beta}|\tau &\sim MVN(\mathbf{0},\tau V_\rho) \\ V_\rho &= (1-\rho)\mathbf{K}+\rho\mathbf{1}\mathbf{1}^T, \quad 0\le\rho\le 1 \end{aligned} \tag{2.2.4}$$

where V_ρ is an $n\times n$ (scaled) covariance matrix of $\boldsymbol{\beta}$. We note that V_ρ is a linear combination of two matrices, $\mathbf{1}\mathbf{1}^T$ and $\mathbf{K}$, with coefficient ρ that determines the degree of heterogeneity. $\rho = 0$ indicates that the covariance structure of β_is is the same as the kernel matrix $\mathbf{K}$, and $\rho = 1$ indicates that β_is are perfectly correlated (and hence homogeneous).

Our proposed model includes the three frequentist meta-analysis approaches as special cases. For example, if $\rho = 1$ (i.e $V_\rho = \mathbf{1}\mathbf{1}^T$), the model is effectively the same as FE since all β_is should be the same under the alternative hypothesis. We show in Section 2.2.4 that the variance component score test for $\tau = 0$ with $\rho = 1$ is exactly the same as the inverse-variance weighted meta-analysis test, the most popular test for the FE approach. As a result, one of the important features of our model is that it includes FE regardless of the choice of $\mathbf{K}$. We believe this is a desirable feature since numerous disease-associated SNPs in various meta-analysis scenarios including trans-ethnic meta-analysis exhibit homogeneous genetic effects across studies (Marigorta and Navarro, 2013). RE and RE-HE are equivalent to testing for $\tau_2 = 0$ and $\tau_1 = \tau_2 = 0$ under (2.2.3), respectively, with $\mathbf{K} = \mathbf{I}$. This indicates that RE is equivalent to testing for $\rho = 0$, and RE-HE is equivalent to testing for $\tau = 0$ while adaptively selecting ρ under the re-parameterized model (2.2.4) with $\mathbf{K} = \mathbf{I}$.

2.2.3 Choice of the Kernel Matrix K for Trans-Ethnic Meta-Analysis

Suppose the GWAS meta-analysis has B ancestry groups from n studies, based on this assumption, we propose two choices for the kernel structure $\mathbf{K}$:

Choice 1. Group-wise independent kernel structure:

We consider a simple assumption in which genetic effect sizes are independently distributed across ancestry groups, but homogeneous within the same ancestry group.

In particular,

$$K_{ij} = \begin{cases} 1 & \text{if study i and j belong to the same ancestry group} \\ 0 & \text{otherwise} \end{cases},$$

where $i, j \in \{1, \ldots, n\}$. In Supplementary Materials Section 2.5.1, we provide the general form of matrix $\mathbf{K}$ under this group-wise independent structure.

Choice 2. Genetic similarity (F_{st}) kernel structure:

The fixation index (F_{st}) is a widely used measure of population differentiation due to genetic structure (Wright, 1949). $F_{st} = 0$ indicates there is no allele frequency differentiation between populations, whereas a large value of F_{st} indicates that populations are genetically very different. F_{st} has been used as a genetic distance among populations. For example, MANTRA uses F_{st} to group studies to ethnic clusters. For each cluster, it is assumed that studies share the same genetic effect. We adapt the strategy of using F_{st} in constructing the kernel matrix $\mathbf{K}$ to incorporate genetic similarity into modeling the genetic effect similarity. In particular, we set

$$K_{ij} = 1 - \frac{F_{st_{bb'}}}{D}, \quad \text{with } D = \max_{b,b' \in \{1,\ldots,B\}} \{F_{st_{bb'}}\},$$

where study i and j belong to ancestry group b and $b^{'}$ respectively, and $F_{st_{bb'}}$ is the pairwise F_{st} between the corresponding ancestry groups. In Supplementary Materials Section 2.5.1, we provide the general form of $\mathbf{K}$ under this genetic similarity (F_{st}) kernel structure. Unlike MANTRA, which adaptively groups studies based on the prior model of relatedness and observed effect sizes via the Bayesian partition model, our method constructs the genetic similarity (F_{st}) kernel using only the genotype data and fixes it prior to carrying out the data analysis.

2.2.4 Hypothesis Test

Under the proposed model (2.2.4), testing for $H_0 : \beta_1 = \cdots = \beta_n = 0$ is the same as testing for the variance component $\tau = 0$ (i.e. $H_0 : \tau = 0$). We first consider a situation in which ρ is given before carrying out the test. Following Zhang and Lin (2003), the score test statistics of the variance component τ with a given ρ is

$$S_\rho = \widehat{\boldsymbol{\beta}}^T \widehat{\Sigma}^{-1} V_\rho \widehat{\Sigma}^{-1} \widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^T \widehat{\Sigma}^{-1} [(1-\rho)\mathbf{K} + \rho \mathbf{1}\mathbf{1}^T] \widehat{\Sigma}^{-1} \widehat{\boldsymbol{\beta}}, \tag{2.2.5}$$

where $\widehat{\Sigma} = diag(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_n^2)$, and $\widehat{\sigma}_i^2$ is an estimate of σ_i^2. When $\rho = 1$, the test statistic S_ρ becomes $\left(\sum_{i=1}^n \widehat{\beta}_i / \widehat{\sigma}_i^2\right)^2$, which is the test statistics of the inverse variance weighting.

For any given ρ, S_ρ asymptotically follows a mixture of χ^2 distributions under the null hypothesis. Specifically, if $(\lambda_1, \ldots, \lambda_n)$ are the eigenvalues of $\widehat{\Sigma}^{-1/2} V_\rho \widehat{\Sigma}^{-1/2}$, the null distribution of S_ρ can be closely approximated by $\sum_{j=1}^n \lambda_j \chi_{1,j}^2$, where $\{\chi_{1,j}^2\}$ are independent χ_1^2 random variables. Several methods exist to obtain tail probabilities of the mixture of χ^2 distributions. Among them, the method to invert a characteristic function (Davies, 1980) provides very accurate estimates of tail probabilities and is widely used in many recently developed genetic association tests (Wu et al., 2011). We employ this approach to approximate the asymptotic distribution of S_ρ when ρ is given.

In practice, however, we rarely have prior information on which ρ is optimal in terms of maximizing power. Lee et al. (2012) have studied a similar problem within a context of rare variant association analysis; they proposed to use the minimum p-value over a grid of ρ as a test statistics. We adopt the same strategy here. Specifically, we set the test statistic as $T = \underset{0 \le \rho \le 1}{inf}\, p_\rho$, where p_ρ is the corresponding p-value of S_ρ for the given ρ. T can be obtained by a simple grid search across a range of ρ values: set a grid $0 \le \rho_1 \le \rho_2 \le \ldots \le \rho_\nu \le 1$, then the test statistic becomes

$$T = \min\{p_{\rho_1}, \ldots, p_{\rho_\nu}\},$$

and the optimal ρ is set as the one whose corresponding p-value (p_ρ) equals to T. We observe that a dense grid of ρ does not necessarily improve power (Supplementary Figure 2.10). Therefore, we suggest using $\rho = (0, 0.3^2, 0.5^2, 1)$ for simulations and real data analysis. Once the test statistic T is calculated, the next step is to obtain the corresponding p-value for ascertaining the association evidence. If we had just used the minimum p-value (which is denoted as our test statistic T) to assess significance, we would ignore the multiple comparisons between different p_ρ values, which would result in inflated type I error control. Thus, we propose to use numerical integration to approximate the asymptotic distribution of T, details provided in Supplementary Materials Section 2.5.2.

2.2.5 Using Z-scores Instead of Effect-size Estimates

In previous sections, we construct our methods based on estimates of effect sizes and their standard errors. However, Z-score based approaches are also very popular in GWAS. Z-score based approaches use p-values (p_i), sample sizes (n_i) and direction of effects (Δ_i) to construct Z-scores for each study, and then calculate a weighted sum of Z-scores to carry out meta-analysis. A major advantage of the Z-score based approach is that it allows meta-analysis of data when effect size estimates are not available or measurements of traits are difficult to standardize, ex. tobacco or alcohol use (Evangelou and Ioannidis, 2013). In this section, we extend TransMeta to incorporate Z-score based data input.

Given the input summary statistics (p_i, n_i, Δ_i), a signed Z-score is constructed as $Z_i = \Phi^{-1}(1 - p_i/2) * sign(\Delta_i)$ for each study, where $\Phi(\cdot)$ is the standard normal distribution function. For continuous traits, it can be shown that the effect size estimate $\widehat{\beta}_i$ is asymptotically equivalent to $Z_i/\sqrt{n_i q_i(1 - q_i)}$ (up to a scalar factor), where q_i is a minor allele frequency (MAF) of the SNP (details provided in Supplementary Materials Section 2.5.3). For binary traits, the log odds ratio estimate $\widehat{\beta}_i$ is asymptotically equivalent to $Z_i/\sqrt{n_i r_i(1 - r_i) q_i(1 - q_i)}$, where $r_i = n_{case,i}/n_i$ is a proportion of case samples (Supplementary Materials Section 2.5.3). If all studies have similar ratios of cases and controls, the $r_i(1 - r_i)$ term can thus be ignored. Consequently,

$\tilde{\beta}_i = Z_i/\sqrt{n_i q_i(1-q_i)}$ and its standard error $\tilde{\sigma}_i = 1/\sqrt{n_i q_i(1-q_i)}$ can be used as inputs for both continuous and binary traits. To differentiate between the two types of input summary statistics, if effect size estimates $\widehat{\beta}_i$s and the corresponding standard errors $\widehat{\sigma}_i$s are used as input data for our proposed method, we denote it as effect-size based TransMeta; if transformed Z-scores $\tilde{\beta}_i$s and the corresponding standard errors $\tilde{\sigma}_i$s are used as input data for our proposed method, we denote it as Z-score based TransMeta.

2.3 Results

2.3.1 Simulation Studies

To investigate the performance of TransMeta, we ran a series of simulations with varying assumptions on genetic effect heterogeneity across multiple ancestry groups. To generate SNPs with realistic MAF spectrums across different ancestry groups, we used Phase III of the HapMap Project (HMP3) data (Consortium et al., 2010). HMP3 consists of approximately 1.6 million SNPs, obtained from 1,184 subjects from 11 populations. We excluded the admixed African American population, combined the Japanese and Chinese as one population, and used the resulting 9 populations as seed populations to generate SNP genotypes.

The retrospective binary phenotype Y_{ik} of the kth individual in the ith study was generated using the following logistic regression model

$$\text{logit } Pr(Y_{ik} = 1) = \beta_0 + \beta_i g_{ik}, \tag{2.3.1}$$

where g_{ik} is a genotype of the selected SNP, and β_i is a log odds ratio parameter. The intercept β_0 was chosen to have disease prevalence 0.05. In each replication, we randomly chose a SNP with a MAF of at least 1% in all populations, and generated SNP genotypes as $g_{ik} \sim Binomial(2, q_i)$, where q_i denotes the MAF of the selected SNP. We also used model (2.3.1) to estimate log odds ratio $\widehat{\beta}_i$ and its standard error

Table 2.1: **Type-I error rate estimates for TransMeta at different α levels, with three studies in each ancestry group.** Type-I error rate estimates at different α levels based on 20 million replicates. Each entry represents an estimated type I error rate calculated using the proportion of empirical p-values smaller than the given level α. Three studies are simulated per ancestry group, and each study had 500 cases and 500 controls.

	$\alpha = 10^{-2}$	10^{-3}	10^{-4}	10^{-5}	10^{-6}
TransMeta.Fst	9.7×10^{-3}	1.1×10^{-3}	9.6×10^{-5}	1.0×10^{-5}	9.5×10^{-7}
TransMeta.Indep	9.8×10^{-3}	0.9×10^{-3}	7.6×10^{-5}	5.8×10^{-6}	4.0×10^{-7}

$\widehat{\sigma}_i$ as the input data. In addition, we recorded $\Delta_i = sign(\widehat{\beta}_i)$, the direction of effect and the p-value p_i for testing $H_0 : \beta_i = 0$. We generated 500 cases and 500 controls for each of the 9 ancestry groups in triplicate, which resulted in a total of 27 studies with a total sample size of 13,500 cases and 13,500 controls.

2.3.2 Type I Error Simulations

To estimate type I error rates at stringent α levels, we generated 20 million replicates from model (2.3.1) with $\beta_i = 0$. Table 2.1 showed that the proposed methods yields controlled type I error rates at different significance levels under the F_{st} kernel (denoted as TransMeta.Fst), although slightly conservative under the independent kernel (denoted as TransMeta.Indep). We also considered a setting where there is only one study per ancestry group. Each study then had 1500 cases and 1500 controls. We again used model (2.3.1) with $\beta_i = 0$ to simulate a total of 100 million replicates, and observed that empirical type I error rates were well controlled (Table 2.2 in the Supplementary Materials).

2.3.3 Power Simulations

Recently, Wang et al. (2013) carried out comparisons of trans-ethnic meta-analysis methods under five different scenarios, which cover a wide range of possible scenarios of genetic effect heterogeneity. We adopted these five scenarios to compare performances of TransMeta with existing approaches:

(a) ‘Trans-ethnic fixed-effect’, where no heterogeneity exists in genetic effects at the causal SNP between populations, specifically that, each of the 27 studies carries a genetic relative risk of 1.12 at the causal SNP.

(b) ‘Out-of-Africa effect’, where each of the 18 studies from the non-African populations carries a genetic relative risk of 1.08, whereas the 9 studies from the African populations (LWK, MKK and YRI) present no genetic effects.

(c) ‘Europe and south Asia effect’, where the 12 studies from the European and south Asian populations (CEU, GIH, MEX and TSI) share the same genetic relative risk of 1.2, whereas the 15 studies from the remaining populations present no genetic effects.

(d) ‘Heterogeneous Out-of-Africa effect’, where the causal variant has genetic effects only in non-African populations, with the 6 studies from the east Asian populations (CHB+JPT and CHD) each carrying a genetic relative risk of 1.15 while the European and south Asian populations carry a genetic relative risk of 1.12.

(e) ‘Environment modifying effect’, where the causal variant has a genetic effect only in the populations living in Europe and USA, with the 9 studies from CHD, CEU and TSI each carry a genetic relative risk of 1.2.

In all scenarios, causal SNPs had the same direction of associations across ancestry groups. For each scenario, we generated 2,000 replicates to obtain empirical power. To perform a fair comparison between the frequentist and Bayesian methods, we generated 20 million SNPs under the null hypothesis and compute Bayes factors using MANTRA. We observed that a log10 Bayes factor threshold larger than 5 corresponds to a p-value threshold less than $\alpha = 1.8 \times 10^{-6}$. To find a log10 Bayes factor threshold corresponding to the genome-wide significance level, we carried out a simple regression analysis between empirical type I error rates and log10 Bayes factors, and observed that log10 Bayes factor= 6.34 corresponds to $\alpha = 5 \times 10^{-8}$ (see

Supplementary Materials Section 2.5.5 for details).

Figure 2.1 showed the empirical power of TransMeta as well as existing approaches (FE, RE, RE-HE and MANTRA) under all five scenarios. It can be seen that TransMeta.Fst yielded the highest or near highest power among the five methods, except in scenario (e). In scenario (a) where no heterogeneity exists, all five methods performed similarly, with FE having the highest power, as expected. In the remaining three scenarios with heterogeneous genetic effects that are not due to the environment modification, TransMeta.Fst outperformed the four existing meta-analysis methods. Unsurprisingly, RE yielded the lowest power across all five approaches. In scenario (e) where the genetic effect was influenced by environmental exposures, populations that shared similar genetic architectures did not necessarily share similar genetic effects. This violated the assumption of using the F_{st} to take account of the variability in genetic effects, and in this case, TransMeta.Indep yielded the highest power.

Figure 2.2 showed the empirical power of the five methods with one integrated study per ancestry group. The patterns of empirical power in this setting were very similar to what we observe in Figure 2.1 with three sub-studies per ancestry group, except for RE-HE, which had slightly higher power than that of TransMeta.Indep. Since TransMeta.Indep has the identity matrix as the kernel structure (i.e $\mathbf{K} = \mathbf{I}$) under this setting, the similar performance of TransMeta.Indep and RE-HE is not surprising. Overall, TransMeta.Fst attained similar or higher power over competing methods except in scenario (e).

The barplots in Figure 2.7 and 2.8 of the Supplementary Materials summarized the power of the five methods at the more stringent level $\alpha = 5 \times 10^{-8}$; the results were quantitatively similar to the patterns we observe in Figure 2.1 and 2.2.

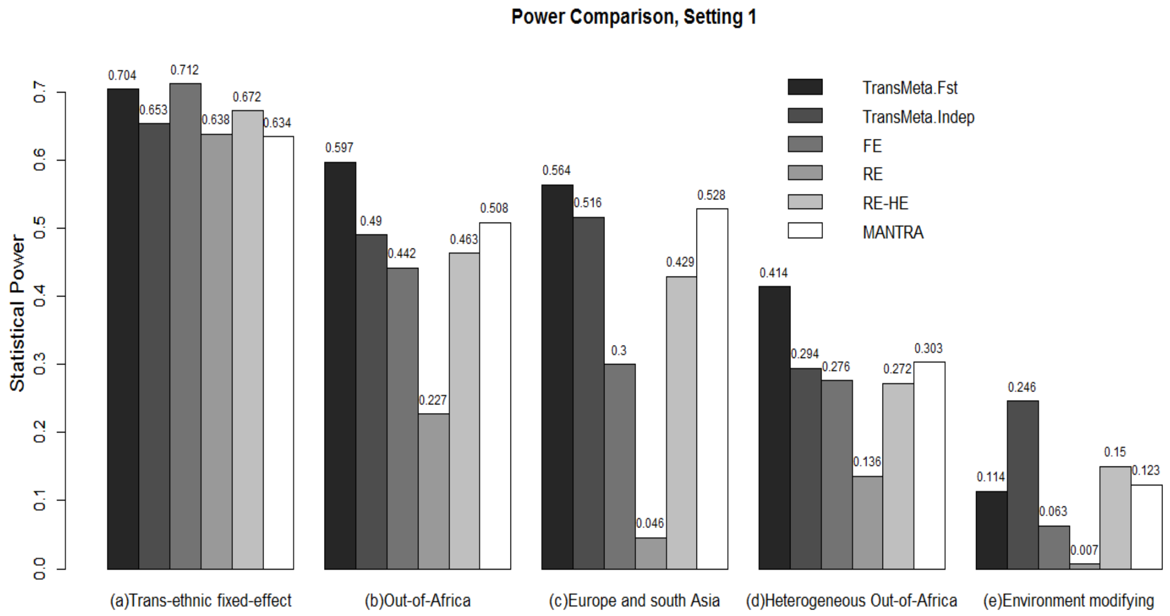


Figure 2.1: **Empirical power for TransMeta under different effect-size heterogeneity configurations, with three sub-studies in each ancestry group, and significance level at $\alpha = 1.8 \times 10^{-6}$.** Empirical power for TransMeta and existing methods under the five effect-size scenarios. Three studies are simulated per ancestry group, each with 500 cases and 500 controls. The empirical power is obtained based on 2000 replicates with the level of significance defined as a p-value less than 1.8×10^{-6} or as a log10 Bayes factor larger than 5. The five-effect size scenarios are (a) 'Trans-ethnic fixed-effect', where no heterogeneity exists in allelic effects at the causal SNP between populations; (b) 'Out-of-Africa effect', where only studies from the non-African populations carry the causal variant; (c) 'Europe and south Asia effect', where only studies from the European and south Asian populations carry the causal variant; (d) 'Heterogeneous Out-of-Africa effect', where the causal variant has genetic effects only in non-African populations, but the effect size in the east Asian populations is different from that in the European and south Asian populations; (e) 'Environment modifying effect', where the causal variant has genetic effect only in the populations living in Europe and USA.

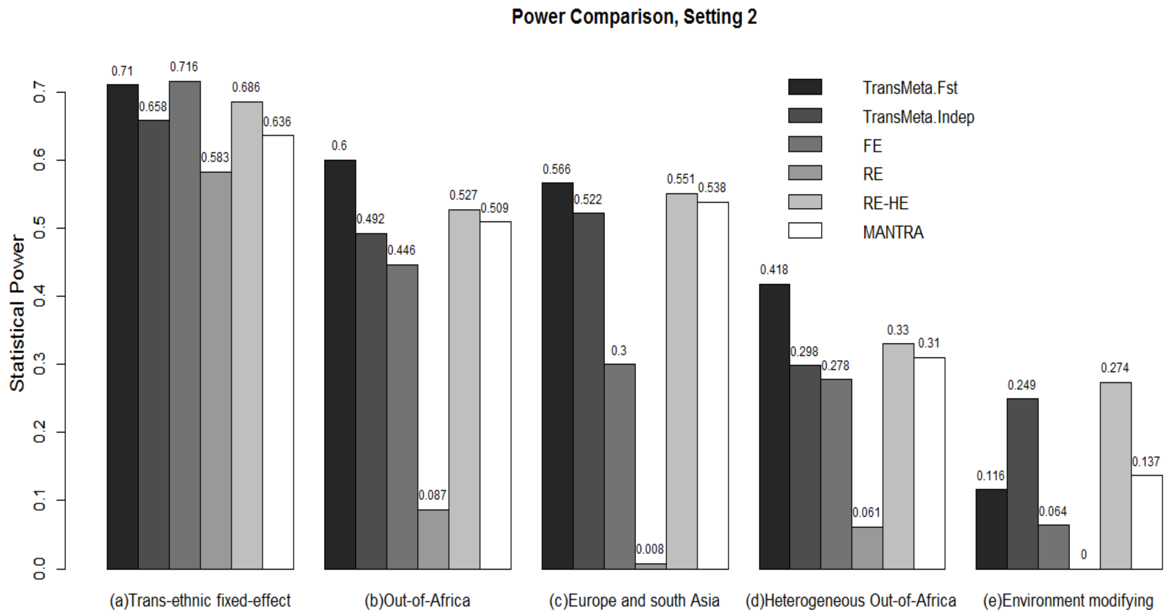


Figure 2.2: **Empirical power for TransMeta under different effect-size heterogeneity configurations, with one integrated study in each ancestry group, and significance level at $\alpha = 1.8 \times 10^{-6}$.** Empirical power for TransMeta and existing methods under the five effect-size scenarios. One integrated study is simulated per ancestry group, each with 1500 cases and 1500 controls. The empirical power is obtained based on 2000 replicates with the level of significance defined as a p-value less than 1.8×10^{-6} or as a log10 Bayes factor larger than 5. The five effect-size scenarios are (a) 'Trans-ethnic fixed-effect', where no heterogeneity exists in allelic effects at the causal SNP between populations; (b) 'Out-of-Africa effect', where only studies from the non-African populations carry the causal variant; (c) 'Europe and south Asia effect', where only studies from the European and south Asian populations carry the causal variant; (d) 'Heterogeneous Out-of-Africa effect', where the causal variant has genetic effects only in non-African populations, but the effect size in the east Asian populations is different from that in the European and south Asian populations; (e) 'Environment modifying effect', where the causal variant has genetic effect only in the populations living in Europe and USA.

2.3.4 Comparison Between Effect-size-based and Z-score-based TransMeta

To demonstrate that Z-scores can be used as input summary statistics for TransMeta without loss of efficiency, we compared the power of the effect-size based and Z-score based TransMeta. Since the proportion of case samples was one (i.e $r_i = 1$) for all studies, we ignored r_i in the transformation. We also considered using only the transformed Z-scores and sample sizes as the input, which is equivalent to assume that MAFs of SNPs are the same across all studies. In this case, the transformation simplifies to $\tilde{\beta}_i = Z_i/\sqrt{n_i}$ with standard error $\tilde{se}_i = 1/\sqrt{n_i}$. We included this setting because Z-scores are typically obtained without MAFs.

The scatter plot in Figure 2.3 compared the power of the effect-size-based and the Z-score-based TransMeta under the five scenarios as outlined in Section 2.3.3. The plot was generated under the settings where we had three sub-studies per ethnic group, with the level of significance as a p-value less than 1.8×10^{-6}. The power of these two approaches was nearly identical when we incorporate both sample sizes and MAFs in the Z-score transformations, and the power of the Z-score based TransMeta was slightly lower than the effect-size based TransMeta when only sample sizes are used in the Z-score transformations. For the one integrated study per ancestry group setting, the results were quantitatively similar to the patterns in Figure 2.3 (Figure 2.9 in the Supplementary Materials). At the genome-wide significance level, we again observed similar patterns as in Figure 2.3 and Supplementary Figure 2.9 (data not shown).

2.3.5 Computation Time

TransMeta provides scalable computation time for genome-wide datasets. To analyze 2,000 SNPs in the power simulations, both TransMeta.Fst and TransMeta.Indep took 20 seconds on average on a Linux cluster node with 2.80 GHz CPU. To analyze one million SNPs in a genome-wide dataset, TransMeta would require less

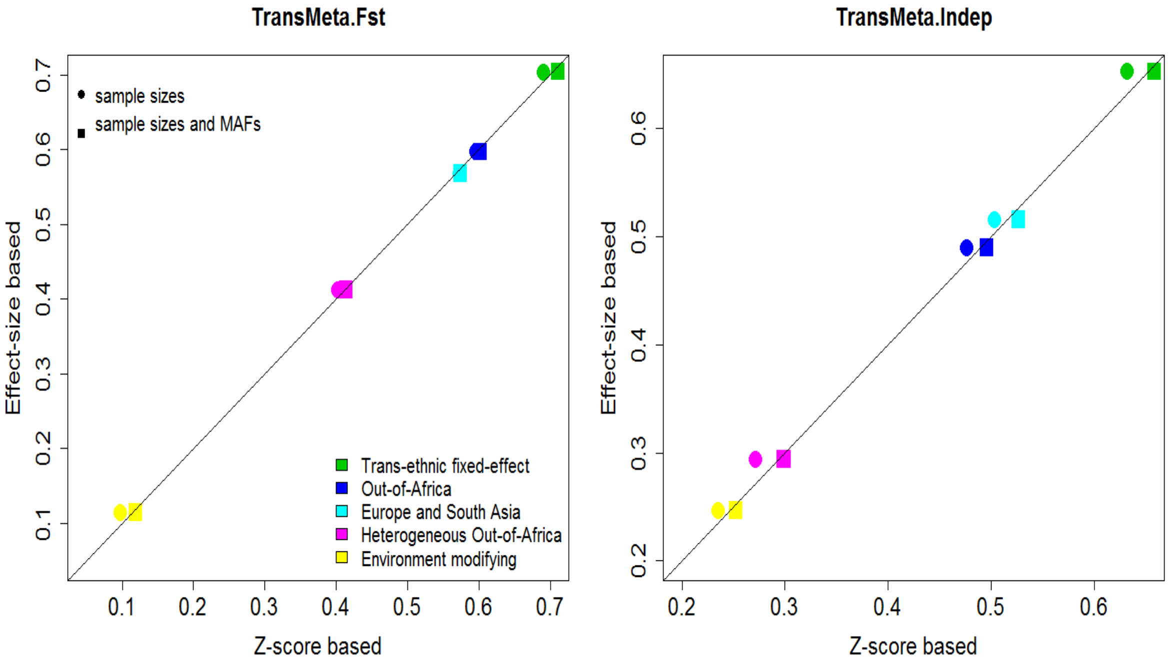


Figure 2.3: **Power comparison of the effect-size-based and Z-score-based TransMeta, with three sub-studies in each ancestry group, and significance level at $\alpha = 1.8 \times 10^{-6}$.** Power comparison of the effect-size-based and Z-score-based TransMeta under the five effect size scenarios. Three studies are simulated per ancestry group, each with 500 cases and 500 controls. The empirical power is obtained based on 2000 replicates with the level of significance defined as a p-value less than 1.8×10^{-6}. The left panel is based on TransMeta.Fst and the right panel is based on TransMeta.Indep. In each plot, the x-axis denotes empirical power of the the Z-score-based TransMeta and the y-axis denotes empirical power of effect-size-based TransMeta. The solid dots represent the power of transformed Z-scores using only sample sizes, and the solid squares represent transformed Z-scores using both sample sizes and MAFs.

than 3 hours. Among the competing methods, MANTRA was computationally expensive and took 45 and 95 minutes on average to analyze 2,000 SNPs with 9 and 27 studies, respectively. An R package 'TransMeta' has been developed to implement our proposed method and can be downloaded at the authors' website (https://sites.google.com/a/umich.edu/leeshawn/software).

2.3.6 Application to Type 2 Diabetes (T2D) GWAS

Large scale GWAS of T2D have successfully identified many risk-associated loci, including a landmark meta-analysis on T2D by the DIAGRAM consortium with over 110,000 genotyped individuals (Mahajan et al., 2014). Most of those studies have applied one or a combination of the classical FE or RE meta-analysis approaches, with limited use of the more powerful RE-HE or MANTRA methods. In this section, we re-analyzed the published T2D GWAS meta-analysis (Mahajan et al., 2014), in which the FE was employed in the trans-ethnic discovery-stage GWAS meta-analysis. The aggregated data included 69 lead SNPs from the previously established T2D susceptibility loci, with 26,488 cases and 83,964 controls from four major ancestry groups of Europeans (12,171 cases and 56,862 controls), east Asians (6,952 cases and 11,865 controls), south Asians (5,561 cases and 14,458 controls), and Mexican and Mexican-Americans (1,804 cases and 779 controls). Association summary statistics – such as MAFs, effect size estimates, and standard errors – of the lead 69 SNPs were available for all four ancestry groups (Supplementary Table 3 of Mahajan et al. (2014)).

We applied TransMeta to the aggregated data along with the other existing meta-analysis approaches. Due to the small number of SNPs in the aggregated dataset, estimates of F_{st} may be unreliable. Instead, we used the pairwise F_{st} from HMP3 to construct the genetic similarity kernel (Table 2.3 in the Supplementary Materials). Supplementary Tables 2.4 and 2.5 listed p-values (or Bayes factors) of the 69 SNPs

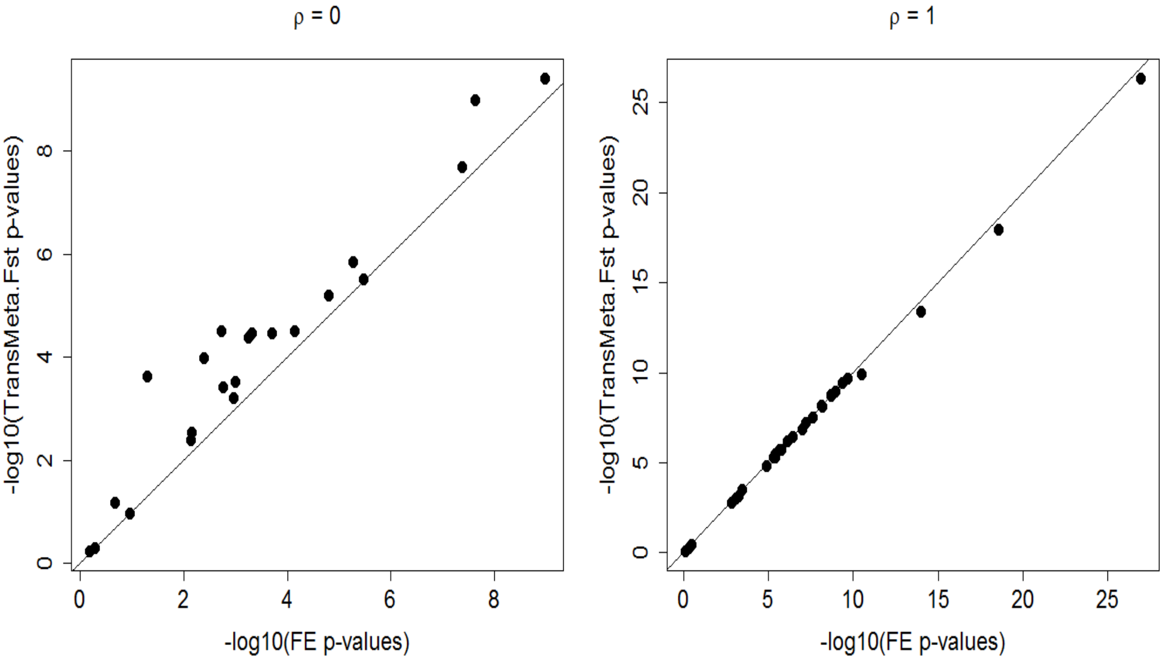


Figure 2.4: **Comparison of p-values of TransMeta.Fst and FE for 69 lead SNPs in the T2D meta-analysis data.** The left panel displays p-values of SNPs whose TransMeta.Fst ρ is zero; the right panel displays p-values of SNPs whose TransMeta.Fst ρ is one. In each plot, the x-axis denotes $-\log_{10}$(FE p-values), and the y-axis denotes $-\log_{10}$(TransMeta.Fst p-values).

with selected optimal ρs of TransMeta. Among those 69 SNPs, 37 had optimal $\rho < 1$ under TransMeta.Fst. Figure 2.4 compares p-values of TransMeta.Fst and FE for different selected optimal ρs. When the selected optimal $\rho = 0$, our method yielded a smaller p-value than FE, which indicated that TransMeta can be more powerful than FE. When the selected $\rho = 1$, and hence FE was the optimal test, FE yielded a smaller p-value than TransMeta, but the difference was minimal.

At the significance level $\alpha = 1.8\times10^{-6}$ or a log10 Bayes factor > 5, TransMeta.Fst, TransMeta.Indep, FE and RE-HE all identified 31 SNPs, while RE and MANTRA identified 18 and 28 SNPs, respectively. At the genome-wide significance level of $\alpha = 5\times10^{-8}$ or a log10 Bayes factor > 6.34, both TransMeta.Fst and TransMeta.Indep identified 24 SNPs, while FE, RE, RE-HE and MANTRA identified 23, 12, 22 and

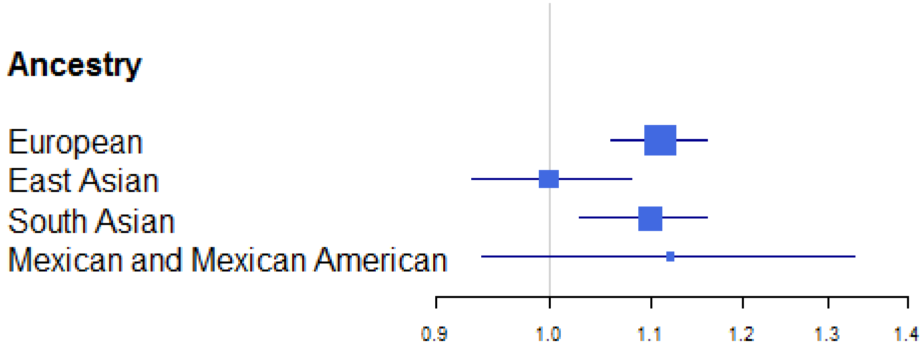


Figure 2.5: **Forest plot of the estimated OR and 95 % CI for rs10830963 in each ancestry group in the T2D meta-analysis data.** The association signal of rs10830963 is detected by TransMeta only.

19 SNPs respectively.

At the genome-wide significance level, TransMeta was able to identify one more SNP, rs10830963, with TransMeta.Fst p-value= 2.98×10^{-8} (selected optimal $\rho = 0.25$) and TransMeta.Indep p-value=3.76×10^{-8} (selected optimal $\rho = 0.25$), respectively. In contrast, p-values of FE, RE and RE-HE were all larger than 10^{-7}, and MANTRA log10 Bayes factor was 5.6. The SNP rs10830963 is located in Melatonin receptor 1-B, which belongs to the seven transmembrane G protein-coupled receptor superfamily, and previous studies have shown that this SNP is associated with fasting glycemia and T2D (Rönn et al., 2009; Sparsø et al., 2009; Kan et al., 2010; Vlassi et al., 2012).

Figure 2.5 displayed a forest plot of odds ratios and their corresponding confidence intervals for this SNP (extracted from Supplementary Table 3 in Mahajan et al. (2014)). The odds ratios of Europeans, south Asians and Mexicans were all close to 1.1, although the odds ratio for Mexicans was non-significant due to small sample size. In contrast, the odds ratio in east Asians was close to one. Since east Asians are genetically more distant than other populations (Table 2.3 in the Supplementary Materials), this result indicated that our approach to modeling genetic effect heterogeneity using genetic distance can increase power.

2.4 Discussion

In this chapter, we proposed a novel trans-ethnic meta-analysis framework to flexibly model the genetic effect heterogeneity across ancestry groups. The framework incorporates the genetic distances to model the genetic effect heterogeneity and adaptively uses variance component test to achieve robust power. Simulations and the trans-ethnic T2D GWAS application suggest that our approach can improve power when genetic effect-size heterogeneity exists.

Since TransMeta.Fst accommodates genetic similarity to model the effect size similarity, we recommend TransMeta.Fst as the primary test. However, if there is evidence suggesting that the genetic effects are modified by non-genetic exposures (such as environmental or lifestyle factors), TransMeta.Indep may be a better choice. To avoid data fishing, the choice of using TransMeta.Fst or TransMeta.Indep needs to be made prior to data analysis. For the sequence of ρ values used in the grid search, we observe that using a dense grid of ρs does not necessarily increase power. In fact, in Supplementary Figure 2.10 , we employ a denser grid with eleven evenly spaced points of $\rho = (0, 0.1, \ldots, 0.9, 1)$ in the power simulations and observe that the resulting power is very similar or even identical to the power based on $\rho =$ (0, 0.09, 0.25, 1). So we suggest using $\rho =$ (0, 0.09, 0.25, 1) as the default sequence of ρ values. We note that it is not required to select ρ from the grid prior to perform the analysis, since TransMeta automatically selects the optimal ρ, and calculate p-values while accounting for the selection.

Unlike the I^2 statistic (Higgins et al., 2002), which is developed to measure the extent of heterogeneity, the optimal ρ is set as the value (over a pre-specified grid) whose score statistic has the smallest p-value among all. As a result, the optimal ρ should not be interpreted as a measurement of heterogeneity. For example, we count the number of optimal ρ values in each of the five scenarios in the power simulations (Table 2.7 in the Supplementary Materials) and observe that in the homogeneous effect size scenario, only less than half of the optimal ρ values in TransMeta.Fst are determined to be 1. (Please recall that ρ equals to 1 models homogeneous effect sizes; the closer ρ is to 0, the stronger the indication of heterogeneity.) However, the optimal ρ does provide some insights into the extent of heterogeneity. For example, in our power simulations, we observe that the I^2 statistic tends to decrease as the optimal ρ increases, as shown in Supplementary Table 2.6. (Please recall that $I^2 = 0$ means homogeneity; and the level of heterogeneity increases as I^2 approaches to 1.) In addition, we observe in Supplementary Table 2.7 that when heterogeneity does exist, such as scenarios (b) - (e) in the power simulations, the majority of the optimal ρ values in TransMeta.Fst are selected to be 0. Similar trends are observed in TransMeta.Indep, data not shown.

Our score statistics S_ρ is a linear combination of two components, each models the genetic effect homogeneity and the genetic effect heterogeneity, respectively. As a result, although TransMeta is designed to tackle heterogeneous effect sizes situations, it can also handle homogeneity scenarios. In fact, the right panel of Figure 2.4 demonstrates that under genetic effect homogeneity, our approach achieves almost the same statistical significance as FE.

We note that the empirical power of MANTRA is similar or lower than that of TransMeta.Fst in scenarios (b)-(d), but is higher in scenario (e)(Figure 2.1 and 2.2). This occurred because under scenario (e), the genetic distance does not provide guidance to the genetic effect similarity, which violates the key assumption in both MANTRA and TransMeta.Fst. Since MANTRA groups studies into clusters data-adaptively, it is more robust than TransMeta.Fst under this situation. As a result, MANTRA had higher power than TransMeta.Fst.

RE-HE is equivalent to testing for $\tau = 0$ while adaptively selecting ρ under model (2.2.4) with $\mathbf{K} = \mathbf{I}$. When we have one integrated study per ancestry group, the $\mathbf{K}$ matrix in TransMeta.Indep is exactly equal to $\mathbf{I}$, which makes RE-HE equivalent to

TransMeta.Indep in terms of testing (although they use different approaches to obtain p-values). As a result, RE-HE and TransMeta.Indep have similar power in all five scenarios in Figure 2.2. When we have three studies per ancestry group (Figure 2.1), RE-HE treats each study as its own cluster. In contrast, TransMeta.Indep groups studies in the same ancestry. As a result, TransMeta.Indep yields higher power than RE-HE in nearly all scenarios in Figure 2.1.

Our proposed method is based on the score test which does not require estimating parameters under the alternative hypothesis. Score test enables fast computation of p-values, however, it does not provide an estimate for the overall effect size. Estimating the overall effect size under our proposed model framework may be considered as one possible direction for future work.

The rapid technological advances in high-throughput sequencing platforms have made it possible to test for rare variant associations (here defined as alleles with a frequency less than 1%) to accelerate our knowledge of complex trait genetics. One of the challenge in the design and analysis of sequencing-based GWAS is that rare variant tests are usually underpowered without exceptionally large sample size or large enough number of rare alleles captured (Bansal et al., 2010). A popular strategy to boost the study power is to combine rare variants based on a gene or a region to bring a synergy of information (Lee et al., 2013; Tang and Lin, 2014; Liu et al., 2014). In the next chapter, we will extend the framework of using genetic distance for modeling the genetic effect heterogeneity to gene/region based rare variant tests in trans-ethnic meta-analysis.

2.5 Supplementary Materials

2.5.1 Kernel Matrix K for Trans-Ethnic Meta-Analysis

In Section 2.2.3, we propose two choices of **K** for trans-ethnic meta-analysis. In this section, to help better understand the two proposed kernel matrix **K**, we provide more details on how to construct those two kernels. Suppose that the first n_1 studies belong to the first ancestry group, followed by another n_2 studies belonging to the second ancestry group, and so on. With B ancestry groups among the n studies, we have $n_1 + n_2 + \ldots + n_B = n$. Let A_b denote a set of indices for the studies in the bth ancestry group, $b = 1, \ldots, B$. According to the order of arranging the studies, we thus have $A_1 = \{1, \ldots, n_1\}, A_2 = \{n_1 + 1, \ldots, n_1 + n_2\}, \ldots, A_b = \{n_1 + \ldots + n_{b-1} + 1, \ldots, n_1 + \ldots + n_{b-1} + n_b\}$. Based on those notations, we propose two choices for **K**:

Choice 1. Group-wise independent kernel structure

Define each entry of the **K** matrix as

$$K_{ij} = \begin{cases} 1 & \text{if } i, j \in A_b \text{ for some b} \in \{1, \ldots, B\} \\ 0 & \text{otherwise} \end{cases},$$

where $i, j \in \{1, \dots, n\}$. The $\mathbf{K}$ matrix can be written as

$$\mathbf{K} = \left(\begin{array}{ccc|ccc|c|ccc}
1 & \dots & 1 & 0 & \dots & 0 & & 0 & \dots & 0 \\
 & \vdots & & & \vdots & & \ddots & & \vdots & \\
1 & \dots & 1 & 0 & \dots & 0 & & 0 & \dots & 0 \\
\hline
0 & \dots & 0 & 1 & \dots & 1 & & 0 & \dots & 0 \\
 & \vdots & & & \vdots & & \ddots & & \vdots & \\
0 & \dots & 0 & 1 & \dots & 1 & & 0 & \dots & 0 \\
\hline
 & \ddots & & & \ddots & & \ddots & & \ddots & \\
\hline
0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1 \\
 & \vdots & & & \vdots & & \vdots & & \vdots & \\
0 & \dots & 0 & 0 & \dots & 0 & \dots & 1 & \dots & 1
\end{array}\right).$$

2. Genetic similarity (F_{st}) kernel structure

Define each entry of the $\mathbf{K}$ matrix as

$$K_{ij} = 1 - \frac{F_{st_{bb'}}}{D}, \quad \text{with } D = \max_{b,b' \in \{1,\dots,B\}} \{F_{st_{bb'}}\},$$

where $i \in A_b$ for some $b \in \{1, \dots, B\}$, $j \in A_{b'}$ for some $b^{'} \in \{1, \dots, B\}$, and $F_{st_{bb'}}$ is the pairwise F_{st} between ancestry group b and $b^{'}$. Since $F_{st_{bb'}} \leq D, \forall\ t$ and $t^{'}$, as a consequence, $0 \leq K_{ij} \leq 1, \forall\ i$ and j. In general, the $\mathbf{K}$ matrix under this assumption can be written as

$$\mathbf{K} = \left(\begin{array}{ccc|ccc|c|ccc}
1 & \dots & 1 & 1-\frac{F_{st_{12}}}{D} & \dots & 1-\frac{F_{st_{12}}}{D} & & 1-\frac{F_{st_{1B}}}{D} & \dots & 1-\frac{F_{st_{1B}}}{D} \\
 & \vdots & & & \vdots & & \ddots & & \vdots & \\
1 & \dots & 1 & 1-\frac{F_{st_{12}}}{D} & \dots & 1-\frac{F_{st_{12}}}{D} & & 1-\frac{F_{st_{1B}}}{D} & \dots & 1-\frac{F_{st_{1B}}}{D} \\
\hline
1-\frac{F_{st_{21}}}{D} & \dots & 1-\frac{F_{st_{21}}}{D} & 1 & \dots & 1 & & 1-\frac{F_{st_{2B}}}{D} & \dots & 1-\frac{F_{st_{2B}}}{D} \\
 & \vdots & & & \vdots & & \ddots & & \vdots & \\
1-\frac{F_{st_{21}}}{D} & \dots & 1-\frac{F_{st_{21}}}{D} & 1 & \dots & 1 & & 1-\frac{F_{st_{2B}}}{D} & \dots & 1-\frac{F_{st_{2B}}}{D} \\
\hline
 & \ddots & & & \ddots & & \ddots & & \ddots & \\
\hline
1-\frac{F_{st_{B1}}}{D} & \dots & 1-\frac{F_{st_{B1}}}{D} & 1-\frac{F_{st_{B2}}}{D} & \dots & 1-\frac{F_{st_{B2}}}{D} & & 1 & \dots & 1 \\
 & \vdots & & & \vdots & & \ddots & & \vdots & \\
1-\frac{F_{st_{B1}}}{D} & \dots & 1-\frac{F_{st_{B1}}}{D} & 1-\frac{F_{st_{B2}}}{D} & \dots & 1-\frac{F_{st_{B2}}}{D} & & 1 & \dots & 1
\end{array}\right).$$

2.5.2 Derivation of the Asymptotics

After computing the test statistic T, the next step is to obtain the corresponding p-value for ascertaining the association evidence. If we had just used the minimum p-value (which is denoted as our test statistic T) to assess the significance, we would ignore the multiple comparisons between different p_ρ values, which then leads to an inflated type I error control. Thus, we derive the asymptotic distribution of T to obtain its p-value, details provided as follows:

Recall that the score test statistics can be written as:

$$S_\rho = (1-\rho)\widehat{\boldsymbol{\beta}}^T \widehat{\Sigma}^{-1}\mathbf{K}\widehat{\Sigma}^{-1}\widehat{\boldsymbol{\beta}} + \rho\widehat{\boldsymbol{\beta}}^T \widehat{\Sigma}^{-1}\mathbf{1}\mathbf{1}^T\widehat{\Sigma}^{-1}\widehat{\boldsymbol{\beta}}. \tag{2.5.1}$$

And for any given ρ, the null distribution of S_ρ asypototically follows

$$\sum_{j=1}^{n} \lambda_j \chi^2_{1,j}, \tag{2.5.2}$$

where $(\lambda_1, \ldots, \lambda_n)$ are the eigenvalues of $\widehat{\Sigma}^{-1/2}V_\rho\widehat{\Sigma}^{-1/2}$, and $\{\chi^2_{1,j}\}$ are independent χ^2_1 random variables.

Let $\mathbf{u} = \widehat{\Sigma}^{-1/2}\widehat{\boldsymbol{\beta}}$, $\mathbf{Z} = \widehat{\Sigma}^{-1/2}\mathbf{1}$ and $\mathbf{M} = \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$, then $\mathbf{M}$ is a projection matrix onto the space spanned by $\mathbf{Z}$. Based on those notations, the first term of the right side of (2.5.1) can be written as:

$$\begin{aligned}
(1-\rho)\widehat{\boldsymbol{\beta}}^T\widehat{\Sigma}^{-1}\mathbf{K}\widehat{\Sigma}^{-1}\widehat{\boldsymbol{\beta}} &= (1-\rho)\mathbf{u}^T\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}\mathbf{u} \\
&= (1-\rho)\mathbf{u}^T(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{M})\mathbf{u} \quad (2.5.3) \\
&+ 2(1-\rho)\mathbf{u}^T(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}\mathbf{M}\mathbf{u} \quad (2.5.4) \\
&+ (1-\rho)\mathbf{u}^T\mathbf{M}\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}\mathbf{M}\mathbf{u}, \quad (2.5.5)
\end{aligned}$$

and the second term of the right side of (2.5.1) can be written as:

$$\begin{aligned}
\rho\widehat{\boldsymbol{\beta}}^T\widehat{\Sigma}^{-1}\mathbf{1}\mathbf{1}^T\widehat{\Sigma}^{-1}\widehat{\boldsymbol{\beta}} &= \rho\mathbf{u}^T\widehat{\Sigma}^{-1/2}\mathbf{1}\mathbf{1}^T\widehat{\Sigma}^{-1/2}\mathbf{u} \\
&= \rho\mathbf{u}^T\mathbf{M}\mathbf{Z}\mathbf{Z}^T\mathbf{M}\mathbf{u}. \quad (2.5.6)
\end{aligned}$$

Following the derivation as in Lee et al. (2012), it can be easily shown that (2.5.3) + (2.5.4) = $(1-\rho)\kappa$ and (2.5.5) + (2.5.6) = $\tau(\rho)\eta_0$, where

$$\begin{aligned}
\kappa &= \mathbf{u}^T(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{M})\mathbf{u} \\
&+ 2\mathbf{u}^T(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}\mathbf{M}\mathbf{u}, \\
\tau(\rho) &= [a^2b(1-\rho)+\rho]/a. \\
\text{with} \quad a &= (\mathbf{Z}^T\mathbf{Z})^{-1}, \ b = \mathbf{Z}^T\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}\mathbf{Z}, \text{ and } \eta_0 = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{u}^T\mathbf{Z}\mathbf{Z}^T\mathbf{u}.
\end{aligned}$$

As a result, we have $S_\rho = (1-\rho)\kappa + \tau(\rho)\eta_0$.

The asymptotic distribution of S_ρ can be approximated as $(1-\rho)(\sum_{k=1}^m \lambda_k^{'}\eta_k + \zeta) + \tau(\rho)\eta_0$, since under the null, each elements of $\mathbf{u}$ has mean 0 and variance 1, $\mathbf{u}^T(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{M})\mathbf{u}$ asymptotically follows $\sum_{k=1}^m \lambda_k^{'}\eta_k$, where $\{\lambda_1^{'}, \ldots, \lambda_m^{'}\}$ are non-zero eigenvalues of $(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{M})$, η_ks are iid χ_1^2 random variables, $\eta_0 = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{u}^T\mathbf{Z}\mathbf{Z}^T\mathbf{u}$ asymptotically follows χ_1^2 distribution. Furthermore, since $\mathbf{M}$ is a projection matrix, $(\mathbf{I}-\mathbf{M})\mathbf{u}$ and $\mathbf{M}\mathbf{u}$ are asymptotically independent. Therefore,

$\zeta = 2\mathbf{u}^T(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}\mathbf{M}\mathbf{u}$ satisfies the following conditions:

$$\begin{aligned} E(\zeta) &= 0, \quad var(\zeta) = 4trace(\widehat{\Sigma}^{-1/2}\mathbf{M}\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}), \\ corr(\eta_0, \zeta) &= 0, \quad \text{and} \quad corr(\mathbf{u}'(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{M})\mathbf{u}, \zeta) = 0 \end{aligned}$$

In addition, due to the asymptotic independence between $(\mathbf{I}-\mathbf{M})\mathbf{u}$ and $\mathbf{M}\mathbf{u}$, it can be shown that $\mathbf{u}^T(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{M})\mathbf{u}$ and $(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{u}^T\mathbf{Z}\mathbf{Z}^T\mathbf{u}$ are also asymptotically independent. Since the Pearson correlation between κ and η_0 is zero, we can approximate S_ρ as the mixture of two independent variables. We can approximate the distribution of κ by using the moment matching or the characteristic function inversion method (Davis, 1980) after adjusting for the extra variance term of ζ.

To estimate the distribution of $T = \min\{p_{\rho_1}, \ldots, p_{\rho_\nu}\}$, let $q_{\min}(\rho)$ denote the $(1-T)$th percentile of the distribution of S_ρ for each ρ in the grid. The p-value of T can be computed from

$$\begin{aligned} 1 &- P(S_{\rho_1} < q_{\min}(\rho_1), \ldots, S_{\rho_\nu} < q_{\min}(\rho_\nu)) \\ = 1 &- E[P(\kappa < \min\{(q_{\min}(\rho_i) - \tau(\rho_i)\eta_0)/(1-\rho_i)\})|\eta_0], \qquad (2.5.7) \end{aligned}$$

which can be obtained by one-dimensional numerical integration.

To sum up, our proposed method can be implemented through the following algorithm:

Step 1: Set a grid $0 \le \rho_1 \le \rho_2 \le \ldots \le \rho_\nu \le 1$.

Step 2: Compute $S_{\rho_1}, \ldots, S_{\rho_\nu}$ using equation (2.5.1).

Step 3: Compute $\mathbf{Z}$, $\mathbf{M}$, λ_k's, $\tau(\rho_i)$,

$$\begin{aligned} \mu_S &= \sum_{k=1}^{m} \lambda_k', \quad \sigma_\zeta = 2\sqrt{trace(\widehat{\Sigma}^{-1/2}\mathbf{M}\widehat{\Sigma}^{-1/2}\mathbf{K}\widehat{\Sigma}^{-1/2}(\mathbf{I}-\mathbf{M})\widehat{\Sigma}^{-1/2}\mathbf{K})}, \\ \text{and } \sigma_S &= \sqrt{2\sum_{k=1}^{m}(\lambda_k')^2 + \sigma_\zeta^2}. \end{aligned}$$

Step 4: For each $\rho_i, i \in \{1, \ldots, \nu\}$, computer p_{ρ_i} and $q_{min}(\rho_i)$ from the mixture of χ^2 distribution in equation (2.5.2), and set $T = \min\{p_{\rho_1}, \ldots, p_{\rho_\nu}\}$.

Step 5: Numerically integrate $F(\delta(x)|\lambda)f(x|\chi_1^2)$, where

$$\delta(x) = (\min\{(q_{min}(\rho_i) - \tau(\rho_i)x)/(1-\rho_i)\} - \mu_S)\frac{\sqrt{\sigma_S^2 - \sigma_\zeta^2}}{\sigma_S} + \mu_S,$$

$f(x|\chi_1^2)$ is the density function of χ_1^2, and $F(\delta(x)|\lambda)$ is a distribution function of a mixture of chi-square distribution $\sum \lambda_k' \chi_k^2$. The p-value is computed as

$$p - value = 1 - \int F(\delta(x)|\lambda)f(x|\chi_1^2)dx.$$

2.5.3 Using Z-scores Instead of Effect-size Estimates

Based on p-values (p_i), sample sizes (n_i) and direction of effects (Δ_i), we can construct a signed Z-score $Z_i = \Phi^{-1}(1 - p_i/2) * sign(\Delta_i)$ for each study, where $\Phi(\cdot)$ is the standard normal distribution function. Now we show how to transform the Z-scores as input data for our proposed method.

2.5.3.1 Continuous Traits

For continuous traits, the linear regression model can be written as

$$y_{ik} = \beta_0 + \beta_i g_{ik} + \epsilon_{ik},$$

where y_{ik} is a trait value of individual k in study i , g_{ik} is a minor allele count, and $\epsilon_{ik} \sim N(0, \omega_i^2)$ is the error term. We further denote $\boldsymbol{x}_{ik} = (1, g_{ik})$ and $\mathbf{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i})^T$. Then the estimator of β_i follows the normal distribution

$$\widehat{\beta}_i \sim N(\beta_i, \sigma_i^2),$$

where $\sigma_i^2 = \omega_i^2(\mathbf{X}_i^T\mathbf{X}_i)^{-1}_{2,2}$ and $(\mathbf{X}_i^T\mathbf{X}_i)^{-1}_{2,2}$ is the (2,2) element of $(\mathbf{X}_i^T\mathbf{X}_i)^{-1}$. The two side p-value is $p_i = 1 - 2\Phi(|\widehat{\beta}_i/\sigma_i|)$, and thus the Z-score Z_i follows $N(\beta_i/\sigma_i, 1)$. This result implies that we can reconstruct $\widehat{\beta}_i$ using a Z-score by estimating σ_i^2. Since

$$(\mathbf{X}_i^T\mathbf{X}_i)^{-1} = \frac{1}{n_i\sum_{k=1}^{n_i} g_{ik}^2 - (\sum_{k=1}^{n_i} g_{ik})^2}\begin{pmatrix} \sum_{k=1}^{n_i} g_{ik}^2 & -\sum_{k=1}^{n_i} g_{ik} \\ -\sum_{k=1}^{n_i} g_{ik} & n_i \end{pmatrix},$$

we then have under the Hardy-Weinberg equilibrium

$$\frac{n_i\omega_i^2}{n_i\sum_{k=1}^{n_i} g_{ik}^2 - (\sum_{k=1}^{n_i} g_{ik})^2} \quad \approx \quad \frac{\omega_i^2}{n_i 2q_i(1-q_i)},$$

where q_i is a minor allele frequency (MAF) for the corresponding queried SNP. As a result, under the Hardy-Weinberg equilibrium $\widehat{\beta}_i$ is equivalent to $\sqrt{\frac{\omega_i^2}{n_i 2q_i(1-q_i)}}Z_i$. With an additional assumption that the variance of error term (ω_i^2) are the same across studies, we can use

$$\tilde{\beta}_i = Z_i/\sqrt{n_i q_i(1-q_i)}$$

and its standard error

$$\tilde{\sigma}_i = 1/\sqrt{n_i q_i(1-q_i)}$$

as inputs for our proposed method.

2.5.3.2 Binary Traits

For binary traits, the logistic regression model can be written as

$$logitPr(y_{ik} = 1) = \beta_0 + \beta_i g_{ik}.$$

Asymptotically, $var(\widehat{\boldsymbol{\beta}}_i) = J^{-1}(\boldsymbol{\beta}_i)$, where $J(\boldsymbol{\beta}_i) = \sum_{k=1}^{n_i} \boldsymbol{x}_{ik}\boldsymbol{x}_{ik}^T \mu_{ik}(1-\mu_{ik})$, and $\mu_{ik} = \frac{\exp(\boldsymbol{\beta}_i^T \boldsymbol{x}_{ik})}{1+\exp(\boldsymbol{\beta}_i^{'} \boldsymbol{x}_{ik})}$. Since

$$\boldsymbol{x}_{ik}\boldsymbol{x}_{ik}^T = \begin{pmatrix} 1 & g_{ik} \\ g_{ik} & g_{ik}^2 \end{pmatrix},$$

we then have

$$J(\boldsymbol{\beta}_i) = \sum_{k=1}^{n_i} \frac{\exp(\boldsymbol{\beta}_i^T \boldsymbol{x}_{ik})}{[1+\exp(\boldsymbol{\beta}_i^T \boldsymbol{x}_{ik})]^2} \begin{pmatrix} 1 & g_{ik} \\ g_{ik} & g_{ik}^2 \end{pmatrix}.$$

If we use $r_i = n_{case,i}/n_i$ to denote the proportion of case samples for study i and assume that its effect size β_i is very small, then $\frac{\exp(\boldsymbol{\beta}_i^T \boldsymbol{x}_{ik})}{[1+\exp(\boldsymbol{\beta}_i^T \boldsymbol{x}_{ik})]^2} \approx r_i(1-r_i)$ for any

$k \in \{1, \ldots, n_i\}$, and $J(\boldsymbol{\beta}_i)$ reduces to

$$J(\boldsymbol{\beta}_i) = r_i(1-r_i) \sum_{k=1}^{n_i} \begin{pmatrix} 1 & g_{ik} \\ g_{ik} & g_{ik}^2 \end{pmatrix} = r_i(1-r_i) \begin{pmatrix} n_i & \sum g_{ik} \\ \sum g_{ik} & \sum g_{ik}^2 \end{pmatrix}.$$

The remaining derivations then follow the same calculation as in the continuous traits case. As a result, for binary traits, the log odds ratio estimate $\widehat{\beta}_i$ is asymptotically equivalent to $Z_i/\sqrt{n_i r_i(1-r_i)q_i(1-q_i)}$. If all studies have similar ratios of cases and controls, the $r_i(1-r_i)$ term can be ignored. Consequently,

$$\tilde{\beta}_i = Z_i/\sqrt{n_i q_i(1-q_i)}$$

and its standard error

$$\tilde{\sigma}_i = 1/\sqrt{n_i q_i(1-q_i)}$$

can be used as inputs for both continuous and binary traits.

2.5.4 Prior Density Function for MANTRA

MANTRA uses the Bayesian partition model to adaptively determine the number of ethnic clusters and the cluster membership and assesses the association evidence by means of the Bayes factor. We use the same prior density functions as suggested in Morris (2011) in our simulation studies and data analysis. Specifically, let M_0 denote the null model with $\boldsymbol{\beta} = \mathbf{0}$, in which there is no association of the variant with the trait in any population, and M_1 denotes the alternative model with $\boldsymbol{\beta} \neq \mathbf{0}$, then the evidence in favor of the alternative model can be assessed by means of the Bayes factor (Kass and Raftery, 1995) given by

$$\begin{aligned}
\Delta &= \frac{f(\hat{\boldsymbol{\beta}}, \Sigma|M_1)}{f(\hat{\boldsymbol{\beta}}, \Sigma|M_0)}, \\
\text{where } f(\hat{\boldsymbol{\beta}}, \Sigma|M) &\propto \int_{\boldsymbol{\beta}} f(\hat{\boldsymbol{\beta}}, \Sigma|\boldsymbol{\beta}) f(\boldsymbol{\beta}|M) \partial \boldsymbol{\beta}, \\
\text{and } f(\hat{\boldsymbol{\beta}}, \Sigma|\boldsymbol{\beta}) &= \prod_{i=1}^{n} f(\hat{\beta}_i, \sigma_i^2|\beta_i), \\
\text{with } f(\hat{\beta}_i, \sigma_i^2|\beta_i) &\propto \frac{1}{\sigma_i} \exp\{\frac{-(\hat{\beta}_i - \beta_i)^2}{2\sigma_i^2}\}.
\end{aligned}$$

Suppose the n populations can be allocated to B cluster centers $\mathbf{C} = \{C_1, C_2, \ldots, C_B\}$ with the corresponding cluster allelic effects $\boldsymbol{\Psi} = \{\Psi_1, \Psi_2, \ldots, \Psi_B\}$. The true effect size $\boldsymbol{\beta}$ is determined by the assignment of populations to ethnic clusters under a Bayesian partition model. The assignment is given by $\mathbf{T}$, where $T_{ib} = 1$ if the i^{th} population is allocated to the cluster with center C_b and 0 otherwise. Under such an assignment, the marginal likelihood can be written as

$$f(\hat{\beta}_i, \sigma_i^2|\beta_i) = f(\hat{\beta}_i, \sigma_i^2|B, \mathbf{C}, \boldsymbol{\Psi}) \propto \frac{1}{\sigma_i} \exp\{\frac{-(\hat{\beta}_i - \sum_{b=1}^{B} T_{ib}\Psi_b)^2}{2\sigma_i^2}\}.$$

Under the null model M_0, the population-specific allelic effect are all zero, and hence any clustering of populations is irrelevant. Consequently, $f(\boldsymbol{\beta}|M_0) = 1$ if $\boldsymbol{\beta} = \mathbf{0}$, and 0 otherwise. Under the alternative model M_1, population-specific allelic effects are determined by the Bayesian partition model, in which the prior density of the number of clusters of populations is given by

$$f(B) = \begin{cases} 1 & \text{if } B = 1 \\ \frac{2^{n-1}}{2^B(2^{n-1}-1)} & \text{otherwise} \end{cases}.$$

Given B, each population is equally likely, a priori, to be a cluster center, and the cluster allelic effect have a prior $N(\mu, \theta)$ distribution, independent of $\mathbf{C}$, where μ has a prior uniform distribution and θ has a prior exponential distribution with expectation 1. Combining the components of the prior density function, it follows that

$$f(\boldsymbol{\beta}|M_1) \propto f(B)(n-B)!\frac{\exp\{-\theta\}}{\theta}\prod_{b=1}^{B}\exp\{\frac{(\Psi_b-\mu)^2}{2\theta^2}\}.$$

2.5.5 Estimation of Bayes Factor Threshold

We carried out 20 million null simulations for MANTRA to find Bayes factor thresholds corresponding to genome-wide p-value significance levels. Following our type I error simulations as in Section 2.3, each simulated dataset had 27 studies (9 ancestry groups in triplicate) and each study has 500 cases and 500 controls. We then applied MANTRA to those 20 million nulls to obtain the Bayes factors, and computed the empirical type I error rates as the proportion of Bayes factors (out of the 20 million) that were greater than a given Bayes factor threshold. When we set log10 Bayes factor = 5 as a threshold, the empirical type I error rate was 1.8×10^{-6} with the exact binomial confidence interval $(1.25 \times 10^{-6}, 2.4 \times 10^{-6})$. Supplementary Figure 2.6 ploted the obtained empirical type I error rates (illustrated in -log10(empirical type I error rate) on the vertical axis) and the Bayes factors (illustrated in log10(Bayes' factor) on the horizontal axis).

Due to our limited computing resources, it would take us months to run MANTRA on billions of null simulations that are required to find a comparable Bayes factor threshold to the commonly used genome-wide significance level ($\alpha = 5 \times 10^{-8}$); therefore, we performed a regression analysis between the Bayes factor thresholds and the empirical type I error rates. We first obtained the empirical type I error rates for a sequence of Bayes factor thresholds, and then fitted a linear regression model using -log10(Empirical type I error rate) as a response variable and the log10 Bayes' factor threshold as a predictor. The resulting regression intercept and slope were 1.08577 (p-value $< 2 \times 10^{-16}$) and 0.98106 (p-value $< 2 \times 10^{-16}$) respectively. Based on those regression parameters, we estimated that Bayes factor threshold (on the log10 base) which corresponds to the genome-wide significance level was 6.34. We noted that from this linear model, the estimated significance level that corresponds to log10 Bayes factor $= 5$ was $\alpha = 1.0 \times 10^{-6}$, which was slightly lower than the observed threshold $\alpha = 1.8 \times 10^{-6}$. We employ both $\alpha = 1.8 \times 10^{-6}$ and 1.0×10^{-6} to the power simulations and observed that the results were very similar (data not shown).

To sum up, we defined the level of significance as a p-value less than 1.8×10^{-6}, or as a log10 Bayes factor larger than 5. We also employed the significance level as a p-value less than 5×10^{-8} or as a log10 Bayes factor larger than 6.34.

2.5.6 Supplementary Tables and Figures

Table 2.2: **Type-I error rate estimates for TransMeta at different α levels, with one study in each ancestry group.** Type-I error rate estimates at different α levels based on 100 million replicates. Each entry represents an estimated type I error rate calculated using the proportion of p-values smaller than the given level α. One integrated study was simulated per ancestry group, and each study had 1500 cases and 1500 controls.

	$\alpha = 10^{-2}$	10^{-3}	10^{-4}	10^{-5}	10^{-6}
TransMeta.Fst	1.051×10^{-2}	1.1×10^{-3}	1.079×10^{-4}	1.1×10^{-5}	1.05×10^{-6}
TransMeta.Indep	1.008×10^{-2}	0.9×10^{-3}	8.589×10^{-5}	7.4×10^{-6}	9.0×10^{-7}

Table 2.3: **Pairwise F_{st} values used for the T2D meta-analysis.** The F_{st} values were extracted from Supplementary Table 6 of International HapMap 3 Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311), 52-58.

Ancestry	European	east Asian	south Asian	Mexican and Mexican-American
European	0	0.111	0.035	0.031
east Asian	0.111	0	0.077	0.070
south Asian	0.035	0.077	0	0.035
Mexican and Mexican-American	0.031	0.070	0.035	0

Table 2.4: **Meta-analysis results for the 24 SNPs with TransMeta.Fst p-value $< 5 \times 10^{-8}$ from the T2D trans-ethnic meta-analysis data.** P-values and Bayes' factors of the six meta-analysis methods for the 24 SNPs with TransMeta.Fst p-value $< 5 \times 10^{-8}$ among the 69 SNPs from the T2D trans-ethnic meta-analysis data. Values in the parenthesis are the optimal ρ values for our proposed method. Values in the last column are the I^2 statistic for measuring the heterogeneity level.

SNP	F.ST (ρ)	INDEP(ρ)	FE	RE	RE-HE	Bayes	I^2
rs7903146	6.17e-77 (0.00)	3.59e-84 (0.25)	6.44e-75	2.89e-07	3.55e-76	74.13	0.83
rs10811661	4.42e-27 (1.00)	4.42e-27 (1.00)	1.11e-27	1.28e-24	2.74e-27	25.36	0.10
rs7756992	4.04e-26 (0.25)	3.57e-31 (0.25)	3.39e-26	2.19e-04	1.88e-27	24.87	0.81
rs3802177	6.55e-19 (0.09)	2.10e-19 (0.25)	1.61e-18	1.61e-18	3.39e-18	16.27	0
rs1111875	1.12e-18 (1.00)	1.33e-20 (0.25)	2.80e-19	2.49e-05	3.29e-19	17.12	0.65
rs4402960	5.51e-18 (0.25)	4.22e-18 (0.25)	7.50e-18	1.54e-17	1.55e-17	15.52	0.01
rs163184	4.12e-14 (1.00)	2.63e-14 (0.25)	1.03e-14	4.80e-07	1.64e-14	12.41	0.55
rs9936385	3.32e-12 (0.25)	9.15e-13 (0.25)	9.65e-13	3.01e-10	1.67e-12	10.63	0.11
rs7178572	5.70e-11 (0.25)	4.91e-11 (0.25)	1.47e-11	1.47e-11	2.45e-11	9.35	0
rs5215	1.25e-10 (1.00)	1.07e-10 (0.25)	3.12e-11	8.47e-05	3.24e-11	8.98	0.57
rs12571751	2.19e-10 (1.00)	2.43e-10 (1.00)	2.19e-10	2.19e-10	3.46e-10	8.22	0
rs1801282	3.86e-10 (1.00)	2.89e-10 (0.25)	4.24e-10	4.24e-10	6.41e-10	7.99	0
rs849135	3.88e-10 (0.00)	2.21e-10 (0.25)	1.06e-09	2.84e-03	1.07e-09	7.62	0.53
rs17791513	1.01e-09 (0.00)	1.11e-08 (0.09)	2.42e-08	4.71e-03	1.76e-08	6.60	0.65
rs4430796	1.06e-09 (1.00)	2.59e-09 (1.00)	1.16e-09	1.10e-07	1.71e-09	7.53	0.24
rs4458523	1.72e-09 (1.00)	1.79e-09 (1.00)	1.91e-09	1.91e-09	2.88e-09	7.32	0
rs11257655	2.06e-09 (1.00)	5.31e-09 (1.00)	1.92e-09	8.71e-04	2.22e-09	7.33	0.61
rs2943640	6.52e-09 (1.00)	6.63e-09 (1.00)	7.01e-09	7.01e-09	9.96e-09	6.73	0
rs7612463	8.25e-09 (1.00)	1.7e-08 (0.25)	6.28e-09	6.28e-09	9.21e-09	6.86	0
rs11717195	1.46e-08 (0.25)	3.17e-08 (1.00)	2.26e-08	2.26e-08	3.25e-08	6.20	0
rs4812829	2.09e-08 (0.00)	1.59e-08 (0.25)	4.21e-08	1.42e-04	5.98e-08	6.07	0.4
rs12970134	2.98e-08 (1.00)	4.79e-08 (1.00)	2.48e-08	2.48e-08	3.55e-08	6.06	0
rs10830963	2.98e-08 (0.25)	3.76e-08 (0.25)	1.99e-07	2.91e-03	2.48e-07	5.60	0.50
rs2261181	3.05e-08 (1.00)	7.51e-09 (0.25)	2.34e-08	1.31e-05	3.11e-08	6.24	0.27

Table 2.5: **Table 2.4 continued: Meta-analysis results for the remaining 45 SNPs from the T2D trans-ethnic meta-analysis data.** P-values and Bayes' factors of the six meta-analysis methods for the remaining 45 SNPs among the 69 SNPs from the T2D trans-ethnic meta-analysis data.

SNP	F.ST (ρ)	INDEP(ρ)	FE	RE	RE-HE	Bayes	I^2
rs7845219	6.56e-08 (1.00)	8.63e-08 (1.00)	5.84e-08	5.84e-08	8.57e-08	5.99	0
rs516946	6.57e-08 (0.09)	6.60e-08 (0.25)	1.11e-07	1.11e-07	1.63e-07	5.59	0
rs1552224	1.35e-07 (1.00)	7.71e-08 (0.25)	9.61e-08	2.11e-03	9.68e-08	5.81	0.63
rs17168486	3.86e-07 (1.00)	4.36e-07 (0.09)	3.74e-07	4.38e-03	3.65e-07	5.08	0.58
rs12899811	6.29e-07 (1.00)	1.40e-06 (1.00)	7.42e-07	2.05e-05	1.09e-06	4.74	0.16
rs2028299	6.48e-07 (1.00)	9.0e-07 (1.00)	7.741e-07	2.35e-04	9.09e-07	4.81	0.42
rs1535500	1.45e-06 (0.00)	1.53e-06 (0.25)	5.36e-06	1.13e-02	5.61e-06	4.10	0.52
rs3923113	1.96e-06 (0.25)	2.29e-06 (1.00)	2.31e-06	1.51e-02	5.47e-07	4.62	0.74
rs2796441	1.96e-06 (1.00)	2.43e-06 (1.00)	1.63e-06	1.63e-06	2.39e-06	4.42	0
rs2075423	2.03e-06 (1.00)	2.52e-06 (1.00)	2.17e-06	9.69e-04	3.17e-06	4.34	0.45
rs12427353	3.11e-06 (0.00)	3.13e-06 (0.25)	3.41e-06	3.41e-06	4.17e-06	4.14	0
rs243088	3.49e-06 (1.00)	3.56e-06 (0.25)	3.73e-06	3.73e-06	5.46e-06	4.25	0
rs7163757	4.76e-06 (1.00)	6.51e-06 (1.00)	4.14e-06	4.14e-06	6.04e-06	4.11	0
rs10842994	4.76e-06 (0.25)	6.92e-06 (1.00)	6.75e-06	6.75e-06	9.84e-06	3.93	0
rs8108269	4.98e-06 (1.00)	6.86e-06 (1.00)	4.60e-06	1.97e-03	6.71e-06	3.90	0.43
rs7041847	5.31e-06 (1.00)	7.21e-06 (1.00)	4.03e-06	4.12e-06	5.88e-06	4.20	0
rs11634397	6.29e-06 (0.00)	7.85e-06 (0.25)	1.60e-05	2.58e-03	2.16e-05	3.62	0.31
rs1359790	9.61e-06 (0.25)	2.46e-06 (0.25)	1.08e-05	8.73e-03	1.07e-05	3.60	0.47
rs780094	1.45e-05 (1.00)	1.62e-05 (1.00)	1.29e-05	2.76e-02	5.34e-06	3.81	0.75
rs10203174	3.29e-05 (0.00)	7.11e-06 (0.09)	7.28e-05	1.64e-01	4.99e-05	2.59	0.65
rs7955901	3.11e-05 (0.00)	1.62e-05 (0.00)	1.86e-03	3.68e-01	1.79e-04	2.15	0.76
rs6795735	3.59e-05 (0.00)	1.41e-04 (0.25)	2.00e-04	4.65e-03	2.80e-04	2.60	0.27
rs7593730	3.6e-05 (0.00)	1.13e-05 (0.00)	4.74e-04	1.89e-01	1.34e-04	2.41	0.68
rs7202877	4.32e-05 (0.00)	2.28e-04 (0.09)	5.53e-04	6.23e-02	2.09e-04	2.43	0.72
rs13233731	1.11e-04 (0.00)	1.97e-06 (0.00)	4.08e-03	3.42e-01	1.35e-05	3.90	0.85
rs16861329	2.68e-04 (0.00)	1.46e-05 (0.00)	5.06e-02	6.95e-01	1.01e-04	2.45	0.90
rs11063069	3.33e-04 (0.00)	4.02e-04 (0.25)	9.97e-04	3.87e-02	1.40e-03	1.78	0.25
rs3786897	3.83e-04 (1.00)	3.20e-04 (0.25)	3.34e-04	2.22e-01	1.45e-05	3.84	0.83
rs9470794	3.95e-04 (0.00)	2.61e-04 (0.09)	1.75e-03	3.48e-01	1.53e-03	1.81	0.68
rs6815464	4.39e-04 (0.00)	4.39e-04 (0.00)	NA	NA	NA	2.13	0
rs6878122	5.81e-04 (0.25)	3.23e-04 (0.25)	5.75e-04	1.36e-01	4.64e-05	2.23	0.82
rs1802295	6.97e-04 (0.00)	1.22e-03 (0.25)	1.10e-03	1.56e-01	1.95e-04	1.97	0.81
rs831571	6.84e-04 (1.00)	3.99e-04 (0.25)	5.26e-04	2.10e-01	4.56e-04	2.25	0.73
rs459193	1.06e-03 (1.00)	1.51e-03 (1.00)	8.20e-04	8.20e-04	1.15e-03	1.84	0
rs2334499	1.61e-03 (1.00)	1.64e-03 (0.25)	1.38e-03	1.38e-03	1.93e-03	1.69	0
rs10923931	3.03e-03 (0.00)	1.01e-03 (0.00)	7.10e-03	3.55e-01	8.61e-03	0.95	0.46
rs10401969	3.91e-03 (0.00)	3.35e-03 (0.09)	7.18e-03	1.62e-01	6.19e-03	1.15	0.68
rs6467136	6.72e-02 (0.00)	5.93e-02 (0.00)	2.14e-01	4.63e-01	1.80e-02	0.80	0.76
rs10278336	1.08e-01 (0.00)	1.21e-01 (0.09)	1.11e-01	1.73e-01	1.33e-01	0.11	0.16
rs7403531	1.54e-01 (0.25)	3.78e-02 (0.00)	1.28e-01	7.23e-01	6.81e-02	0.20	0.68
rs6723108	3.49e-01 (1.00)	4.30e-01 (1.00)	3.17e-01	3.17e-01	3.64e-01	-0.21	0
rs17584499	4.98e-01 (0.00)	4.63e-01 (0.00)	5.20e-01	5.62e-01	5.60e-01	-1.18	0.52
rs7560163	5.76e-01 (1.00)	6.06e-01 (1.00)	4.72e-01	4.72e-01	5.08e-01	-0.37	0
rs10886471	5.86e-01 (0.00)	6.45e-01 (0.00)	6.46e-01	6.46e-01	7.05e-01	-0.45	0
rs391300	8.42e-01 (1.00)	8.78e-01 (1.00)	7.40e-01	7.40e-01	7.90e-01	-0.55	0

Table 2.6: **Summary of the computed I^2 statistic in the power simulations.** The I^2 statistic for each of the 2000 SNPs in the five power comparison scenarios. In each cell of the table, we first present the median of the I^2 statistic for all the SNPs (out of 2000) whose optimal ρ value from TransMeta.Fst is as specified at beginning of the row, then we present the corresponding inter-quartile range (IQR) in the parenthesis.

The optimal ρ value	Scenario (a)	Scenario (b)	Scenario (c)	Scenario (d)	Scenario (e)
$\rho = 0$	0.05 (0.28)	0.55 (0.27)	0.69 (0.17)	0.50 (0.29)	0.70 (0.17)
$\rho = 0.09$	0 (0.17)	0.39 (0.45)	0.65 (0.33)	0.21 (0.45)	0.64 (0.18)
$\rho = 0.25$	0 (0.11)	0.20 (0.51)	0.36 (0.38)	0.11 (0.36)	0.63 (0.23)
$\rho = 1$	0 (0.18)	0.03 (0.32)	0.27 (0.36)	0.12 (0.33)	0.53 (0.32)
Overall median (IQR)	0 (0.21)	0.52 (0.33)	0.68 (0.19)	0.45 (0.38)	0.67 (0.20)

Table 2.7: **Contingency Table of the selected optimal ρ value from TransMeta.Fst for each of the 2000 SNPs in the five power comparison scenarios.** In each cell of the table, the entry represents the total number of SNPs (out of 2000) which has the selected optimal ρ value as listed at the beginning of the row under the scenario specified at the top of the column.

The optimal ρ value	Scenario (a)	Scenario (b)	Scenario (c)	Scenario (d)	Scenario (e)
$\rho = 0$	480	**1674**	**1864**	**1573**	**1385**
$\rho = 0.09$	309	136	58	120	109
$\rho = 0.25$	276	77	36	122	143
$\rho = 1$	**935**	113	142	185	363
Total counts	2000	2000	2000	2000	2000

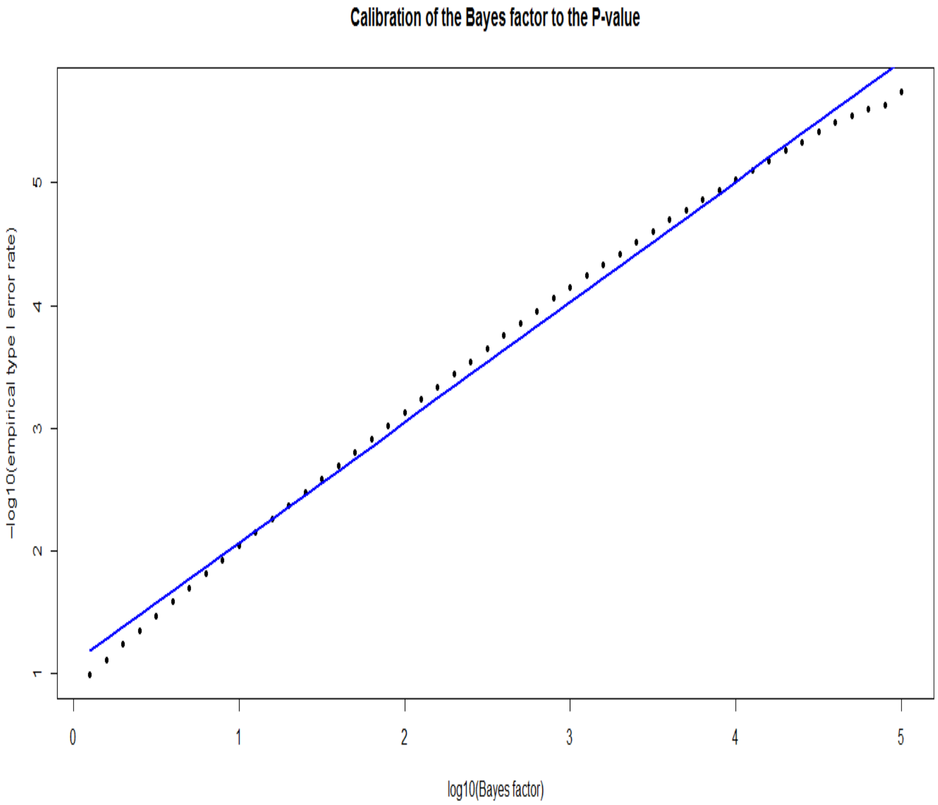


Figure 2.6: **Calibration of the Bayes' factor to the empirical type I error rate.** The vertical axis measures the empirical type I error rate on a -log10 scale, the horizontal axis measures the Bayes' factor on a log10 scale. The blue straight represents the fitted regression line -log10(empirical type I error rate) = 1.08577 + 0.98106 × log10(Bayes' factor).

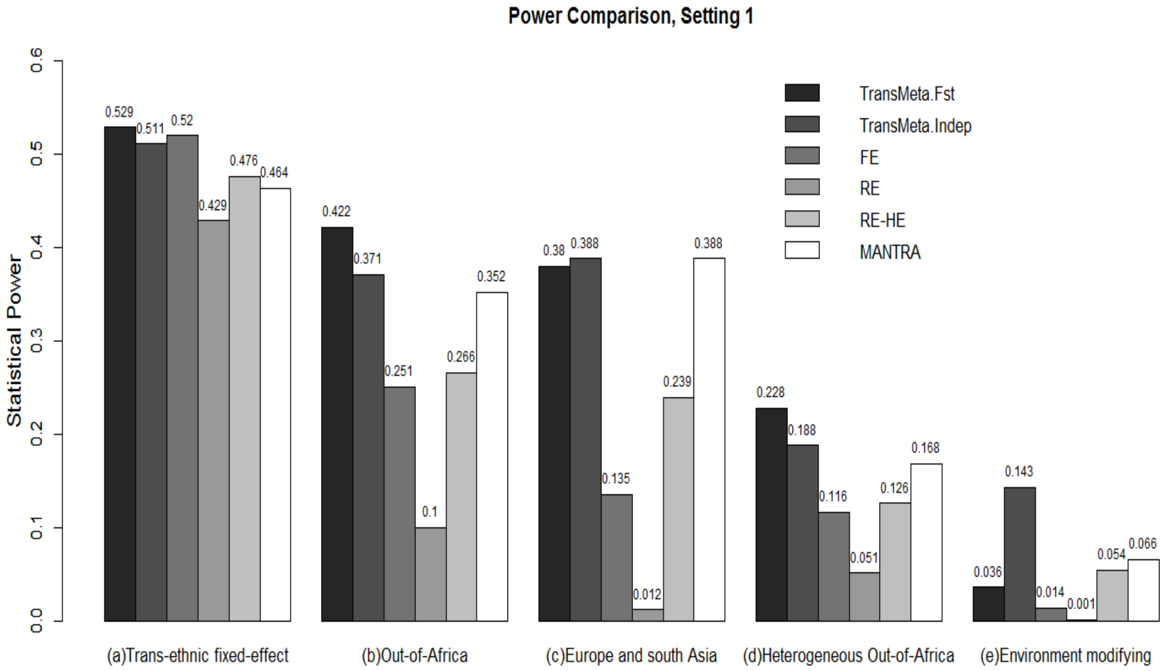


Figure 2.7: **Empirical power for TransMeta under different effect-size heterogeneity configurations, with three sub-studies in each ancestry group, and significance level at $\boldsymbol{\alpha = 5 \times 10^{-8}}$.** Empirical power for TransMeta and existing methods under the five effect size scenarios. Three studies are simulated per ancestry group, each with 500 cases and 500 controls. The empirical power is obtained based on 2000 replicates with the level of significance defined as a p-value less than 5×10^{-8} or as a log10 Bayes' factor larger than 6.34. The five effect size scenarios are (a) 'Trans-ethnic fixed-effect', where no heterogeneity exists in allelic effects at the causal SNP between populations; (b) 'Out-of-Africa effect', where only studies from the non-African populations carry the causal variant; (c) 'Europe and south Asia effect', where only studies from the European and south Asian populations carry the causal variant; (d) 'Heterogeneous Out-of-Africa effect', where the causal variant has genetic effects only in non-African populations, but the effect size in the east Asian populations is different from that in the European and south Asian populations; (e) 'Environment modifying effect', where the causal variant has genetic effect only in the populations living in Europe and USA.

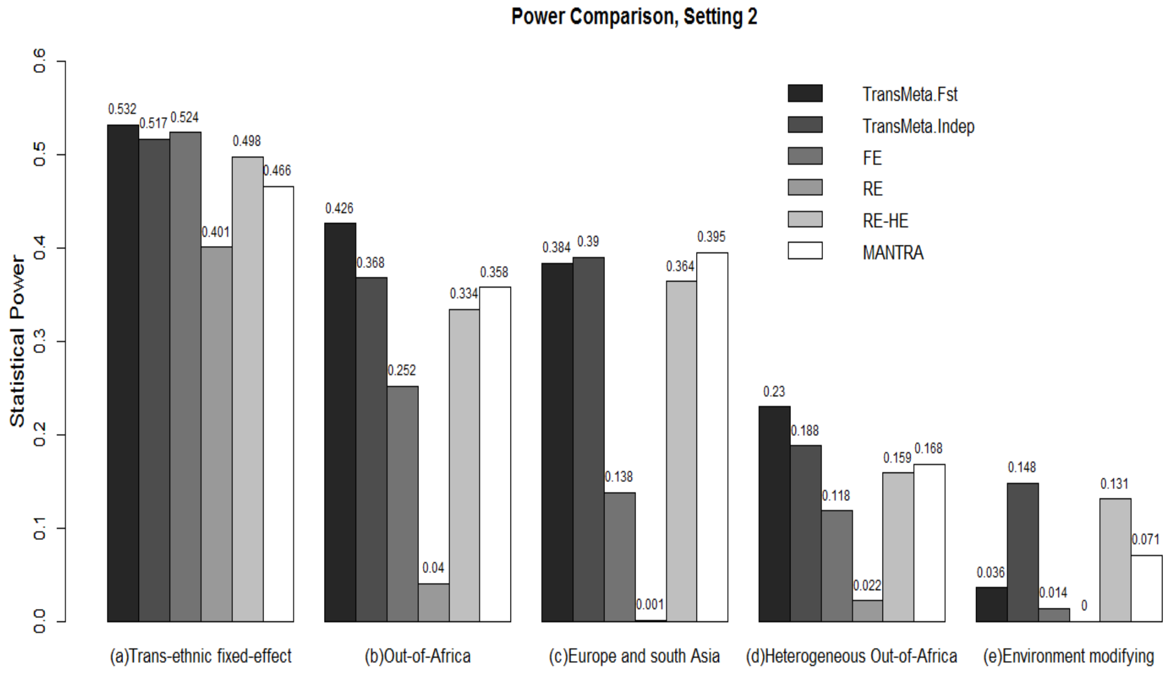


Figure 2.8: **Empirical power for TransMeta under different effect-size heterogeneity configurations, with one integrated study in each ancestry group, and significance level at $\alpha = 5 \times 10^{-8}$.** Empirical power for TransMeta and existing methods under the five effect size scenarios. One integrated study is simulated per ancestry group, each with 1500 cases and 1500 controls. The empirical power is obtained based on 2000 replicates with the level of significance defined as a p-value less than 5×10^{-8} or as a log10 Bayes' factor larger than 6.34. The five effect size scenarios are (a) 'Trans-ethnic fixed-effect', where no heterogeneity exists in allelic effects at the causal SNP between populations; (b) 'Out-of-Africa effect', where only studies from the non-African populations carry the causal variant; (c) 'Europe and south Asia effect', where only studies from the European and south Asian populations carry the causal variant; (d) 'Heterogeneous Out-of-Africa effect', where the causal variant has genetic effects only in non-African populations, but the effect size in the east Asian populations is different from that in the European and south Asian populations; (e) 'Environment modifying effect', where the causal variant has genetic effect only in the populations living in Europe and USA.

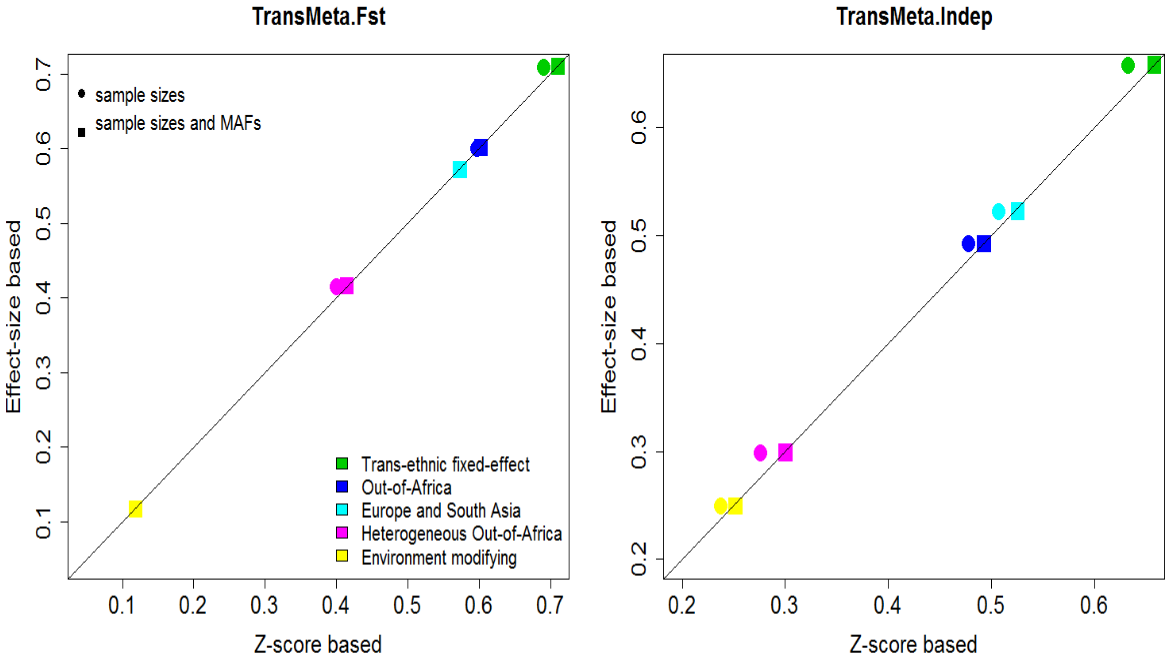


Figure 2.9: **Power comparison of the effect-size-based and Z-score-based TransMeta, with one integrated study in each ancestry group, and significance level at $\alpha = 1.8 \times 10^{-6}$.** Power comparison of the effect-size and Z-score based TransMeta under the five effect size scenarios. One integrated study is simulated per ancestry group, each with 1500 cases and 1500 controls. The empirical power is obtained based on 2000 replicates with the level of significance defined as a p-value less than 1.8×10^{-6}. The left panel is based on TransMeta.Fst and the right panel is based on TransMeta.Indep. In each plot, the x-axis denotes empirical power of the Z-score based TransMeta and the y-axis denotes empirical power of the effect-size based TransMeta. The solid dots represent the power of transformed Z-scores using only sample sizes, and the solid squares represent transformed Z-scores using both sample sizes and MAFs.

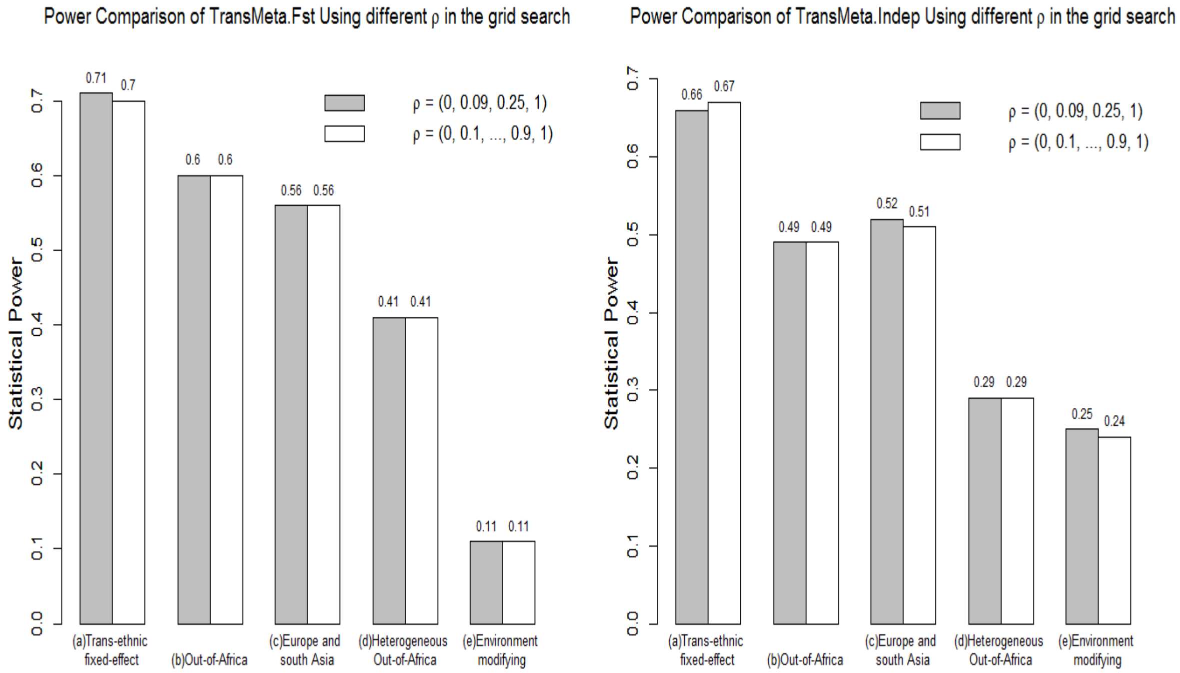


Figure 2.10: **Power comparison of TransMeta under a coarse grid v.s. a dense grid, with three sub-studies in each ancestry group, and significance level at $\boldsymbol{\alpha = 1.8 \times 10^{-6}}$.** Comparison of the empirical power for TransMeta under the five effect size scenarios, using different grid searches for ρ . Three studies are simulated per ancestry group, each with 500 cases and 500 controls. The empirical power is obtained based on 2000 replicates with the level of significance defined as a p-value less than 1.8×10^{-6}. The two grids being compared are: $\rho = (0, 0.09, 0.25, 1)$ v.s $\rho = (0, 0.1, 0.2, \ldots, 0.8, 0.9, 1)$. The left panel is based on TransMeta.Fst and the right panel is based on TransMeta.Indep. The five effect size scenarios are (a) 'Trans-ethnic fixed-effect', where no heterogeneity exists in allelic effects at the causal SNP between populations; (b) 'Out-of-Africa effect', where only studies from the non-African populations carry the causal variant; (c) 'Europe and south Asia effect', where only studies from the European and south Asian populations carry the causal variant; (d) 'Heterogeneous Out-of-Africa effect', where the causal variant has genetic effects only in non-African populations, but the effect size in the east Asian populations is different from that in the European and south Asian populations; (e) 'Environment modifying effect', where the causal variant has genetic effect only in the populations living in Europe and USA.

CHAPTER III

Trans-Ethnic Meta-Analysis of Rare Variants in Sequencing Association Studies

Abstract

Trans-ethnic meta-analysis is a powerful tool at detecting novel loci in genetic association studies. However, under the presence of inter-study genetic effect heterogeneity, existing approaches may be unsatisfactory because they do not consider genetic similarity or dissimilarity among different ancestry groups. In response, we propose a unified score test under a modified random effects model framework for rare variants associations. Specifically, we adapt the kernel regression framework to construct the modified random effects model, and incorporate the genetic similarities across ancestry groups into modeling the heterogeneity structure of the genetic effect coefficients. In addition, we use the adaptive variance component test to achieve robust power regardless of the degree of heterogeneity. A resampling-based copula method is employed to approximate the asymptotic distribution of the proposed test, which enables efficient estimation of p-values. Simulation studies show that our proposed method controls type I error rates at the exome-wide significance level and improves power over existing approaches under the presence of heterogeneity. We further illustrate our method by analyzing the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia data, a multiethnic sample of 12,940 individuals which focuses on exome sequence variations.

Keywords: Effect-size heterogeneity; GWAS; Kernel regression; Random effect model; Rare variants; Trans-ethnic meta-analysis.

3.1 Introduction

The rapid technological advances in high-throughput sequencing platforms have made it possible to test for rare variant (here defined as alleles with a frequency less than 1%) associations to accelerate our knowledge of complex trait genetics. In rare variant association studies, testing for only a single rare variant is usually underpowered due to the limited number of study participants carrying the rare allele and the penalty of multiple testing (Bansal et al., 2010). To enrich the strength of the rare variant association tests, a commonly used strategy is to group the variants in a gene or a functional unit to perform the association tests. For example, the burden test collapses the variants into a burden score and tests its association with the trait of interest (Madsen and Browning, 2009; Morris and Zeggini, 2010); the variable threshold test (VT) performs burden tests at each defined minor allele frequency (MAF) threshold and evaluates the significance for the maximum of these statistics (Price et al., 2010); the variance component tests, such as sequence kernel association test (SKAT), account for variants with opposite effects in a gene through tailored aggregation of individual variant test statistics in a gene (Wu et al., 2011); and the SKAT-O (Lee et al., 2012) and MiST (Sun et al., 2013) tests take convex combination of a burden test and a SKAT variance component test to enhance the robustness and power of the existing approaches.

Since rare variant tests are usually underpowered without an exceptionally large sample size or a sufficient number of rare alleles captured, a practical strategy to enlarge the sample size is to aggregate studies through meta-analysis. To date, most meta-analyses have been undertaken in a single population, usually of European descent; as a result, most existing meta-analysis methods for rare variant associations usually assume that the underlying genetic effects are the same across all studies. Under this homogeneity assumption, Hu et al. (2013) and Liu et al. (2014) proposed practical rare variants meta-analysis approaches to increase the study power by aggregating summary statistics across studies to increase sample sizes.

Trans-ethnic meta-analysis is a natural extension of the single-ancestry-based meta-analysis, as it aims to include samples from as many studies as possible, even if they come from different ancestries. With the further increased sample size, trans-ethnic meta-analysis is expected to be more powerful at detecting novel loci (Cooper et al., 2008; Li and Keating, 2014). However, in performing trans-ethnic analysis, one of the challenges is to properly account for the expected heterogeneity across studies from different ancestry groups. Heterogeneity can arise due to several reasons. First, for a gene-level test, different studies may present different sets of rare variants, due to the fact that rare variants are likely to be population specific and thus may not exist in all populations. Therefore, the gene-level association strengths will likely be unequal among studies, even when effect size across studies is the same for each variant. Second, if environmental risk factors differ among ancestry groups and interact with the causal variants, it is possible that the marginal genetic effects would vary between populations (Morris, 2011) due to the gene-environment interaction.

In the presence of between-study heterogeneity in multi-ethnic meta-analysis, traditional fixed effects meta-analysis approaches, which assume the same genetic effects among all participating studies, do not account for the expected variability in genetic effects. In response, several researchers have proposed using the random effects meta-analysis approach, which assumes a different underlying genetic effect for each population. For example, Lee et al. (2013) developed unified score tests that combine features of both the burden test and SKAT to incorporate genetic effect heterogeneity; Tang and Lin (2014) derived the random effects version of all commonly used rare variants association tests – such as the burden test, VT and SKAT – to allow varying genetic effects among studies.

However, these random effects-based meta-analysis approaches ignore some of the underlying characteristics of trans-ethnic meta-analysis. Specifically, these methods only assume varying genetic effects between studies, but do not consider that studies which share more similar genetic architectures can have more homogeneous genetic effects than those which consist of very disparate ancestries. In addition, these methods were developed for unrelated subjects, and thus cannot properly handle the correlated/clustered structure when the participating studies contain samples of related individuals. To avoid the type I error inflation due to failure of handling the correlated structure, a typical strategy is to remove the related individuals for analysis, which may result in power loss.

To take full advantage of the strengths of multi-ethnic meta-analysis, in this chapter, we propose a unified score test under a modified random effects model framework for rare variants associations, which can adjust for sample relatedness. The proposed method is capable not only of accounting for the expected heterogeneous genetic effects among studies, but also of flexibly modeling varying levels of heterogeneity due to the difference of genetic architectures between the populations. Specifically, we adapt the kernel regression framework to construct the modified random effects model, and incorporate the genetic similarity among ancestry groups into modeling the heterogeneity structure of the genetic effect coefficients. In addition, we use the adaptive variance component test to achieve robust power regardless of the degree of genetic effect heterogeneity. When meta-analyzing family-based studies, to account for sample relatedness, we incorporate the generalized linear mixed model association test (GMMAT) developed by Chen et al. (2016) into our score statistic to properly handle the correlated/clustered structure among samples. We employ a resampling-based copula approach to estimate the asymptotic distribution of the proposed test, which enables time-efficient estimation of p-values.

The rest of this chapter is organized as follows. In Section 3.2, we propose a unified score test, TransMeta-Rare, under a modified random effects model to conduct rare variants association tests in trans-ethnic meta-analysis. We then evaluate the size and power of our proposed method and report results from simulation studies under different scenarios in Section 3.3. As a real data application, we further apply our proposed method to the Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia data in Section 3.4. We conclude this chapter with a discussion in Section 3.5. Supplementary texts, tables and figures are presented in Section 3.6.

3.2 Methods

Suppose one conducts a meta-analysis of K independent studies to investigate the effects of rare variants on a particular phenotype. In the kth study, a total of m_k SNPs are sequenced in a region for each of the n_k subjects. For simplicity, we assume that all variants are observed in all K studies, so that $m = m_1 = m_2 = \ldots = m_k$. We also assume that the K studies come from different ancestries. We will relax these assumptions later. Let y_{ki} be the phenotype value of the ith subject in the kth study (for $i = 1, \ldots, n_k$ and $k = 1, \ldots, K$); let $\mathbf{G}_{ki\cdot} = (g_{ki1}, \ldots, g_{kim})^T$ be the genotype vector of m variants in the region, where g_{kij} denotes the number of rare allele the ith subject carries at the jth SNP in the kth study ($g_{kij} \in \{0, 1, 2\}$); and let $\mathbf{X}_{ki}$ denote a set of q_k covariates including an intercept. For the kth study, to associate the rare variants in a region to the phenotype, we consider the linear regression model

$$y_{ki} = \mathbf{X}_{ki}^T \boldsymbol{\alpha}_k + \mathbf{G}_{ki\cdot}^T \boldsymbol{\beta}_{k\cdot} + \epsilon_{ki}, \ \epsilon_{ki} \sim N(0, \sigma_k^2) \qquad \text{(Model: Linear)}$$

for continuous traits and the logistic regression model

$$logit\mathbf{P}(y_{ki}=1)=\mathbf{X}_{ki}^T\boldsymbol{\alpha}_k+\mathbf{G}_{ki\cdot}^T\boldsymbol{\beta}_{k\cdot} \qquad \text{(Model: Logistic)}$$

for binary traits, where $\boldsymbol{\alpha}_k=(\alpha_{k1},\ldots,\alpha_{kq_k})^T$ is the vector of regression coefficients for the q_k covariates; $\boldsymbol{\beta}_{k\cdot}=(\beta_{k1},\ldots,\beta_{km})^T$ is the vector of regression coefficients for the m observed SNPs in the region; and ϵ_{ki} is an error term with a mean 0 and variance ϵ_k^2. Under both linear and logistic regression models, evaluation of no genetic associations between variants in the region and the phenotype across the K studies corresponds to testing the null hypothesis

$$H_0:\boldsymbol{\beta}_{1\cdot}=\ldots=\boldsymbol{\beta}_{K\cdot}=\mathbf{0}.$$

To construct our proposed test TransMeta-Rare, we first present the random effects model rare variants association test for a single study. We then extend the model to the meta-analysis framework.

3.2.1 The Random Effects Model for a Single Study

In this section, we summarize the random effects model for rare variants association test in a single study. For the kth study, denote $S_{kj}=\sum_{i=1}^{n_k}g_{kij}(y_{ki}-\hat{\mu}_{ki})/\hat{\phi}_k$ as the score statistic of the jth variant obtained from a linear regression model (for continuous traits) or a logistic regression model (for binary traits), where $\hat{\mu}_{ki}$ is the estimated mean of y_{ki} under either the null linear model $y_{ki}=\mathbf{X}_{ki}^T\boldsymbol{\alpha}_k+\epsilon_{ki},\ \epsilon_{ki}\sim N(0,\sigma_k^2)$ or the null logistic model $logit\mathbf{P}(y_{ki}=1)=\mathbf{X}_{ki}^T\boldsymbol{\alpha}_k$; $\hat{\phi}_k=\hat{\sigma}_k^2$ for continuous traits with $\hat{\sigma}_k^2$ being estimated under the null linear model, and $\hat{\phi}_k=1$ for binary traits.

The random effects model assumes

$$\frac{1}{w_{kj}}\beta_{kj}=\mu_j+\eta_{kj},\ \eta_{kj}\sim N(0,\tau_1), \tag{3.2.1}$$

where μ_j represents the mean genetic effect of the jth variant, η_{kj} represents the deviation of the genetic effect from μ_j in the kth study. Under this model framework, the Burden (Madsen and Browning, 2009), SKAT (Wu et al., 2011), and SKAT-O

(Lee et al., 2012) tests for ascertaining $H_0 : \boldsymbol{\beta}_{k\cdot} = 0$ in study k can all be summarized into the following score statistic:

$$U(\rho) = \mathbf{S}_{k\cdot}^T \mathbf{W}_{k\cdot}^T \cdot \mathbb{R}(\rho) \cdot \mathbf{W}_{k\cdot} \mathbf{S}_{k\cdot},$$

where $\mathbf{S}_{k\cdot} = (S_{k1}, S_{k2}, \ldots, S_{km})^T$ is the score vector of the m variants in study k, $\mathbf{W}_{k\cdot} = diag\{w_{k1}, w_{k2}, \ldots, w_{km}\}$ is a diagonal weighting matrix, and the matrix $\mathbb{R}(\rho)$ models the correlation structure of the effect sizes among the variants. The Burden test is constructed through testing $\mu_j = 0$ when fixing $\tau_1 = 0$ in the random effects model in Equation (3.2.1), which corresponds to setting $\mathbb{R}(\rho) = \mathbf{1}_m \mathbf{1}_m^T$, where $\mathbf{1}_m = (1, \ldots, 1)^T$ is an $m \times 1$ vector with all elements being 1. The SKAT test is constructed through testing $\tau_1 = 0$ when fixing $\mu_j = 0$ in Equation (3.2.1), which corresponds to setting $\mathbb{R}(\rho) = \mathbf{I}_m$, the $m \times m$ diagonal matrix. The SKAT-O test is constructed through jointly testing $\mu_j = 0$ and $\tau_1 = 0$ in Equation (3.2.1), which corresponds to $\mathbb{R}(\rho) = \rho \mathbf{1}_m \mathbf{1}_m^T + (1 - \rho)\mathbf{I}_m$, a convex combination of the correlation structures for Burden and SKAT.

3.2.2 The Random Effects Model for Meta-analyzing K Studies

We now extend the random effects model for rare variant association tests in a single study into the a meta-analysis over multiple studies. Let $\boldsymbol{\beta}_{\cdot j} = (\beta_{1j}, \beta_{2j}, \ldots, \beta_{Kj})^T$ denote the vector of regression coefficients for the jth SNP among the K independent studies, then the random effects model in Equation (3.2.1) can be written as

$$\mathbf{W}_{\cdot j}^{-1} \boldsymbol{\beta}_{\cdot j} = \mu_j \mathbf{1}_K + \boldsymbol{\eta}_{\cdot j},\ \boldsymbol{\eta}_{\cdot j} \sim MVN(\mathbf{0}, \tau_1 \mathbf{I}_K), \tag{3.2.2}$$

with $j \in \{1, \ldots, m\}$, $\mathbf{W}_{\cdot j} = diag\{w_{1j}, w_{2j}, \ldots, w_{Kj}\}$, $\mathbf{1}_K = (1, \ldots, 1)_{K \times 1}^T$, $\mathbf{I}_K$ as the identity matrix with dimension $K \times K$. To account for the expected genetic effect heterogeneity of common variants in GWAS trans-ethnic meta-analysis, Shi and Lee (2016) adapted the kernel machine framework to flexibly model the genetic effect distributions. For a given variant j, instead of treating $\{\eta_{kj}\}$ $(k \in \{1, \ldots, K\})$

as identically and independently distributed normal samples, they assumed that $\boldsymbol{\eta}_{\cdot j}$ follows a mean $\mathbf{0}$ Gaussian process with kernel function $\tau_1\Psi(\cdot,\cdot)$, where $\Psi(\cdot,\cdot)$ is a bivariate function representing the genetic similarity between two groups. In addition, they imposed a hierarchical structure in Equation (3.2.2) by treating μ_j as a random variable with distribution $N(0,\tau_2)$, which is independently distributed with $\boldsymbol{\eta}_{\cdot j}$. We adopt their modeling strategy and extend it to meta-analysis of rare variants associations in sequencing studies as follow.

Given the hierarchical random effects model

$$\mathbf{W}_{\cdot j}^{-1}\boldsymbol{\beta}_{\cdot j} = \mu_j\mathbf{1}_K + \boldsymbol{\eta}_{\cdot j}, \text{ with } \boldsymbol{\eta}_{\cdot j} \sim MVN(\mathbf{0},\tau_1\Psi),\ \mu_j \sim N(0,\tau_2),\ \mu_j \perp \boldsymbol{\eta}_{\cdot j}, \quad (3.2.3)$$

we first re-parameterize τ_1, τ_2 as $\tau_1 = \tau(1-\rho_K)$ and $\tau_2 = \tau\rho_K$, with $\tau \geq 0$ and $0 \leq \rho_K \leq 1$. τ measures the size of the average genetic effect μ_j; ρ_K reflects the level of heterogeneity among the studies at any given variant. As ρ_K approaches 1, the size of the genetic effect $\boldsymbol{\beta}_{\cdot j}$ is primarily due to the population effect μ_j, with negligible contribution from the deviation measurement $\boldsymbol{\eta}_{\cdot j}$. Conversely, the closer ρ_K approaches 0, the larger degree of variability there is among the deviation $\{\eta_{kj}\}$ $(k = 1,\ldots,K)$, and the population average effect μ_j becomes minuscule. Under such re-parameterization, the null hypothesis $H_0 : \boldsymbol{\beta}_{1\cdot} = \ldots = \boldsymbol{\beta}_{K\cdot} = \mathbf{0}$ corresponds to testing $H_0 : \mu_j = 0, \boldsymbol{\eta}_{\cdot j} = \mathbf{0}$ for any j, or equivalently, $H_0 : \tau = 0$. Thus, one can show that the score test statistic for $H_0 : \tau = 0$ is

$$U_\tau(\rho_m,\rho_K) = vec(\mathbf{S})^T \cdot \mathbf{W}^T \cdot [\mathbb{R}_K(\rho_K) \otimes \mathbb{R}_m(\rho_m)] \cdot \mathbf{W} \cdot vec(\mathbf{S}), \quad (3.2.4)$$

where $vec(\cdot)$ denotes the vectorization function with $vec(\mathbf{S}) = (\mathbf{S}_{1\cdot}^T, \mathbf{S}_{2\cdot}^T, \cdots, \mathbf{S}_{K\cdot}^T)^T$; $\mathbf{W} = diag\{\mathbf{W}_{1\cdot}, \mathbf{W}_{2\cdot}, \ldots, \mathbf{W}_{K\cdot}\}$ is a diagonal weighting matrix of the variants across K studies; $\mathbb{R}_m(\rho_m)$ and $\mathbb{R}_K(\rho_K)$ are two kernel matrices with nuisance parameters $\rho_m \in \{0,1\}$ and $0 \leq \rho_K \leq 1$; and $\otimes$ denotes the Kronecker product. In Section 3.6.1 of the Supplementary Materials, we provide detailed derivations of the score statistic $U_\tau(\rho_m,\rho_K)$ under different modeling assumptions for the kernels $\mathbb{R}_m(\rho_m)$ and $\mathbb{R}_K(\rho_K)$.

We note that $\mathbf{S}_{k\cdot} = (\mathbf{S}_{k1}, \mathbf{S}_{k2}, \cdots, \mathbf{S}_{km})^T$, the score vector of all the variants in study k, can be obtained from

$$\mathbf{S}_{k\cdot} = \mathbf{G}_{k\cdot}^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_k - \hat{\boldsymbol{\mu}}_k), \tag{3.2.5}$$

where $\mathbf{G}_{k\cdot} = (\mathbf{G}_{k1\cdot}, \mathbf{G}_{k2\cdot}, \ldots, \mathbf{G}_{kn_k\cdot})^T$ is the $n_k \times m$ genotype matrix in study k; $\boldsymbol{\Sigma}_k = \hat{\phi}_k \mathbf{I}_{n_k}$ with $\hat{\phi}_k = \hat{\sigma}_k^2$ for continuous traits ($\hat{\sigma}_k^2$ computed under the null linear regression model) and $\hat{\phi}_k = 1$ for binary traits; $\mathbf{y}_k = (y_{k1}, \ldots, y_{kn_k})^T$ is an $n_k \times 1$ vector of the observed phenotype values; and $\hat{\boldsymbol{\mu}}_k = (\hat{\mu}_{k1}, \ldots, \hat{\mu}_{kn_k})$ is an $n_k \times 1$ vector of estimated means of $\mathbf{y}_k$ under the null regression model.

We note that the two kernel matrices $\mathbb{R}_m(\rho_m)$ and $\mathbb{R}_K(\rho_K)$ model the between-variant and between-ancestry heterogeneity respectively. Specifically, $\mathbb{R}_m(\rho_m)$ models the correlation structure of the average genetic effects $\{\mu_j\}$ $(j = 1, \ldots, m)$ across the m variants in the hierarchical random effects model (Equation (3.2.3)); in contrast, $\mathbb{R}_K(\rho_K)$ models the correlation structure of the deviation $\{\eta_{kj}\}$ $(k = 1, \ldots, K)$ for the jth variant across the K studies in Equation (3.2.3). The two kernels provide us double-flexibility in controlling the dependence of genetic effects. From the test statistic construction perspective, any positive semi-definite matrix can be used as $\mathbb{R}_m(\rho_m)$ and $\mathbb{R}_K(\rho_K)$. From the modeling perspective, to properly account for relationships of the variants within and across studies, we propose several choices for $\mathbb{R}_m(\rho_m)$ and $\mathbb{R}_K(\rho_K)$ in the following sections.

3.2.2.1 The Kernel Structure $\mathbb{R}_K(\rho_K)$

Following the approaches in SKAT and SKAT-O on modeling the effect sizes of multiple variants in a single study, here, we adapt the kernel matrix $\mathbb{R}(\rho)$ from the SKAT-O test to construct $\mathbb{R}_K(\rho_K)$. Specifically, we propose to use

$$\mathbb{R}_K(\rho_K) = \rho_K \mathbf{1}_K \mathbf{1}_K^T + (1 - \rho_K)\Psi, \quad 0 \le \rho_K \le 1, \tag{3.2.6}$$

where Ψ is a kernel matrix which models the correlation structure of the deviation $\{\eta_{kj}\}$ for the jth variant across the K studies, and we provide two choices for modeling Ψ.

Choice 1. Group-wise independent kernel structure:

We first consider a simple scenario in which the deviation measurement $\boldsymbol{\eta}_{\cdot j}$s for the genetic effects are independently distributed across ancestry groups. It can be easily shown that such an assumption is equivalent to assuming $\Psi = \mathbf{I}_K$.

Choice 2. Genetic similarity kernel structure:

Rather than assuming independently distributed genetic effect deviations between studies, an alternative strategy is to consider that studies which share more similar genetic architectures can have more homogeneous genetic effects than those which consist of very disparate ancestries. Under this assumption, we propose to use the proportion of shared variants over all the target gene regions between two ancestry groups as a measure of their genetic similarity. We then accommodate the genetic similarity measure in constructing Ψ to model effect size similarity. Specifically, for two different studies k and $k^{'}$ ($k, k^{'} \in \{1, \cdots, K\}$), the corresponding element in the kernel matrix Ψ is calculated as

$$\Psi_{k,k'} = \frac{\sum_{Gene} \sum_{variant \in Gene} I(\text{the variant is observed in both study } k \text{ and } k^{'})}{\sum_{Gene} \sum_{variant \in Gene} 1}.$$

The element $\Psi_{k,k'}$ defines an overall measure of genetic similarity between any two studies based on their proportion of shared variants among all the gene regions of interest. We propose to use this kernel to model the situation that the studies' genetic effects are more homogeneous when their genetic architectures are more similar. In Sections 3.6.2 and 3.6.3 of the Supplementary Materials, we provide a mock example to illustrate how to construct the genetic similarity kernel Ψ, and justification that the proposed kernel structure is positive-definite.

3.2.2.2 The Kernel Structure $\mathbb{R}_m(\rho_m)$

To properly model the between-variant correlation structure, we adapt the kernel structure $\mathbb{R}(\rho)$ from the SKAT method. Specifically, we propose to use

$$\mathbb{R}_m(\rho_m) = \rho_m \mathbf{1}_m \mathbf{1}_m^T + (1-\rho_m)\mathbf{I}_m, \quad \rho_m \in \{0,1\}. \tag{3.2.7}$$

It can be easily shown that when setting $\rho_m = 1$, the kernel structure is equivalent to assuming that the average genetic effects μ_js are homogeneous with $\mu_1 = \mu_2 = \ldots = \mu_m \sim N(0, \tau_2)$ in Equation (3.2.3); in contrast, setting $\rho_m = 0$ is equivalent to assuming the μ_js are heterogeneous across the m variants with $\mu_1,\ \mu_2, \ldots, \mu_m \overset{i.i.d}{\sim} N(0, \tau_2)$ in Equation (3.2.3).

3.2.3 The Score Test Statistic: TransMeta-Rare

In Section 3.2.2, we proposed a score test statistic for rare variants association tests in trans-ethnic meta-analysis:

$$\begin{aligned} U_\tau(\rho_m, \rho_K) &= vec(\mathbf{S})^T \cdot \mathbf{W}^T \cdot [\mathbb{R}_K(\rho_K) \otimes \mathbb{R}_m(\rho_m)] \cdot \mathbf{W} \cdot vec(\mathbf{S}), \quad (3.2.8) \\ \text{where } \mathbb{R}_K(\rho_K) &= \rho_K \mathbf{1}_K \mathbf{1}_K^T + (1-\rho_K)\Psi,\ 0 \le \rho_K \le 1, \\ \mathbb{R}_m(\rho_m) &= \rho_m \mathbf{1}_m \mathbf{1}_m^T + (1-\rho_m)\mathbf{I}_m,\ \rho_m \in \{0,1\}. \end{aligned}$$

The parameters ρ_m and ρ_K model different aspects of heterogeneity. ρ_m models whether the population average genetic effects among the observed variants are homogeneous ($\rho_m = 1$) or not ($\rho_m = 0$); whereas ρ_K models the degree of heterogeneity for the deviation measurements by accounting for the genetic similarities between populations. Specifically, it incorporates the assumption that studies which share more similar genetic architectures can have more homogeneous genetic effects than those which consist of very disparate ancestries.

Given ρ_m and ρ_K, it can be shown that $U_\tau(\rho_m, \rho_K)$ asymptotically follows a mixture of chi-square distribution $\sum \lambda_l \chi^2_{l,1}$, where $\chi^2_{l,1}$s are independent chi-square random variables with one degree of freedom; and the λ_ls are the non-zero eigenvalues of $\Phi^{\frac{1}{2}}[\mathbb{R}_K(\rho_K) \otimes \mathbb{R}_m(\rho_m)]\Phi^{\frac{1}{2}}$, where Φ is the covariance matrix of $\mathbf{W} \cdot vec(\mathbf{S})$. For the kth study, let $\mathbf{G}_{k\cdot} = (\mathbf{G}_{k\cdot 1}, \ldots, \mathbf{G}_{k\cdot m})$ be its $n_k \times m$ genotype matrix, $\mathbf{X}_k$ be its $n_k \times q_k$ covariate matrix. Under the null linear regression model $y_{ki} = \mathbf{X}_{ki}^T \boldsymbol{\alpha}_k + \epsilon_{ki}$ or null logistic regression model $logit\mathbf{P}(y_{ki} = 1) = \mathbf{X}_{ki}^T \boldsymbol{\alpha}_k$, the the study-specific score vector $\mathbf{S}_{k\cdot} = (S_{k1}, \ldots, S_{km})^T$ has mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Phi}_k = \mathbf{G}_{k\cdot}^T \mathbf{P}_k \mathbf{G}_{k\cdot}$, where $\mathbf{P}_k = \mathbf{V}_k^{-1} - \mathbf{V}_k \mathbf{X}_k (\mathbf{X}_k^T \mathbf{V}_k^{-1} \mathbf{X}_k)^{-1} \mathbf{X}_k^T \mathbf{V}_k^{-1}$ is the projection matrix; $\mathbf{V}_k = \boldsymbol{\Sigma}_k = \hat{\sigma}_k^2 \mathbf{I}_{n_k}$ for continuous traits, and $\mathbf{V}_k = diag\{\frac{1}{\hat{\mu}_{k,1}(1-\hat{\mu}_{k,1})}, \ldots, \frac{1}{\hat{\mu}_{k,n_k}(1-\hat{\mu}_{k,n_k})}\}$ for binary traits, with $\hat{\mu}_{k,i}$ being the estimated mean of y_{ki} under the null logistic regression model. And finally, the covariance matrix Φ has the form $\boldsymbol{\Phi} = diag\{\mathbf{W}_{1\cdot} \boldsymbol{\Phi}_1 \mathbf{W}_{1\cdot}, \ldots, \mathbf{W}_{K\cdot} \boldsymbol{\Phi}_K \mathbf{W}_{K\cdot}\}$.

In practice, however, we rarely have prior information on which set of (ρ_m, ρ_K) is optimal in terms of maximizing power. Shi and Lee (2016) approached a similar problem by using the minimum p-value over a grid values of the nuisance parameter ρ as their test statistic. We adopt the same strategy here and set the test statistic as

$$T_{\text{TransMeta-Rare}} = \inf_{\rho_m \in \{0,1\}, 0 \leq \rho_K \leq 1} p_{\rho_m, \rho_K}, \tag{3.2.9}$$

where p_{ρ_m,ρ_K} is the p-value of $U_\tau(\rho_m, \rho_K)$ in Equation (3.2.8) for a given set of (ρ_m, ρ_K). We name the infimum p-value test in Equation (3.2.9) as **TransMeta-Rare**, our proposed rare variants association test for trans-ethnic meta-analysis.

TransMeta-Rare can be obtained by a simple grid search over sets of (ρ_m, ρ_K)s: given grid $\rho_m \in \{0, 1\}, 0 \leq \rho_{K_1} \leq \rho_{K_2} \leq \ldots \leq \rho_{K_v} \leq 1$, the test statistic is

$$T_{\text{TransMeta-Rare}} = \min\{p_{\rho_m=0,\rho_K=\rho_{K_1}}, \ldots, p_{\rho_m=0,\rho_K=\rho_{K_v}}, p_{\rho_m=1,\rho_K=\rho_{K_1}}, \ldots, p_{\rho_m=1,\rho_K=\rho_{K_v}}\},$$

and the optimal (ρ_m, ρ_K) set is the one which yields $T_{\text{TransMeta-Rare}}$. We observe that a dense grid of ρ_K does not necessarily improve power comparing to a coarse grid. Therefore, we suggest using $\rho_K = (0, 0.09, 0.25, 1)$ for simulations and real data analysis.

3.2.4 Adjusting for Sample Relatedness

In Section 3.2.2 and Section 3.2.3, we formulate the score vector in Equation (3.2.5) under the assumption that each of the participating study is population-based with unrelated subjects. However, when the studies are family-based or contain related individuals, the score vector in Equation (3.2.5) is no longer appropriate, since it ignores sample relatedness. As a result, the score test in Equation (3.2.9) may have inflated type I error rates if the score vector in Equation (3.2.5) were used. To appropriately model sample relatedness, we use the framework of generalized linear mixed model to incorporate an additional random effect term to account for the correlation structure among related individuals (Chen et al., 2013, 2016).

3.2.4.1 Linear Mixed Models (LMM) and Score Statistic

Following Chen et al. (2013), for study k, we consider the following LMM for continuous traits: Following Chen et al. (2013), for study k, we consider the following LMM for continuous traits:

$$y_{ki} = \mathbf{X}_{ki}^T \boldsymbol{\alpha}_k + \mathbf{G}_{ki\cdot}^T \boldsymbol{\beta}_{k\cdot} + b_{ki} + \epsilon_{ki}, \qquad \text{(Model: Linear Mixed)}$$

where $\mathbf{b}_k = (b_{k1}, \ldots, b_{kn_k})^T$ is an $n_k \times 1$ genetic effect vector for the random effects of familial correlation, and the remaining notations are as defined before. We assume the genetic effect vector $\mathbf{b}_k$ is normally distributed with mean 0 and covariance $\sigma_{kG}^2 \boldsymbol{\Omega}_k$, where $\boldsymbol{\Omega}_k$ is twice the kinship matrix of size $n_k \times n_k$ obtained from familial information only. We assume $\mathbf{b}_k$ is uncorrelated with the error term $\boldsymbol{\epsilon}_k = (\epsilon_{k1}, \ldots, \epsilon_{kn_k})^T$, which models the non-shared environmental effects. In summary,

$$\mathbf{b}_k \sim MVN(0, \sigma_{kG}^2 \boldsymbol{\Omega}_k),\ \boldsymbol{\epsilon}_k \sim MVN(0, \sigma_k^2 \mathbf{I}_{n_k}),\ \text{and } \mathbf{b}_k \perp \boldsymbol{\epsilon}_k.$$

Under those assumptions, the score statistic for $H_0 : \boldsymbol{\beta}_{k\cdot} = 0$ can be obtained as

$$\mathbf{S}_{k\cdot} = \mathbf{G}_{k\cdot}^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_k - \hat{\boldsymbol{\mu}}_k),$$

where $\boldsymbol{\Sigma}_k = \hat{\sigma}_{kG}^2 \boldsymbol{\Omega}_k + \hat{\sigma}_k^2 \mathbf{I}_{n_k}$, with $\hat{\sigma}_{kG}^2$ and $\hat{\sigma}_k^2$ estimated from the null LMM $y_{ki} =$

$\mathbf{X}_{ki}^T\boldsymbol{\alpha}_k + b_{ki} + \epsilon_{ki}$; and $\hat{\boldsymbol{\mu}}_k = (\hat{\mu}_{k1}, \dots, \hat{\mu}_{kn_k})^T$ is the estimated mean vector of $\mathbf{y}_k$ under the null LMM. It can be easily shown that under the null LMM, $\mathbf{S}_{k\cdot}$ has mean 0 and covariance matrix $\boldsymbol{\Phi}_k = \mathbf{G}_{k\cdot}^T\mathbf{P}_k\mathbf{G}_{k\cdot}$, where $\mathbf{P}_k = \mathbf{V}_k^{-1} - \mathbf{V}_k\mathbf{X}_k(\mathbf{X}_k^T\mathbf{V}_k^{-1}\mathbf{X}_k)^{-1}\mathbf{X}_k^T\mathbf{V}_k^{-1}$ is the projection matrix with $\mathbf{V}_k = \boldsymbol{\Sigma}_k = \hat{\sigma}_{kG}^2\boldsymbol{\Omega}_k + \hat{\sigma}_k^2\mathbf{I}_{n_k}$.

3.2.4.2 Logistic Mixed Models and Score Statistic

To adjust for the familial correlation with binary traits, we consider the following logistic mixed model for study k:

$$logit\mathbf{P}(y_{ki} = 1) = \mathbf{X}_{ki}^T\boldsymbol{\alpha}_k + \mathbf{G}_{ki\cdot}^T\boldsymbol{\beta}_{k\cdot} + b_{ki}. \qquad \text{(Model: Logistic Mixed)}$$

Following Chen et al. (2016), one can show that the corresponding score statistic for $H_0 : \boldsymbol{\beta}_{k\cdot} = 0$ is

$$\mathbf{S}_{k\cdot} = \mathbf{G}_{k\cdot}^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{Y}_k - \hat{\boldsymbol{\mu}}_k),$$

where $\mathbf{Y}_k$ is the "working vector" from the working null LMM $\mathbf{Y}_k = \mathbf{X}_k^T\boldsymbol{\alpha}_k + \mathbf{b}_k + \boldsymbol{\epsilon}_k$, $\boldsymbol{\epsilon}_k \sim MVN(0, \widetilde{\mathbf{W}}_k^{-1})$ with $\widetilde{\mathbf{W}}_k = diag\{\hat{\mu}_{k1}(1 - \hat{\mu}_{k1}), \dots, \hat{\mu}_{k1}(1 - \hat{\mu}_{k1})\}$; $\hat{\boldsymbol{\mu}}_k = (\hat{\mu}_{k1}, \dots, \hat{\mu}_{kn_k})^T$ is the estimated mean vector of $\mathbf{Y}_k$; and $\boldsymbol{\Sigma}_k = \widetilde{\mathbf{W}}_k^{-1} + \hat{\sigma}_{kG}^2\boldsymbol{\Omega}_k$, with $\hat{\sigma}_{kG}^2$ and $\widetilde{\mathbf{W}}_k$ estimated from the working null LMM. Similarly for a continuous trait, the covariance matrix of $\mathbf{S}_{k\cdot}$ is $\boldsymbol{\Phi}_k = \mathbf{G}_{k\cdot}^T\mathbf{P}_k\mathbf{G}_{k\cdot}$, where $\mathbf{P}_k = \mathbf{V}_k^{-1} - \mathbf{V}_k\mathbf{X}_k(\mathbf{X}_k^T\mathbf{V}_k^{-1}\mathbf{X}_k)^{-1}\mathbf{X}_k^T\mathbf{V}_k^{-1}$ is the projection matrix with $\mathbf{V}_k = \boldsymbol{\Sigma}_k = \hat{\sigma}_{kG}^2\boldsymbol{\Omega}_k + \widetilde{\mathbf{W}}_k$.

With those modified score statistic and the corresponding covariance matrices, our score test in Equation (3.2.9) now have controlled type I error rates under the presence of sample relatedness.

3.2.5 Weights and Missing Variants

Proper choice of weights can increase power in rare variant association analysis. We adopt the flexible beta density function proposed by Wu et al. (2011) as the pre-specified weights for the variants. Specifically, $w_{kj} = Beta(MAF_{kj}, 1, 25)$ for the

jth variant in the kth study. This weight function increases the contributions of rare variants to the score test while keeping decent contributions of variants with MAF $1\% - 5\%$. To compute TransMeta-Rare, we use the ancestry-specific MAFs.

Since rare variants tend to be population specific, if the jth variant is not present in the kth study, we impose a zero weight on it ($w_{kj} = 0$), or equivalently, set $S_{kj} = 0$ and $\phi_{kj'j} = 0$ for all $j^{'} \in \{1, \ldots, m\}$, where $\phi_{kj'j}$ is the $(j^{'}, j)$th element of $\mathbf{\Phi}_k$.

3.2.6 Multiple Studies from the Same Ancestry Group

In trans-ethnic meta-analysis, studies can be naturally grouped based on their ethnicities. Suppose now the K studies can be grouped into B ancestries and the bth ancestry contains K_b studies, $b \in \{1, \ldots, B\}$. Without loss of generality, we assume that the first K_1 studies belong to the first ancestry group, followed by the next K_2 studies belonging to the second ancestry group, and so forth. Let $\tilde{K}_0 = 0, \tilde{K}_b = K_1 + K_2 + \ldots + K_b$, for $b \in \{1, \ldots, B\}$. We assume that studies from the same ancestry group share the same underlying genetic effects, whereas studies from different ancestry groups have different underlying genetic effects, so that $\boldsymbol{\beta}_{k\cdot} = \boldsymbol{\beta}_{k'\cdot}$ if and only if study k and $k^{'}$ belong to the same ethnicity. Under these assumptions, the unified score test in Equation (3.2.4) becomes

$$\begin{aligned} U_\tau(\rho_m, \rho_K) &= vec(\tilde{\mathbf{S}})^T \cdot [\mathbb{R}_B(\rho_B) \otimes \mathbb{R}_m(\rho_m)] \cdot vec(\tilde{\mathbf{S}}), \qquad (3.2.10) \\ \text{where } \mathbb{R}_B(\rho_B) &= \rho_B \mathbf{1}_B \mathbf{1}_B^T + (1-\rho_B)\Psi, \ 0 \le \rho_B \le 1, \\ \mathbb{R}_m(\rho_m) &= \rho_m \mathbf{1}_m \mathbf{1}_m^T + (1-\rho_m)\mathbf{I}_m, \ \rho_m \in \{0, 1\}. \end{aligned}$$

Here, we define $vec(\tilde{\mathbf{S}}) = (\tilde{\mathbf{S}}_{1\cdot\cdot}^T, \tilde{\mathbf{S}}_{2\cdot\cdot}^T, \ldots, \tilde{\mathbf{S}}_{B\cdot\cdot}^T)^T$ with $\tilde{\mathbf{S}}_{b\cdot\cdot} = \mathbf{W}_{\tilde{K}_{b-1}+1,\cdot} \cdot \mathbf{S}_{\tilde{K}_{b-1}+1,\cdot} + \mathbf{W}_{\tilde{K}_{b-1}+2,\cdot} \cdot \mathbf{S}_{\tilde{K}_{b-1}+2,\cdot} + \ldots + \mathbf{W}_{\tilde{K}_{b_1},\cdot} \cdot \mathbf{S}_{\tilde{K}_{b_1},\cdot}$ for $b \in \{1, \ldots, B\}$, $\mathbf{1}_B$ is a $B \times 1$ vector with all elements being 1, and and Ψ is a $B \times B$ kernel.

It can be seen that the unified score test in Equation (3.2.10) first collapses the weighted study-specific score vectors in the same ancestry group and then aggregates the collapsed scores with the kernel matrices which account for the between-ancestry and between-variant relationships. Also notice that Equation (3.2.8) is a special case of Equation (3.2.10), in which $B = K$ and $K_1 = K_2 = \ldots = K_B = 1$.

3.2.7 Asymptotic Distribution Approximation

In this section, we propose a resampling-based algorithm to approximate the asymptotic null distribution of $T_{TransMeta-Rare}$. Denote t as the observed value of $T_{TransMeta-Rare}$. The p-value can be obtained as

$$
\begin{aligned}
&P(T_{TransMeta-Rare} \geq t)\\
&= P(\min\{p_{\rho_m=0,\rho_K=\rho_{K_1}}, \ldots, p_{\rho_m=0,\rho_K=\rho_{K_v}}, p_{\rho_m=1,\rho_K=\rho_{K_1}}, \ldots, p_{\rho_m=1,\rho_K=\rho_{K_v}}\} \geq t)\\
&= P(p_{\rho_m=0,\rho_K=\rho_{K_1}} \geq t, \ldots, p_{\rho_m=0,\rho_K=\rho_{K_v}} \geq t, p_{\rho_m=1,\rho_K=\rho_{K_1}} \geq t, \ldots, p_{\rho_m=1,\rho_K=\rho_{K_v}} \geq t).
\end{aligned}
$$

Since we know the marginal distribution of p_{ρ_m,ρ_K} for any given (ρ_m, ρ_K) follows a Uniform(0,1) distribution under the null hypothesis, here we adapt a re-sampling based algorithm to estimate the correlation structure among p_{ρ_m,ρ_K}s and employ the copula method to approximate the their joint distribution.

Under the null hypothesis, $\mathbf{\Phi}^{-\frac{1}{2}} \cdot \mathbf{W} \cdot vec(\mathbf{S})$ approximately follows an uncorrelated normal distribution $MVN(\mathbf{0}, \mathbf{I}_{mK})$, where $\mathbf{I}_{mK}$ is an identity matrix with dimension $mK \times mK$. Therefore, the following resampling-based algorithm can be implemented to estimate the correlation structure between the $U_\tau(\rho_m, \rho_K)$s:

Step 1: Generate n samples, say $\mathbf{u}$, from the multivariate normal distribution $MVN(\mathbf{0}, I_{mK})$. Here, we use $n = 500$ for the simulation studies and data application.

Step 2: Calculate the null scores as

$$U_\tau^0(\rho_m, \rho_K) = \mathbf{u}^T \cdot \mathbf{\Phi}^{\frac{1}{2}}\{[(1-\rho_K)\mathbf{\Psi} + \rho_K \mathbf{1}_K \mathbf{1}_K^T] \otimes [(1-\rho_m)\mathbf{I}_m + \rho_m \mathbf{1}_m \mathbf{1}_m^T]\}\mathbf{\Phi}^{\frac{1}{2}} \cdot \mathbf{u}$$

for each $(\rho_m, \rho_K) \in \{(0, \rho_{K_1}), \ldots, (0, \rho_{K_v}), \ldots, (1, \rho_{K_1}), \ldots, (1, \rho_{K_v})\}$.

Step 3: Calculate the correlation matrix $\Sigma_{2v\times 2v}$ of the null score $U_\tau^0(\rho_m, \rho_K)$s, where v is the length of the ρ_K grid.

Step 4: With the estimated correlation matrix $\Sigma_{2v\times 2v}$, we use the Gaussian copula to approximate the null joint distribution of the p_{ρ_m,ρ_K}s, which yields

$$P(T_{TransMeta-Rare} \geq t) = 1 - \mathbf{F}_{\Sigma_{2v\times 2v}}(\mathbf{F}^{-1}(1-t), \ldots, \mathbf{F}^{-1}(1-t)),$$

where $\mathbf{F}^{-1}$ is the inverse cumulative distribution function of a standard normal, $\mathbf{F}_{\Sigma_{2v\times 2v}}$ is the joint cumulative distribution of a multivariate normal with zero mean vector and covariance matrix equal to $\Sigma_{2v\times 2v}$.

When calculating the correlation matrix $\Sigma_{2v\times 2v}$ in the resampling-based algorithm, Pearson's correlation coefficient can yield unreliable estimates due to its strong dependent on the normality and homoscedasticity assumptions (Hauke and Kossowski, 2011). Instead, we use Spearman's correlation, a non-parametric version of the Pearson's correlation based on ranks of the random variables.

3.2.8 A Backward Elimination Algorithm to Order Relative Contributions of Participating Studies to Association Strength

After identifying the gene or region that is associated with the phenotype of interest, one important question to answer for further follow-up is to pinpoint those studies with true association signals. Inspired by the work of Ionita-Laza et al. (2014) for identifying rare causal variants in sequence-based study, here we present a simple backward elimination algorithm to iteratively remove relatively less important participating studies. We note that although our procedure does not narrow down a subset of the studies with true associations, it provides us the relative prioritization of the studies in driving the association strength.

Step 1: Start with a set of k studies $StudySet_{current} = \{1, 2, \ldots, k\}$ and compute the TransMeta-Rare meta-analysis p-value using the studies that are included in $StudySet_{current}$. Denote the p-value as $p_{current}$.

Step 2: Remove each of the study one at a time from the set $StudySet_{current} = \{1, 2, \ldots, k\}$. Denote the resulting set as $StudySet_{-i} = \{1, 2, \ldots, i-1, i+1, \ldots, k\}$

for $i = 1, 2, \ldots, k$, and compute the corresponding p-value p_{-i} using TransMeta-Rare.

Step 3: Identify the study j that leads to the smallest p-value, in other words, $j = argmin\{p_{-1}, p_{-2}, \ldots, p_{-k}\}$. Remove the identified study j from the participating studies and update $StudySet_{current}$ to $StudySet_{-j}$.

Step 4: Continue removing the participating studies till only one study is left.

3.3 Simulation Studies

To evaluate the performance of TransMeta-Rare, we ran a series of simulations with varying assumptions on genetic effect heterogeneity among multiple ancestries. We generated 10,000 haplotypes of length 250kb under a calibrated coalescent model using Cosi2 (Shlyakhter et al., 2014) to mimic LD patterns and MAFs observed in the European (EU), African-American (AA), Asian (AS) and African (AF) populations. For each simulated data set, we randomly selected a 3kb sub-region to generate causal variants and test for association strengths between the selected sub-region and phenotypes. We only keep rare variants with MAFs $< 1\%$. To assess type I error rate calibration and power estimation, we conducted meta-analysis of four studies with either equal or different sample sizes. Specifically, we considered an equal sample size scenario with 2,000 subjects from each of the EU, AA, AS and AF populations (8,000 samples in total), and an unequal sample sizes scenario with 2,000 subjects from each of the AA, AS and AF populations and 6,000 subjects from the EU population (12,000 samples in total).

For population-based studies, we generated the phenotypes according to the linear regression model (Model: Linear) for continuous traits and the logistic regression model (Model: Logistic) for binary traits. We set the covariates $\mathbf{X}_{ki}$ as a vector of length 2, in which the first covariate was generated from a standard normal distribution and the second covariate was generated from a Bernoulli distribution with 0.5 probability of success. The associated regression coefficient $\boldsymbol{\alpha}_k$ was set as

$\boldsymbol{\alpha}_k = (0.5, 0.5)^T$. The $\mathbf{G}_{ki\cdot}$ genotype vector contains genotypes of all causal variants and $\boldsymbol{\beta}_{k\cdot}$ is the regression coefficient vector of genetic effects for the causal variants. For continuous traits, we generated the random error ϵ_{ki}s from a standard normal distribution. For binary traits, we set the prevalence rate to be 0.05 when there is no genetic effect and there is a balanced case-control ratio in each of the four studies.

For family-based studies, we generated the phenotypes from the linear mixed model (Model: Linear Mixed) for continuous traits and the logistic mixed model (Model: Logistic Mixed) for binary traits. We assumed each family has 10 members with a pedigree10 structure (Figure 3.1). Consequently, for the equal sample size scenario, each of the four populations contained 200 families; in contrast, for the unequal sample size scenario, EU contained 600 families, whereas the remaining three populations each contained 200 families. To generate the genotypes for the family members, we carried out gene-dropping simulations (Abecasis et al., 2002) using the selected sequences from Cosi2 as founder haplotypes which propagate through the pedigree10 structure. The random effect term b_{ki} which accounts for the correlation structure among related individuals was generated from a standard normal distribution; the remaining model specifications were the same as in the population-based studies.

3.3.1 Type I Error Rates

We evaluated type I error rates of the proposed method by generating 2.5×10^7 datasets under the null model of no associations. To reduce the computational burden, we first generated 50,000 sets of genotypes for randomly selected sub-regions from the coalescent model, and then generated 500 phenotype sets for each of the genotype data sets. We evaluated the type I error rate at various nominal levels α from $10^{-3}, 10^{-4}, 10^{-5}$, to 2.5×10^{-6}, where $\alpha = 2.5 \times 10^{-6}$ corresponds to exome-wide studies of 20,000 genes. The empirical type I error rate was estimated as the proportion of p-values that are less than the nominal level α. Results of TransMeta-

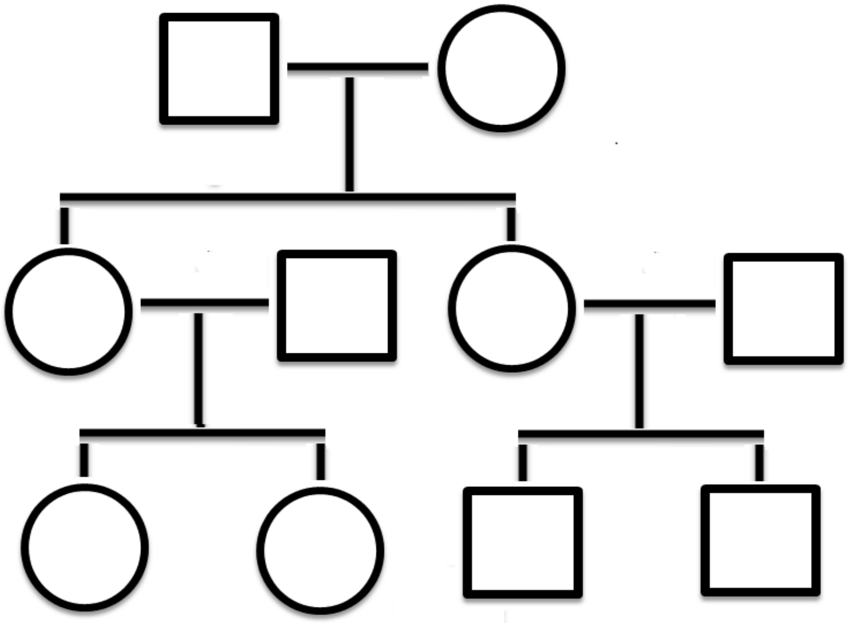


Figure 3.1: **Pedigree of families, each with 10 members, in the family-based simulation studies.**

Rare were presented in Table 3.1a and Table 3.1b for the equal sample size and unequal sample size scenarios, respectively. For both continuous and dichotomous phenotypes, regardless of the kernel used, the proposed method yielded controlled type I error rates under both the population-based and family-based study designs.

3.3.2 Power Comparisons

For power simulations, to allow the possibility of rare variants having large effects, we modeled the genetic regression coefficient $\boldsymbol{\beta}$ in the linear model and logistic model as $\beta = c|\log_{10}(MAF)|$. We considered several possible configurations of the genetic effect heterogeneity to compare performances of TransMeta-Rare with existing meta-analysis approaches. In the first scenario, we simulated a homogenous genetic effect case by assuming that the causal variants were observed in all four ancestry groups. In the second scenario, we mimiced a heterogeneous genetic effect case which assumed that the causal variants were only present in the African-American and African popu-

(a) Type I error rates of TransMeta-Rare at different α levels based on 2.5×10^7 simulations. Data are generated under the equal sample size scenario with 2,000 subjects from each of the EU, AA, AS and AF populations. The 'Genetic Similarity' and 'Indep' categories refer to the two kernel choices we provide for Ψ in Section 3.2.2.1.

Data Type	Size α	Population-based		Family-based	
		Genetic Similarity	Indep	Genetic Similarity	Indep
Gaussian	10^{-3}	9.79×10^{-4}	9.99×10^{-4}	9.86×10^{-4}	9.96×10^{-4}
	10^{-4}	1.02×10^{-4}	1.03×10^{-4}	9.75×10^{-5}	9.84×10^{-5}
	10^{-5}	1.06×10^{-5}	1.12×10^{-5}	1.09×10^{-5}	1.10×10^{-5}
	2.5×10^{-6}	2.83×10^{-6}	2.84×10^{-6}	2.85×10^{-6}	2.88×10^{-6}
Binary	10^{-3}	9.85×10^{-4}	1.01×10^{-3}	9.88×10^{-4}	9.87×10^{-4}
	10^{-4}	1.03×10^{-5}	1.03×10^{-5}	9.84×10^{-5}	9.90×10^{-5}
	10^{-5}	1.00×10^{-5}	1.02×10^{-5}	1.06×10^{-5}	1.07×10^{-5}
	2.5×10^{-6}	2.12×10^{-6}	2.12×10^{-6}	2.58×10^{-6}	2.67×10^{-6}

(b) Type I error rates at different α levels based on 2.5×10^7 simulations. Data are generated under the unequal sample size scenario with 2,000 subjects from each of the AA, AS and AF populations and 6,000 subjects from the EU population. The 'Genetic Similarity' and 'Indep' categories refer to the two kernel choices we provide for Ψ in Section 3.2.2.1.

Data Type	Size α	Population-based		Family-based	
		Genetic Similarity	Indep	Genetic Similarity	Indep
Gaussian	10^{-3}	9.87×10^{-4}	9.86×10^{-4}	1.03×10^{-3}	1.14×10^{-3}
	10^{-4}	9.82×10^{-5}	9.93×10^{-5}	1.09×10^{-4}	1.14×10^{-4}
	10^{-5}	1.10×10^{-5}	1.13×10^{-5}	1.15×10^{-5}	1.16×10^{-5}
	2.5×10^{-6}	2.73×10^{-6}	2.81×10^{-6}	2.86×10^{-6}	2.84×10^{-6}
Binary	10^{-3}	9.75×10^{-4}	9.84×10^{-4}	9.74×10^{-4}	9.75×10^{-3}
	10^{-4}	9.87×10^{-5}	9.94×10^{-5}	9.98×10^{-5}	1.09×10^{-4}
	10^{-5}	9.96×10^{-6}	9.97×10^{-6}	1.05×10^{-5}	1.03×10^{-5}
	2.5×10^{-6}	2.04×10^{-6}	2.05×10^{-6}	2.56×10^{-6}	2.65×10^{-6}

Table 3.1: **Type I error rates of TransMeta-Rare at different α levels based on 2.5×10^7 simulations.**

lations. In the third scenario, we generated another heterogeneous genetic effect case by assuming that the causal variants were only present in the Asian population. In each setting, we assumed either all causal variants were risk increasing or 80% are risk increasing and the remaining 20% were risk decreasing. We illustrated in Figure 3.2 the pool of candidate SNPs where the causal variants can be drawn from for each of those three scenarios. We assumed a spectrum of varying percentages of the causal rare variants. At any given percentage of causal variants, the expected genetic variations were assumed as 0.005 and 0.05 for the continuous and binary traits respectively. We calculated the constant c in the regression coefficient β as the one that yields the desired variation level (i.e. 0.005 for continuous traits and 0.05 for binary traits). In Table 3.7 of Supplementary Materials, we summarized the percentages of causal variants and the corresponding c values in each of the three effect size heterogeneity scenarios.

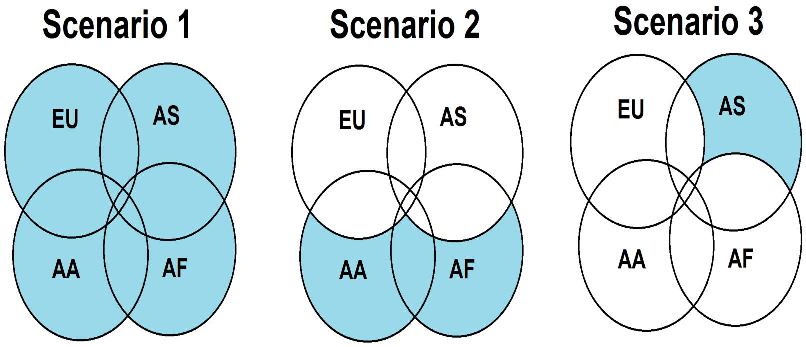


Figure 3.2: **Venn diagrams to illustrate causal variants selections.** The pool of candidate SNPs from which the causal variants are drawn from. Each circle in each Venn diagram represents the observed variants in a population, and the area colored in blue represents the pool of candidate SNPs where the causal variants can be drawn from for each of these three scenarios.

We ran 1,000 replicates to evaluate the power at the exome-wide significance level $\alpha = 2.5\times10^{-6}$. The plots in Figure 3.3 and Figure 3.4 summarized the empirical power of TransMeta-Rare as well as the competing methods for continuous traits and binary traits respectively, under the setting where all causal variants were risk increasing and the sample sizes were the same among the four ancestries with population-based studies. For the 80% risk increasing causal variants configuration, we summarized the results in Figure 3.5 and Figure 3.6 for continuous traits and binary traits respectively, in which the four populations had the same sample size with unrelated individuals.

Figure 3.3 to 3.6 showed that the performance of TransMeta-Rare varied depending on the genetic effect heterogeneity level and the percentage of causal variants. Under all scenarios, TransMeta-Rare with the genetic similarity kernel consistently achieved comparable or higher power than the group-wise independent kernel. We note that when the meta-analysis consists of one study per ancestry group, TransMeta-Rare with the group-wise independent kernel ($\Psi = \mathbf{I}_K$) is equivalent to RE-VC-O, the random-effect version of SKAT-O developed by Tang and Lin (2014). Although the two approaches have equivalent test statistics, RE-VC-O uses an adaptive Monte Carlo procedure to approximate the asymptotic distribution, which can be computationally expensive when estimating the tail probabilities.

In the first scenario where the underlying genetic effects were homogeneous across ancestries, TransMeta-Rare achieved comparable results to the most powerful competing test Hom-MetaSKAT-O by Lee et al. (2013), which assumes homogeneous genetic effects. In the two scenarios which assumed heterogeneous genetic effects across studies, TransMeta-Rare, especially with the genetic similarity kernel, outperformed the existing methods across different percentages of the causal variants. In addition, the power gain in Scenario 3 was generally higher than in Scenario 2, which was in line with the underlying genetic structures among the four group, as summarized in Table 3.2 for the genetic similarity kernel used in the simulation studies. Those values indicated that the genetic structure of Asians was very different from the remaining groups; African Americans and Africans were genetically more closely related; Europeans and African Americans had moderate genetic similarity compare

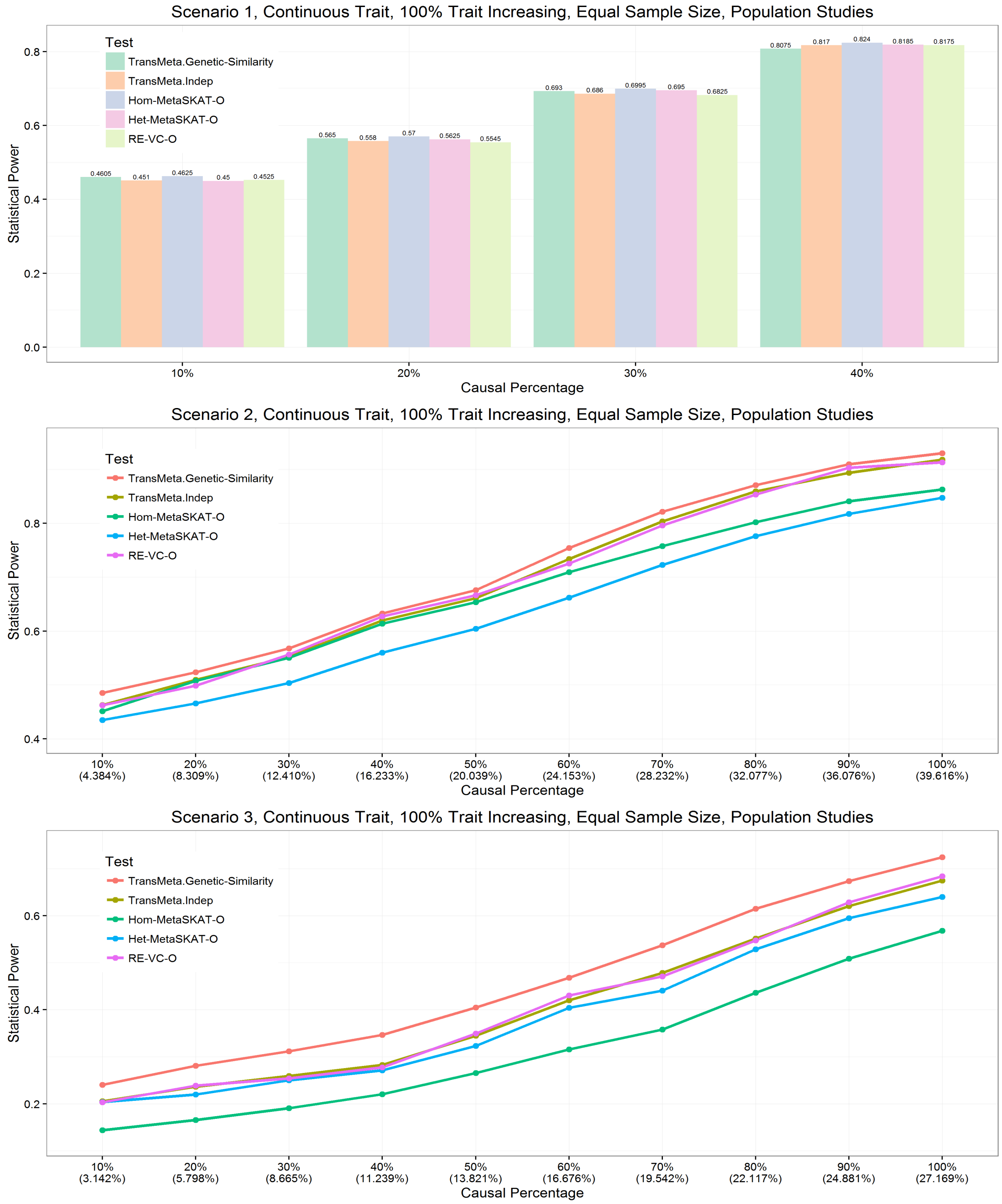


Figure 3.3: **Power comparison results for the continuous traits under different heterogeneity configurations, with population-based study design, equal sample size and all the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refers to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013); RE-VC-O refers to the method proposed by Tang and Lin (2014). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

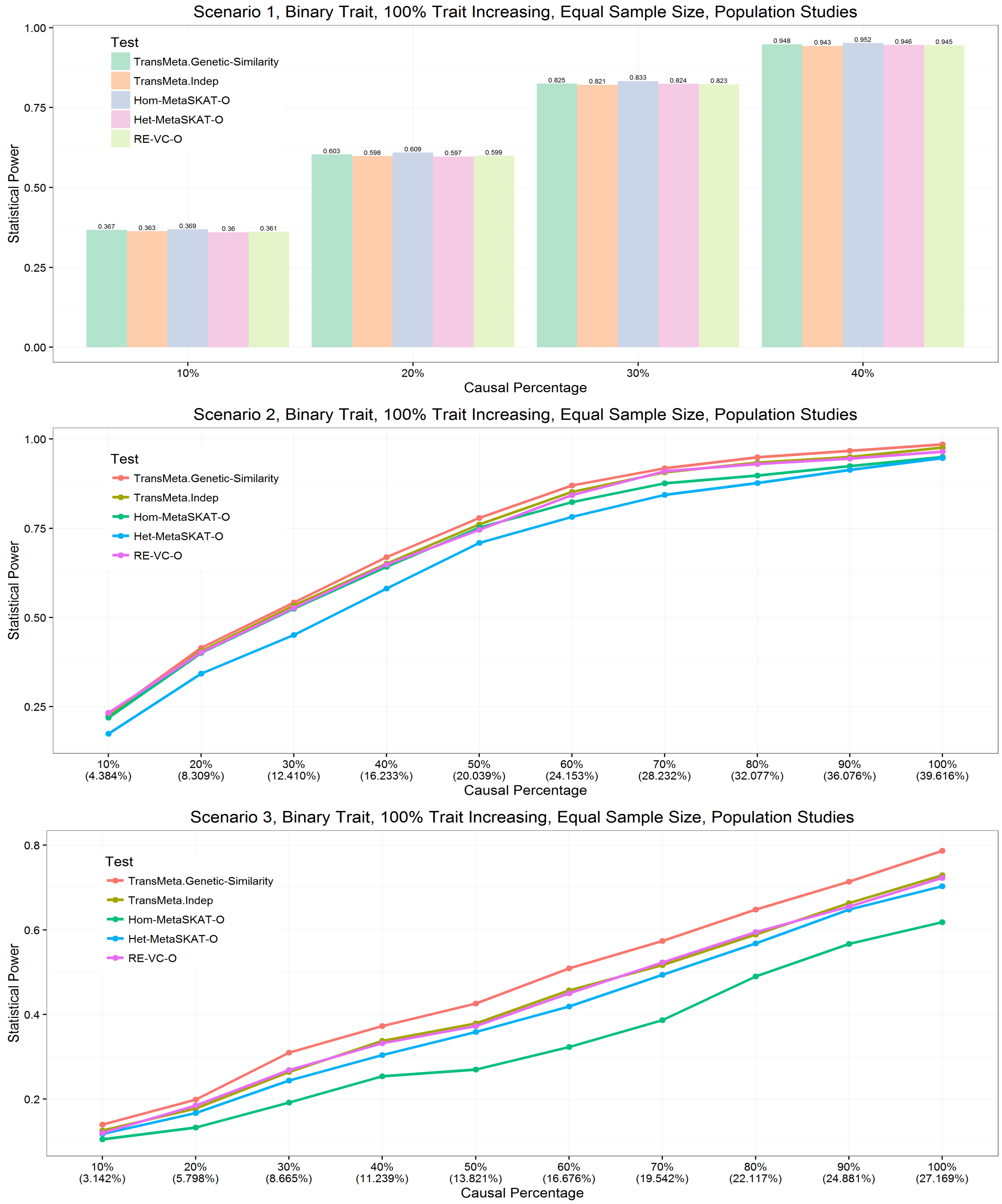


Figure 3.4: **Power comparison results for the binary traits under different heterogeneity configurations, with population-based study design, equal sample size and all the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013); RE-VC-O refers to the method proposed by Tang and Lin (2014). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

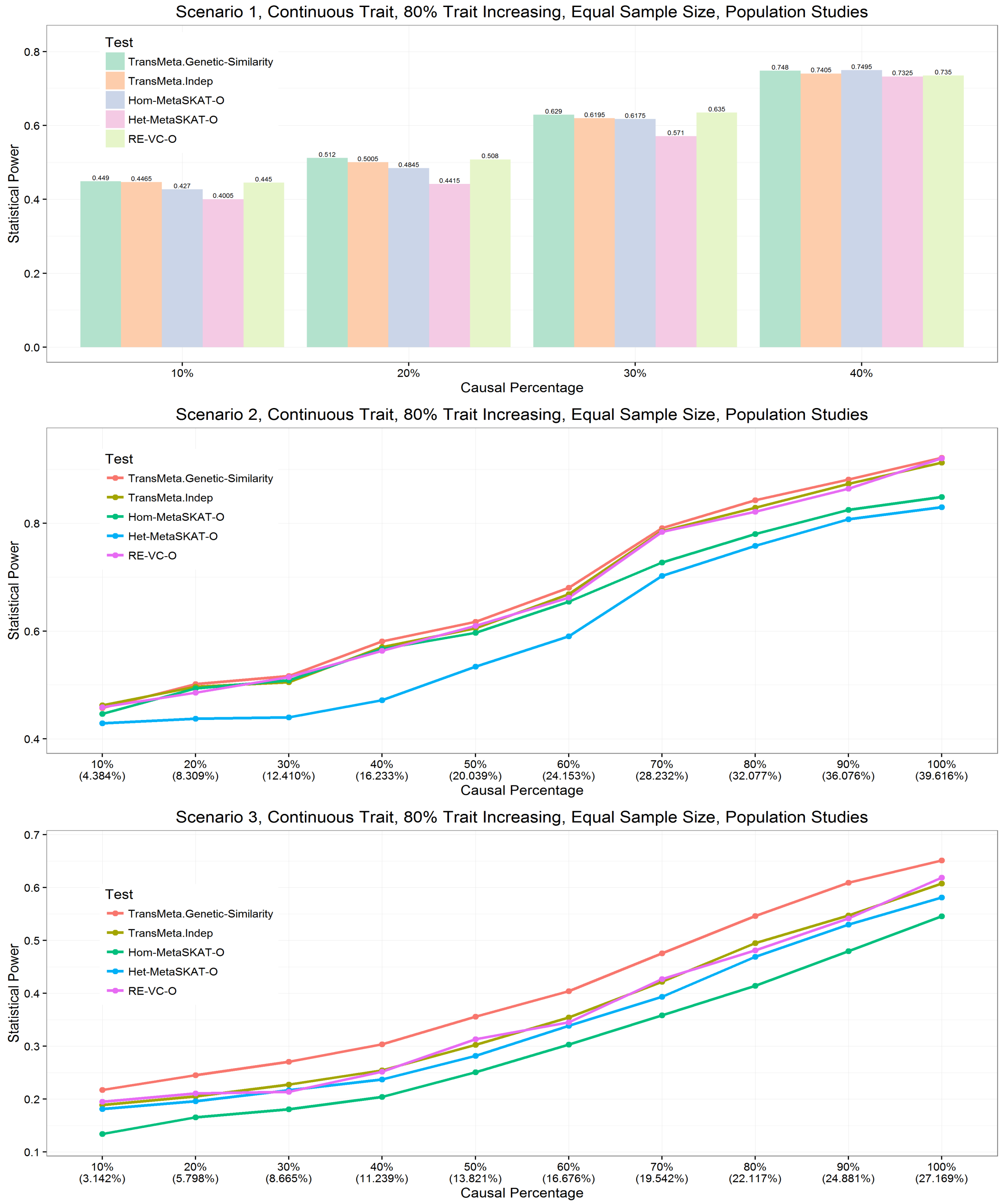


Figure 3.5: **Power comparison results for the continuous traits under different heterogeneity configurations, with population-based study design, equal sample size and 80% of the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013); RE-VC-O refers to the method proposed by Tang and Lin (2014). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

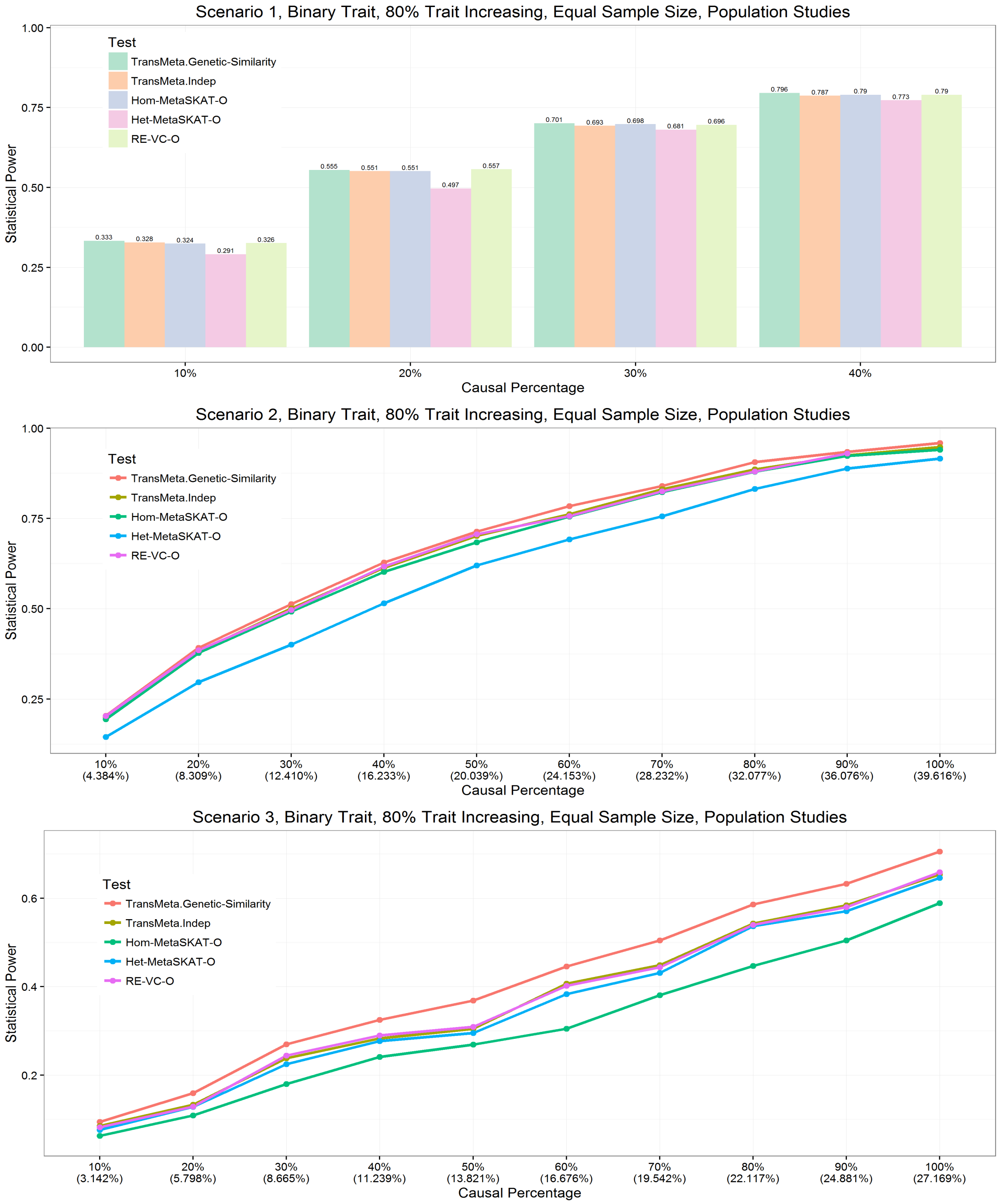


Figure 3.6: **Power comparison results for the binary traits under different heterogeneity configurations, with population-based study design, equal sample size and 80% of the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013); RE-VC-O refers to the method proposed by Tang and Lin (2014). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

to other groups. As a result, when such a genetic similarity kernel was used, the effect size heterogeneity in Scenario 3 is more consistent with the presumption of using the genetic similarity kernel for modeling the effect size heterogeneity. In contrast, although Scenario 2 assumed the genetic effect similarity between African Americans and Africans according to the similarity between their genetic architectures, it ignored the moderate genetic similarity level between African Americans and Europeans by assuming no causal variants in the European group. Consequently, we observed less power gain of TransMeta-Rare over other approaches in Scenario 2 than in Scenario 3.

Table 3.2: **The genetic similarity kernel Ψ for simulation studies.**

Ancestry	European	African American	Asian	African
European	1			
African American	0.108	1		
Asian	0.020	0.002	1	
African	0.024	0.352	0.010	1

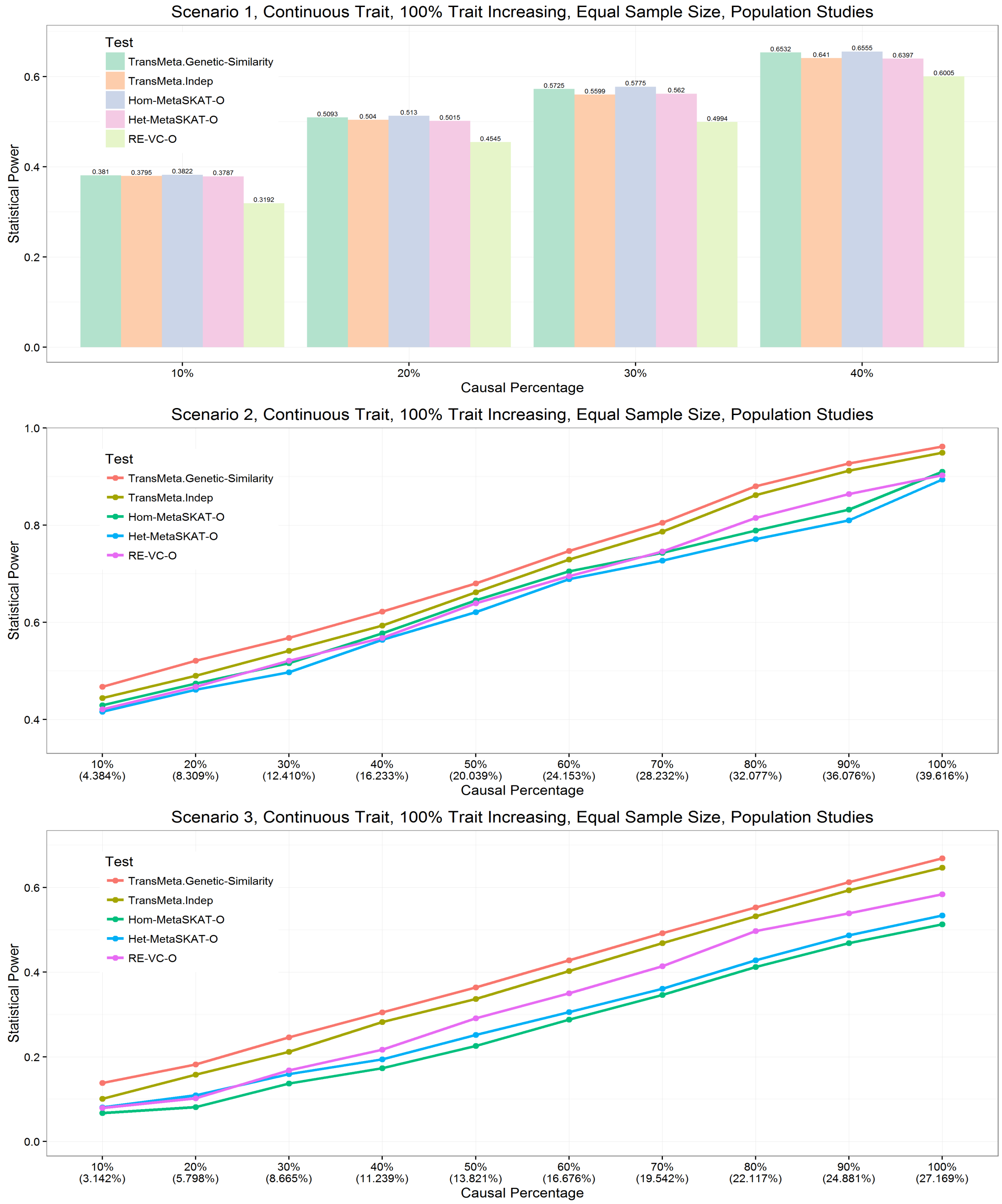
We summarized the power comparison results for the family-based studies and/or unequal sample size scenarios in Figure 3.10 to Figure 3.21 in Section 3.6.4 of the Supplementary Materials. We observed that the patterns in these supplementary figures were very similar to the results shown in Figure 3.3 to Figure 3.6. In addition, we conducted more simulations with three participating studies in each of the four groups. Figure 3.7 summarized the empirical power of TransMeta-Rare as well as the competing methods for continuous traits with three sub-studies per ancestry group, under the setting where all causal variants were risk increasing and sample sizes were the same among the four ancestries in the population-based studies. We observed that the performances had very similar patterns with the single study per ancestry group case in Figure 3.3, except that TransMeta-Rare.Indep yielded higher power than RE-VE-O. This is because rather than collapsing studies within each ancestry, RE-VC-O treated the 12 studies as if they all came from different ancestries, yielding less power than the other approaches. In addition, we observed that the power gain of TransMeta-Rare over existing approaches was even higher in the two heterogeneous configurations. The results for binary traits, family-based studies and unequal sample size scenarios were very similar to the single study per ancestry group case (data not shown).

3.3.3 Computing Time

TransMeta-Rare provides scalable computation time for gene/region-based rare variants meta-analyses. To analyze 1,000 genes/regions in the power simulations, TransMeta-Rare took 50 minutes on average on a Linux cluster node with 2.80 GHz CPU. To analyze 20,000 genes in a genome-wide dataset, TransMeta-Rare would require less than 17 hours. An R package 'TransMeta-Rare' has been developed to implement our proposed method and can be downloaded at the authors' website (https://www.leelabsg.org/).

3.4 Data Application

The Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia focused on sequencing variations that are attributable to T2D risk. T2D-GENES assembled data from a multi-ethnic sample of 12,940 unrelated individuals drawn from 5 ancestry groups: 2,350 cases and 2,168 controls of European origin; 1,012 cases and 1,152 controls of East Asian origin; 1,087 cases and 1,112 controls of South Asian origin; 1,016 cases and 9,22 controls of Hispanic origin and 1,009 cases and 1,016 controls of African American origin. The sequenced DNA samples identified 3.04 million variants, with an 82X mean coverage



Scenario 1, Continuous Trait, 100% Trait Increasing, Equal Sample Size, Population Studies

Scenario 2, Continuous Trait, 100% Trait Increasing, Equal Sample Size, Population Studies

Scenario 3, Continuous Trait, 100% Trait Increasing, Equal Sample Size, Population Studies

Figure 3.7: **Power comparison results for the continuous traits under different heterogeneity configurations, with population-based study design, three substudies per ancestry group, equal sample size and all the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta-Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013); RE-VC-O refers to the method proposed by Tang and Lin (2014). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

across the coding sequence of 18,281 genes.

We applied our proposed method to the T2D-GENES data for multi-ethnic rare variants association analysis. The phenotypes being considered included both the binary trait T2D and the continuous trait BMI. To investigate the associations between rare variants and phenotypes of interest, we employed the variant list ('mask') defined in Fuchsberger et al. (2016) based on MAFs and functional annotations for the genes. The mask was comprised of variants that were predicted to be protein-truncating and protein-altering variants with MAF $< 1\%$ that were predicted to be deleterious by at least one of the five annotation prediction algorithms: Polyphen2-HumDiv, PolyPhen2-HumVar, LRT, Mutation Taster, and SIFT.

In each of the five ancestry groups, to obtain the summary statistics such as the score test statistic and associated information matrix, we first carried out the logistic regression model in (Model: Logistic) for T2D and the linear regression model (Model: Linear) for BMI. We then meta-analyzed the five groups using TransMeta-Rare as well as other existing gene-level meta-analysis methods. For the dichotomous phenotype T2D, the adjusting covariates used in each ancestry-specific logistic regression model included age, gender, BMI and principal components calculated from the sequencing data. The principal components were included in the model to account for potential population stratification. We adjusted for the top four principal components in each of the ancestry groups. For the continuous trait BMI, we applied inverse normal transformation to BMI in order for the transformed responses to more closely approximate the Gaussian distribution.

The QQ plots in Figure 3.8a and Figure 3.8b displayed the TransMeta-Rare p-values with the genetic similarity kernel for both T2D and BMI, while the genetic similarity kernel used in the data analysis were summarized in Table 3.4. The QQ plots showed that TransMeta-Rare had controlled type I error rate for both T2D and BMI. In addition, the protein coding gene *PLCD1* (Phospholipase C Delta 1) achieved the exome-wide significance level for its association with BMI. We presented the meta-analysis p-values for this gene as well as its characteristic in Table 3.3a and Table 3.3b respectively, and the Venn Diagram in Figure 3.9 summarized the number of shared variants among the five populations for this gene. *PLCD1* encodes a member of the phospholipase C family, while phospholipases are a group of enzymes that hydrolyze phospholipids into fatty acids and other lipophilic molecules (Hu, 2011). Previous studies have shown that PLCD1 is involved in obesity. For example, the GTEx Data portal has identified the Adipose tissue specific gene expression for *PLCD1*, where increase in the number and size of adipocytes is viewed as a hallmark of obesity (www.gtexportal.org/home/eqtls/byGene?geneId=PLCD1&tissueName=All); and Hirata et al. (2011) have experimented on *PLCD1* knockout mice and observed protection from diet-induced obesity and higher metabolic rate among those *PLCD1* knockout mice through thermogenesis and adipogenesis regulation.

(a) The meta-analysis p-values of gene *PLCD1* for its association with BMI.

TransMeta-Rare. Genetic Similarity	TransMeta-Rare. Indep	Hom-MetaSKAT-O	Het-MetaSKAT-O
5.67×10^{-7}	6.06×10^{-7}	5.87×10^{-7}	7.78×10^{-7}

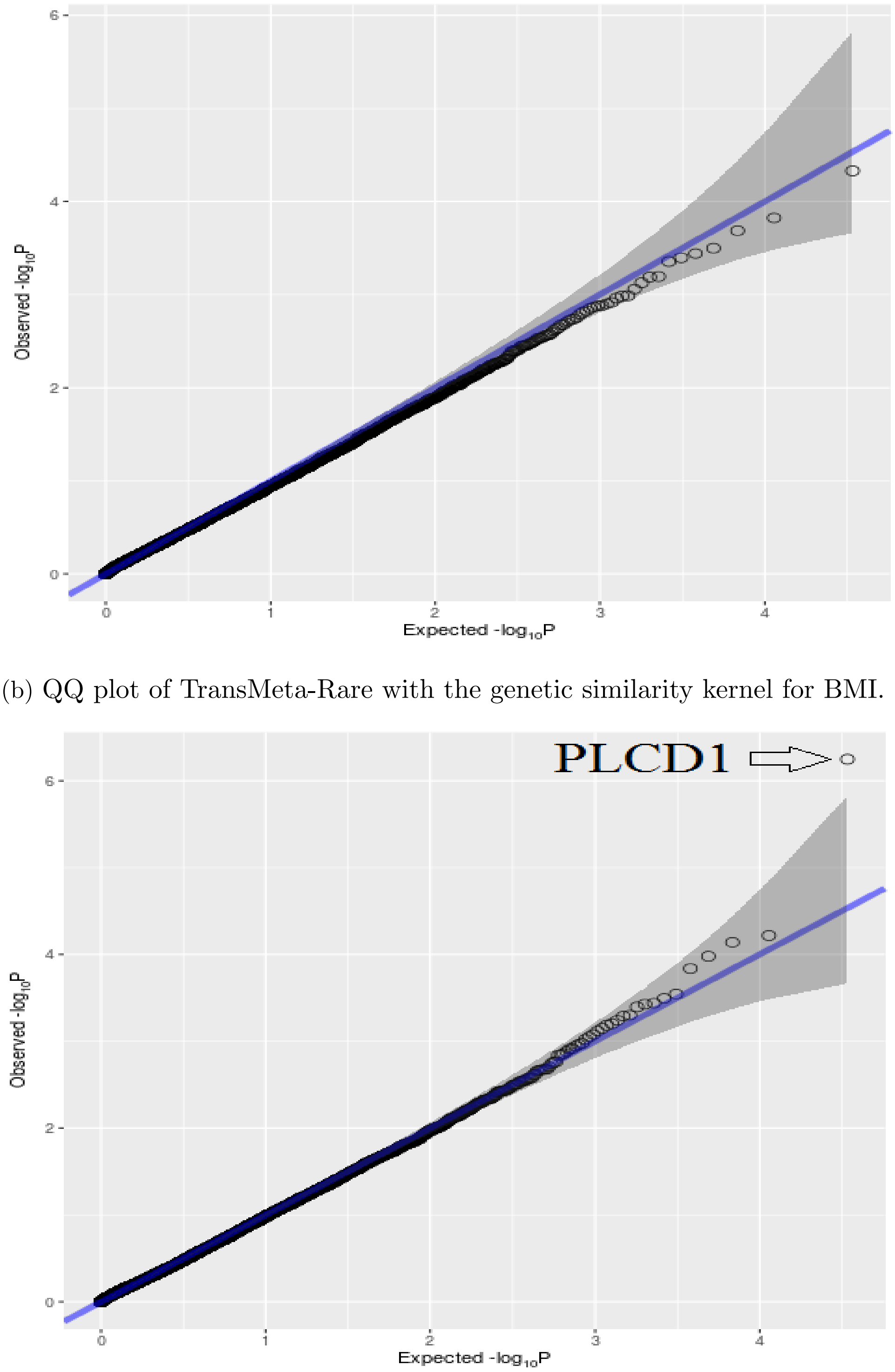
(b) Characteristics of gene *PLCD1*.

Gene	Band	Start (bp)	End (bp)	Size	♯ of SNPs				
					EA	SA	EUR	HIS	AA
PLCD1	3p22.2	38,049,337	38,065,843	16,570	17	22	21	16	24

Table 3.3: **Analysis results and characteristics of gene *PLCD1*.**

We applied the backward elimination algorithm to the *PLCD1* gene to investigate the relative contributions of the five studies in driving the association signal. The elimination results was summarized in Table 3.5. When we removed each of the five

(a) QQ plots of TransMeta-Rare with the genetic similarity kernel for T2D.



(b) QQ plot of TransMeta-Rare with the genetic similarity kernel for BMI.

Figure 3.8: **QQ plot of TransMeta-Rare with the genetic similarity kernel for T2D and BMI.** QQ-plots of -log10 p-values of TransMeta-Rare with the genetic similarity kernel. A total of 18,281 genes are tested for their associations with T2D (top panel) and BMI (bottom panel) in the T2D-GENENS data.

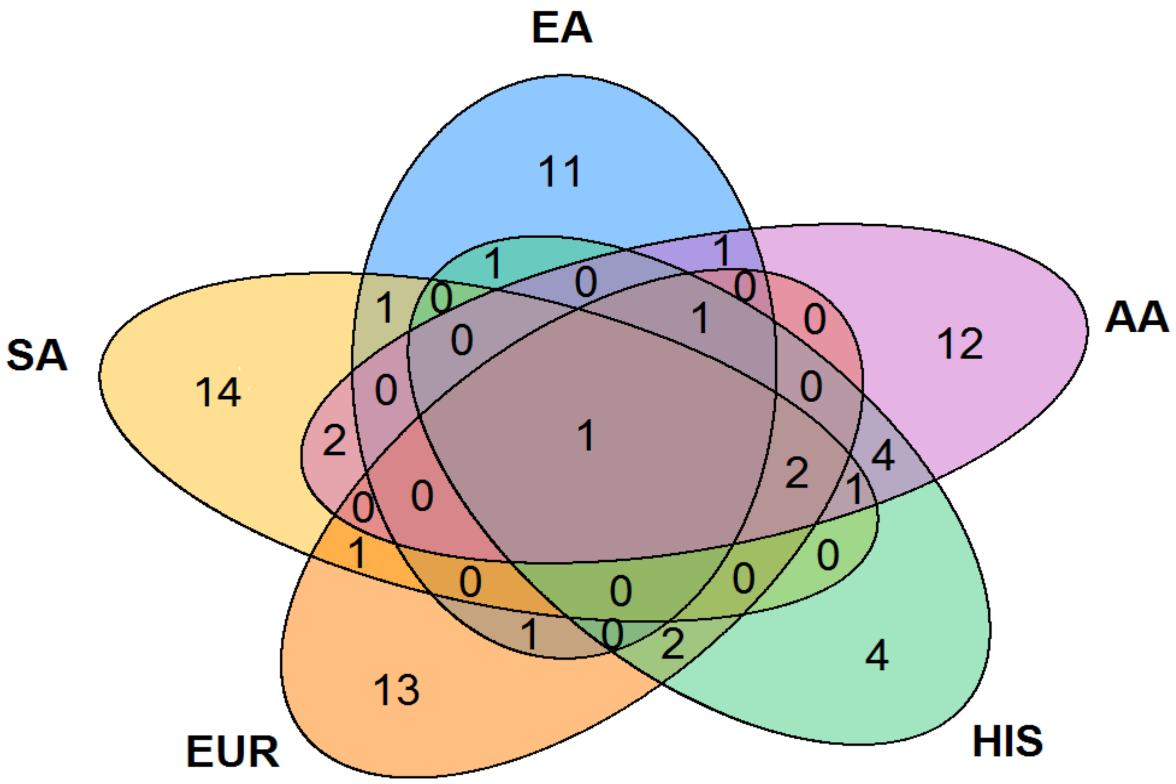


Figure 3.9: **Venn Diagram for the number of shared SNPs in gene *PLCD1* among the five ancestry groups.**

Table 3.4: **The genetic similarity kernel Ψ for T2D-GENES data.**

Ancestry	East-Asian	South-Asian	European	Hispanic	African-American
East-Asian	1				
South-Asian	0.101	1			
European	0.119	0.138	1		
Hispanic	0.095	0.111	0.163	1	
African-American	0.097	0.105	0.153	0.151	1

populations one at a time and computed TransMeta-Rare based on the remaining four populations, we observed that the association strength improves the most when SA was eliminated, with the p-value decreased from 5.67×10^{-7} to 1.50×10^{-7}. This suggested that among the five groups, SA had the least contribution to the association signal. After removing SA, HIS was the next study to be eliminated, which indicated that HIS had the next lowest contribution after SA. Further carrying out the proposed algorithm, we next removed EA and EUR in order. In other words, AA remained as the last study in the elimination sequence, suggesting that it had the strongest driving signal for the association between *PLCD1* and BMI. This was in agreement with the single study SKAT-O results (Table 3.8 in Supplementary Materials). We also applied TransMeta-Rare to each of the 72 variants in *PLCD1* to conduct the single-variant meta-analysis over the five populations (results summarized in Table 3.9, 3.10 and 3.11 in the Supplementary Materials). It can be seen from those tables that none of the single-marker meta-analysis p-values achieved genome-wide significance, which suggested that multiple variants were associated with BMI. We noted that SNP rs116413856, which yielded the smallest single-variant meta-analysis p-value, was an AA-specific variant. This result was in line with our backward elimination order, which suggested that AA was the strongest driver of the association signal for *PLCD1*.

3.5 Discussion

We have proposed a statistical framework, TransMeta-Rare, for meta-analyzing rare variant association tests in multi-ethnic samples. TransMeta-Rare incorporates the kernel regression framework to provide double-flexibility in modeling both between-ancestry and between-variant genetic effect heterogeneity. In addition, it uses the adaptive variance component test to achieve robust power regardless of degree of heterogeneity. To enable efficient approximation of the asymptotic distribution of

Table 3.5: The backward elimination sequence for gene *PLCD1*.

Removed Population	p-value		Note
	Genetic Similarity	Indep	
-	5.67×10^{-7}	6.06×10^{-7}	TransMeta-Rare p-value based on 5 populations
SA	1.50×10^{-7}	3.42×10^{-7}	TransMeta-Rare p-value based on EA, EUR, HIS and AA
HIS	1.14×10^{-6}	2.48×10^{-6}	TransMeta-Rare p-value based on EA, EUR and AA
EA	8.69×10^{-9}	3.63×10^{-8}	TransMeta-Rare p-value based on EUR and AA
EUR	8.15×10^{-5}	8.15×10^{-5}	SKAT-O p-value based on AA

the proposed method, we employed a resampling-based copula approach to estimate the p-values analytically.

In contrast to joint analysis, which requires sharing of individual-level data, the proposed gene-/region-based multi-marker test is based on study-specific summary statistics for each target region. Specifically, it only requires sharing of the single variant score statistics and the between-variant information matrix which accounts for the LD structure of the gene regions.

TransMeta-Rare can be viewed as a multi-marker extension of the modified single-variant random-effect model proposed by Shi and Lee (2016). The difference is that instead of taking the regression coefficients as input data, the rare variant association test is based on score statistics of multiple variants. There are several advantages of using score statistics instead of the regression coefficients for rare variants. First, estimation of the regression coefficients for rare variants in sequencing studies tends to be unstable with large variances, which would make the meta-analysis results unstable if the regression coefficients and the associated variance estimates are used as input data. Second, since rare variants tend to be population-specific, for those variants that are only present in some but not all studies, one can easily modify the summary statistics input by setting the score statistics as 0 for those unobserved variants in a given study.

Shi and Lee (2016) have shown that for a single variant meta-analysis, in the presence of genetic effect heterogeneity, the modified random effect framework can increase power over the traditional fixed-effect as well as the random-effect models, when heterogeneity is properly modeled. Our simulation results are consistent in this regard in multi-variant settings.

One important feature of TransMeta-Rare is that it allows for flexible modeling of the varying levels of genetic effect heterogeneity across studies, and the power simulations confirm that the proposed method can improve power over existing approaches when the genetic effect heterogeneity is properly modeled. Although TransMeta-Rare was developed to account for heterogeneous genetic effects across studies, our simulations demonstrate that when genetic effects are homogeneous across ancestries, the proposed method yields comparable results to those tests which assume homogeneous genetic effects. The T2D-GENES application suggests that TransMeta-Rare works well in practice.

The genetic similarity kernel Ψ is proposed to account for the situation where studies which share more similar genetic architectures can have more homogeneous genetic effects than those which consist of very disparate ancestries. The power simulations demonstrate that when such an underlying assumption is in line with the sources of genetic effect heterogeneity, using the genetic similarity kernel can achieve power gain over the other methods. We recommend using the genetic similarity kernel as the primary choice when fitting TransMeta-Rare. However, if there is evidence suggesting that the genetic effects are modified by non-genetic exposures such as environmental or lifestyle factors, then the group-wise independence kernel may be a better choice under such situations. To avoid data fishing, the choice regarding which kernel structure to use should be determined prior to the data analysis.

3.6 Supplementary Materials

3.6.1 Derivation of the Score Test Statistics

(Model:Linear) and (Model:Logistic) can be summarized as the following generalized linear model with a canonical link function $h(\cdot)$:

$$h[E(y_{ki})] = \mathbf{X}_{ki}^T\boldsymbol{\alpha}_k + \mathbf{G}_{ki\cdot}^T\boldsymbol{\beta}_{k\cdot}, \tag{3.6.1}$$

where $h(\cdot)$ is an identity function for the continuous traits and a logistic function for the binary traits. Based on Equation (3.6.1), the regression model over the K studies can be written as:

$$\begin{pmatrix} h[E(\mathbf{y}_1)] \\ h[E(\mathbf{y}_2)] \\ \vdots \\ h[E(\mathbf{y}_K)] \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{X}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{X}_K \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_K \end{pmatrix} + \begin{pmatrix} \mathbf{G}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{G}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \mathbf{G}_K \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_{1\cdot} \\ \boldsymbol{\beta}_{2\cdot} \\ \vdots \\ \boldsymbol{\beta}_{K\cdot} \end{pmatrix}, \tag{3.6.2}$$

and evaluation of no genetic associations between variants in the region and the phenotype across the K studies corresponds to testing the null hypothesis

$$H_0 : \boldsymbol{\beta}_{1\cdot} = \ldots = \boldsymbol{\beta}_{K\cdot} = \mathbf{0}.$$

We assume that the regression coefficients $\boldsymbol{\beta}_{\cdot j}$s for any given variant j ($j \in \{1, \ldots, m\}$) across the K studies are random variable which follow the modified random effects model

$$\mathbf{W}_{\cdot j}^{-1}\boldsymbol{\beta}_{\cdot j} = \mu_j\mathbf{1}_K + \boldsymbol{\eta}_{\cdot j},\ \boldsymbol{\eta}_{\cdot j} \sim N(\mathbf{0}, \tau_1\Psi), \tag{3.6.3}$$

where μ_j represents the average genetic effect of the jth variant over the K studies, and $\boldsymbol{\eta}_{\cdot j}$ represents the deviation of the genetic effect from μ_j across the K studies.

3.6.1.1 Meta-Analysis Assuming Homogeneous Average Genetic Effects of the Variants

If the average genetic effects μ_js are homogenous among the m variants, the modified random effects model (Equation (3.6.3)) under this assumption can be written as

$$\begin{aligned} \mathbf{W}_{\cdot j}^{-1}\boldsymbol{\beta}_{\cdot j} &= \mu_j \mathbf{1}_K + \boldsymbol{\eta}_{\cdot j},\ \boldsymbol{\eta}_{\cdot j} \sim MVN(\mathbf{0}, \tau_1\Psi) \quad \text{for } j \in \{1, \ldots, m\}, \\ \mu_1 &= \mu_2 = \ldots = \mu_m \sim N(0, \tau_2),\ \mu_j \perp \boldsymbol{\eta}_{\cdot j} \qquad\qquad \text{(Model: Hom)} \end{aligned}$$

We reparameterize τ_1, τ_2 as $\tau_1 = \tau(1 - \rho_K)$ and $\tau_2 = \tau\rho_K$, with $\tau \geq 0$ and $0 \leq \rho_K \leq 1$. τ measures the size of the average genetic effect μ_j, and ρ_K reflects the level of heterogeneity among the studies at any given variant. As ρ_K approaches 1, the size of the genetic effect $\boldsymbol{\beta}_{\cdot j}$ is primarily due to the population effect μ_j, with negligible contribution from the deviation measurement $\boldsymbol{\eta}_{\cdot j}$. Conversely, the closer ρ_K approaches 0, the larger degree of variability there is among the deviation η_{kj}s $(k = 1, \ldots, K)$, and the population average effect μ_j becomes minuscule. From this reparameterization, testing the null hypothesis of no genetic associations between any variant in the region and the phenotype corresponds to testing $H_0 : \mu_j = 0, \boldsymbol{\eta}_{\cdot j} = \mathbf{0}$ for any j, or equivalently, $H_0 : \tau = 0$.

Under (Model: Hom), the meta-analysis score test for $H_0 : \tau = 0$ for a given ρ_K is

$$U_\tau(\rho_K) = vec(\mathbf{S})^T\mathbf{W}^T \cdot [(1 - \rho_K)\Psi \otimes \mathbf{1}_m\mathbf{1}_m^T + \rho_K\mathbf{1}_K\mathbf{1}_K^T \otimes \mathbf{1}_m\mathbf{1}_m^T] \cdot \mathbf{W}vec(\mathbf{S}), \qquad (3.6.4)$$

where $\mathbf{W} = diag\{\mathbf{W}_{1\cdot}, \mathbf{W}_{2\cdot}, \ldots, \mathbf{W}_{K\cdot}\}$ is a diagonal weighting matrix of the variants across K studies; $vec(\mathbf{S})$ is a vector of aggregated study-specific scores, and $vec(\cdot)$ denotes the vectorization function with $vec(\mathbf{S}) = (\mathbf{S}_{1\cdot}^T, \mathbf{S}_{2\cdot}^T, \cdots, \mathbf{S}_{K\cdot}^T)^T$, while $\mathbf{S}_{k\cdot} = (S_{k1}, \ldots, S_{km})^T$ is the score vector of the m variants in the kth study and S_{kj} is the individual score test statistic for testing the marginal effect of the jth marker $(H_0 : \beta_{kj} = 0)$ under the study-specific linear or logistic regression model for the kth study.

3.6.1.2 Meta-Analysis Assuming Heterogeneous Average Genetic Effects of the Variants

If the average genetic effects $\mu_j s$ are heterogeneous among the m variants, the modified random effects model under this assumption can be written as

$$\begin{aligned}
\mathbf{W}_{\cdot j}^{-1}\boldsymbol{\beta}_{\cdot j} &= \mu_j \mathbf{1}_K + \boldsymbol{\eta}_{\cdot j},\ \boldsymbol{\eta}_{\cdot j} \sim MVN(\mathbf{0}, \tau_1 \Psi) \text{ for } j \in \{1, \ldots, m\},\\
\mu_1,\ \mu_2,\ \ \ldots,\ \ \mu_m &\overset{\text{iid}}{\sim} N(0, \tau_2),\ \mu_j \perp \boldsymbol{\eta}_{\cdot j} \qquad\qquad \text{(Model: Het)}
\end{aligned}$$

Applying the same reparameterization $\tau_1 = \tau(1-\rho_K)$ and $\tau_2 = \tau\rho_K$ as in Section 3.6.1.1, testing the null hypothesis of no rare variant associations is again equivalent to testing $H_0 : \tau = 0$.

Under (Model: Het), the meta-analysis score test for $H_0 : \tau = 0$ for a given ρ_K becomes

$$U_\tau(\rho_K) = vec(\mathbf{S})^T \cdot \mathbf{W}^T \cdot [(1-\rho_K)\Psi \otimes \mathbf{I}_m + \rho_K \mathbf{1}_K \mathbf{1}_K^T \otimes \mathbf{I}_m] \cdot \mathbf{W} \cdot vec(\mathbf{S}), \quad (3.6.5)$$

where $\mathbf{I}_m$ is an identity matrix with dimension m, $vec(\mathbf{S})$ and $\mathbf{W}$ have the same definitions as in Section 3.6.1.1.

3.6.1.3 The Unified Score Statistics

It can be easily seen that the derived statistics in (3.6.4) and (3.6.5) are special cases of the following unified score test

$$\begin{aligned}
U_\tau(\rho_m, \rho_K) &= vec(\mathbf{S})^T \cdot \mathbf{W}^T \cdot [\mathbb{R}_K(\rho_K) \otimes \mathbb{R}_m(\rho_m)] \cdot \mathbf{W} \cdot vec(\mathbf{S}), \quad (3.6.6)\\
\text{where } \mathbb{R}_K(\rho_K) &= \rho_K \mathbf{1}_K \mathbf{1}_K^T + (1-\rho_K)\Psi,\ 0 \le \rho_K \le 1,\\
\mathbb{R}_m(\rho_m) &= \rho_m \mathbf{1}_m \mathbf{1}_m^T + (1-\rho_m)\mathbf{I}_m,\ \rho_m \in \{0, 1\}.
\end{aligned}$$

Specifically, the unified score statistics in Equation (3.6.6) reduces to the homogeneous test in Equation (3.6.4) and the heterogeneous test in Equation (3.6.5) by setting $\rho_m = 1$ and 0 respectively.

Table 3.6: **A mock example to demonstrate how to construct the genetic similarity kernel Ψ.**

Gene	SNP	Study 1 MAF	Study 2 MAF	SNP present in both studies or not	♯ SNPs present in both studies	Total ♯ SNPs in a gene
Gene1	rs001	0.001	0	0	1	3
	rs002	0.002	0.003	1		
	rs003	0	0.004	0		
Gene2	rs004	0.005	0.006	1	2	4
	rs005	0.007	0	0		
	rs006	0	0.008	0		
	rs007	0.009	0.009	1		

3.6.2 Construction of the Genetic Similarity Kernel Ψ

In this section, we provide a mock example to illustrate how to construct the genetic similarity kernel Ψ. Recall, given two different studies k and $k^{'}$ $(k, k^{'} \in \{1, \cdots, K\})$, the corresponding element in the kernel matrix Ψ can be computed as

$$\Psi_{k,k'} = \frac{\sum_{Gene} \sum_{variant \in Gene} I(\text{the variant is observed in both study } k \text{ and } k^{'})}{\sum_{Gene} \sum_{variant \in Gene} 1},$$

where the numerator measures the number of variants that are present in both studies over all the targeted genes, and the denominator simply measures the total number of variants that are included among all the targeted genes.

We consider a mock example where the meta-analysis consists of two studies and we are interested in the associations of two target genes with a trait. We construct the genetic similarity kernel Ψ based on the observed MAFs of the two target genes in the two studies (Table 3.6).

The 5^{th} column in Table 3.6 (denoted as “SNP present in both studies or not”) measures whether each variant in the gene is observed in both studies (positive MAFs in both studies) or not (at least one study has MAF = 0). The 6^{th} column (denoted as “♯ SNPs present in both studies”) sums up the number of variants that are observed in both studies in each of the defined gene. The very last column (denoted as "Total $\sharp$ SNPs in a gene") simply counts the number of variants that are listed in each of the gene. Since we assume the meta-analysis only consists of two studies, the kernel matrix Ψ is of dimension 2×2, with the diagonal elements always be 1 (since it measures the % of shared variants over all target genes between a study and itself), and the off-diagonal element calculated as $\Psi_{1,2} = \Psi_{2,1} = (1+2)/(3+4) = 3/7$. So we obtain the kernel matrix as

$$\Psi = \begin{pmatrix} 1 & 3/7 \\ 3/7 & 1 \end{pmatrix}.$$

3.6.3 Positive Definite Matrix Ψ under the Genetic Similarity Kernel

In this section, we demonstrate that the proposed genetic similarity kernel Ψ is a positive definite matrix. We assume that the meta-analysis consists of K studies from K ancestries, and the total number of variants defined among all the genes/regions of interest is M, where $M = \sum_{Gene} \sum_{variant \in Gene} 1$. For the kth study, $k \in \{1, \ldots, K\}$, let z_k be a column vector of length M where each element has value either 0 or 1 which indicates whether the ith variant ($i \in \{1, \ldots, M\}$) is observed in the kth study or not. Let $Z = (z_1, z_2, \ldots, z_K)$ be the $M \times K$ matrix whose columns are collections of those z_k column vectors. Based on these notations, it can be easily shown that the genetic similarity kernel Ψ can be constructed as

$$\Psi = \frac{1}{M}[Z^T Z + D], \tag{3.6.7}$$

where D is a $K \times K$ diagonal matrix such that the diagonal elements of $Z^T Z + D$ are all equal to M. Let v be a non-zero column vector of length K, we now show that $Z^T Z + D$ is a positive definite matrix.

(1). D is a zero matrix

D is a zero matrix iff the M variants are observed in all K studies. In this case,

Ψ reduces to $\Psi = \frac{1}{M} Z^T Z$, with all elements in Z being 1. As a result,

$$v^T (Z^T Z) v = (Zv)^T (Zv) = ||Zv||^2 > 0,$$

since in this case at least one element in Zv has non-zero values. Consequently, $\Psi = \frac{1}{M} Z^T Z$ is positive definite.

(2). D is a diagonal matrix with at least one non-zero entry

If D has at least one non-zero entry, then $v^T D v \geq 0$ and $v^T (Z^T Z) v = ||Zv||^2 \geq 0$ would both hold. Furthermore, due to the constraint that the diagonal elements in $Z^T Z + D$ all have to equal to M, when $v^T D v = 0$, we should have $v^T (Z^T Z) v > 0$, and vice versa. Consequently, $v^T (Z^T Z + D) v > 0$ for any non-zero column vector v.

The above two cases thus suggest that $\Psi = \frac{1}{M}(Z^T Z + D)$ is positive definite.

3.6.4 Supplementary Tables and Figures

Table 3.7: **The percentages of causal variants and the corresponding c values in each of the three effect size heterogeneity scenarios.**

Data Type	Scenario	% of Causal Variants									
		10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Continuous	Scenario 1	0.467	0.324	0.263	0.228	-	-	-	-	-	-
	Scenario 2	0.714	0.474	0.379	0.326	0.292	0.264	0.243	0.227	0.217	0.203
	Scenario 3	1.154	0.827	0.665	0.577	0.519	0.469	0.431	0.403	0.379	0.364
Binary	Scenario 1	1.477	1.026	0.833	0.719	-	-	-	-	-	-
	Scenario 2	2.258	1.499	1.197	1.032	0.925	0.835	0.768	0.717	0.673	0.643
	Scenario 3	3.648	2.615	2.102	1.826	1.641	1.482	1.362	1.276	1.200	1.150

Table 3.8: **The single study multi-variant SKAT-O p-values for gene *PLCD1*.**

EA	SA	EUR	HIS	AA
1.99×10^{-2}	2.35×10^{-1}	1.55×10^{-1}	7.03×10^{-2}	8.15×10^{-5}

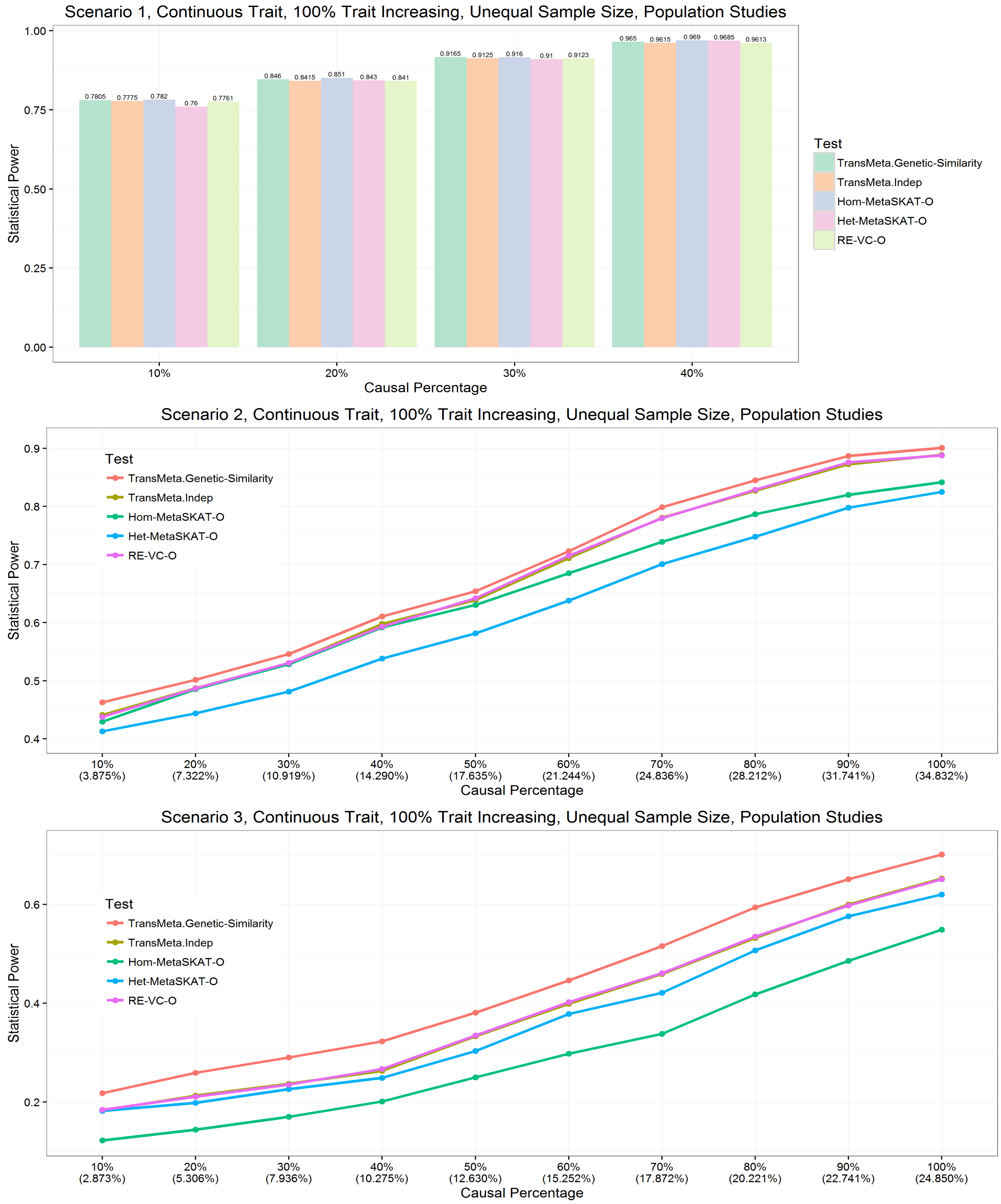


Figure 3.10: **Power comparison results for the continuous traits under different heterogeneity configurations, with population-based study design, unequal sample size and all the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013); RE-VC-O refers to the method proposed by Tang and Lin (2014). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

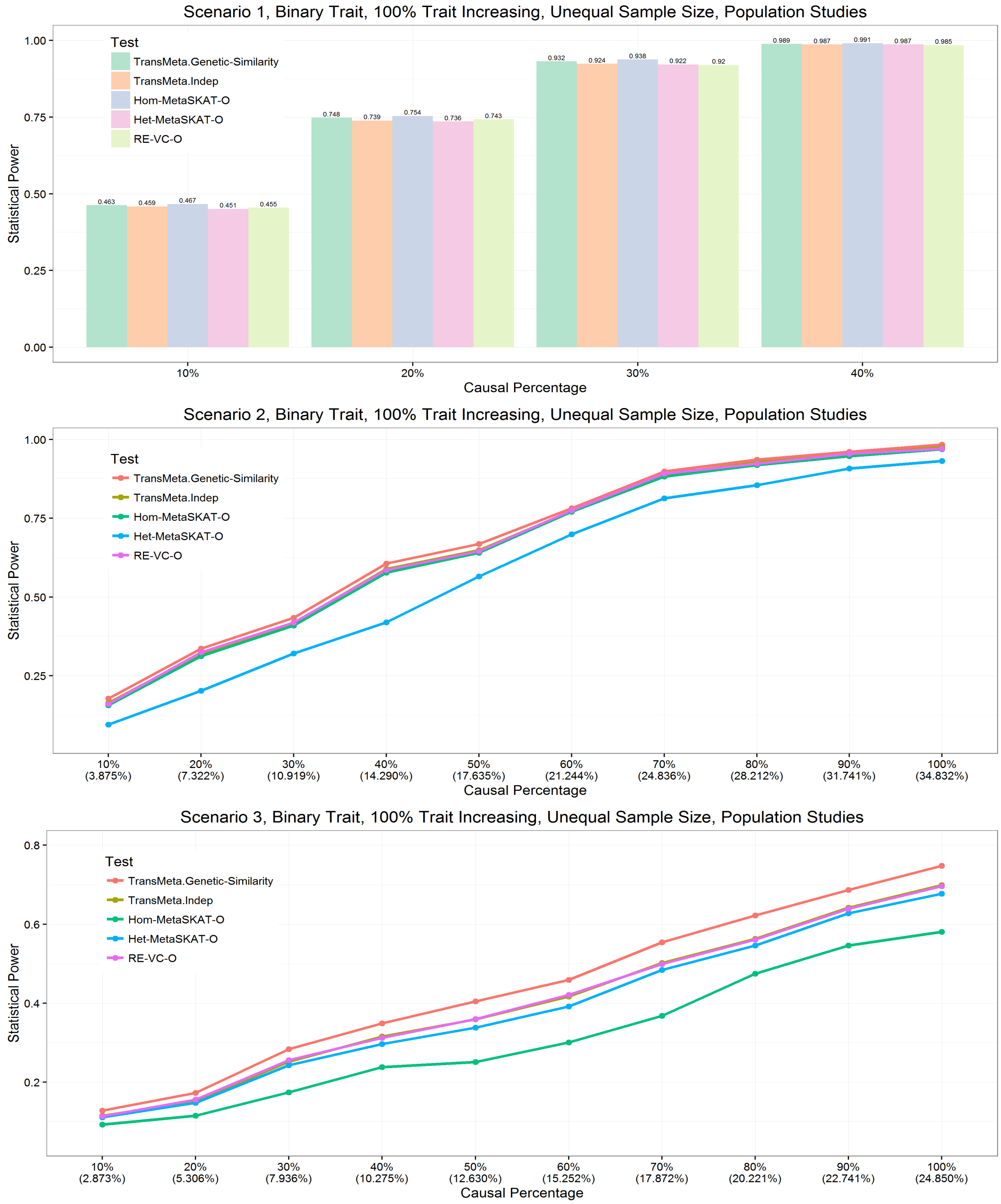


Figure 3.11: **Power comparison results for the binary traits under different heterogeneity configurations, with population-based study design, unequal sample size and all the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013); RE-VC-O refers to the method proposed by Tang and Lin (2014). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

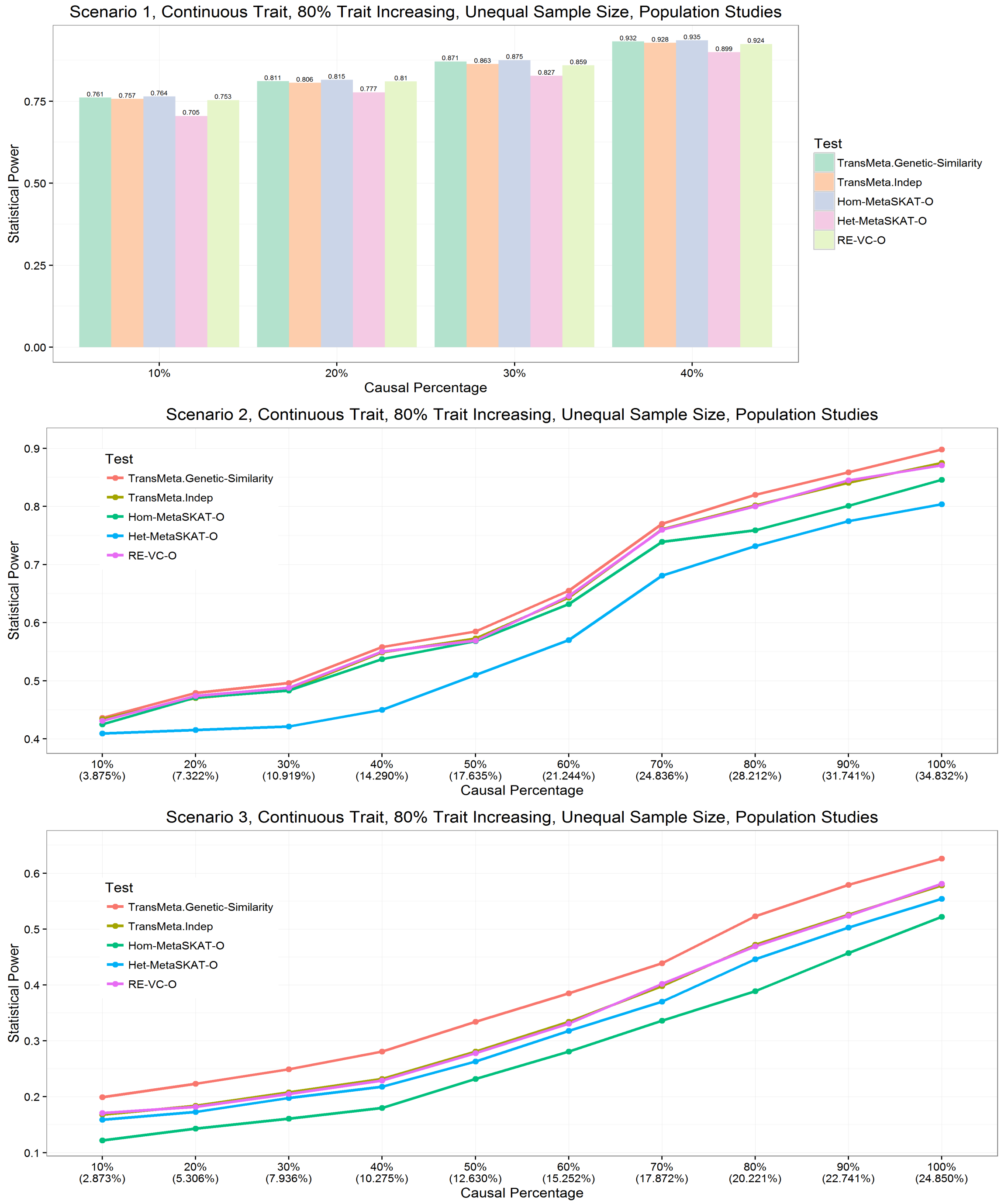


Figure 3.12: **Power comparison results for the continuous traits under different heterogeneity configurations, with population-based study design, unequal sample size and 80% of the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013); RE-VC-O refers to the method proposed by Tang and Lin (2014). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

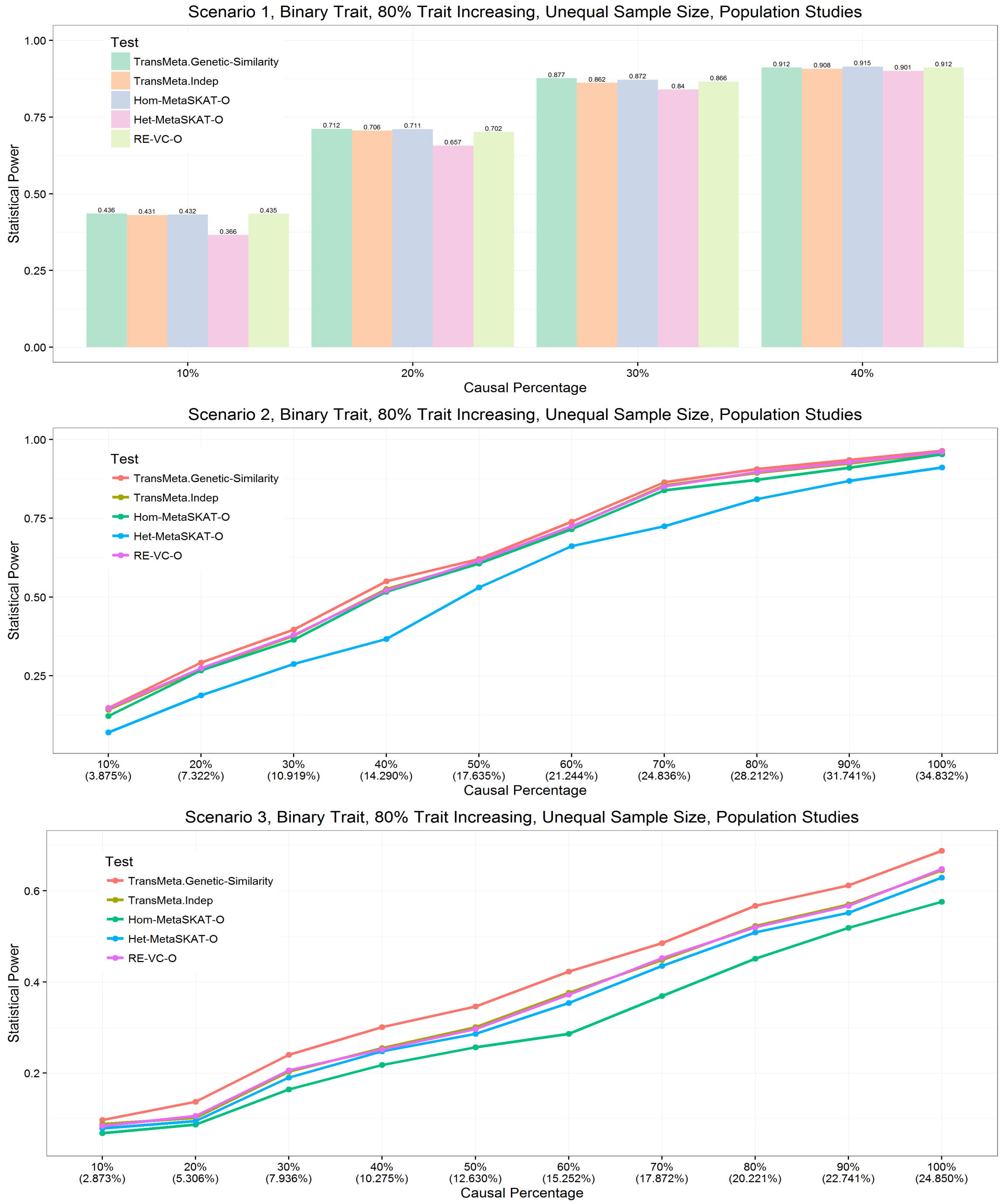


Figure 3.13: **Power comparison results for the binary traits under different heterogeneity configurations, with population-based study design, unequal sample size and 80% of the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013); RE-VC-O refers to the method proposed by Tang and Lin (2014). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

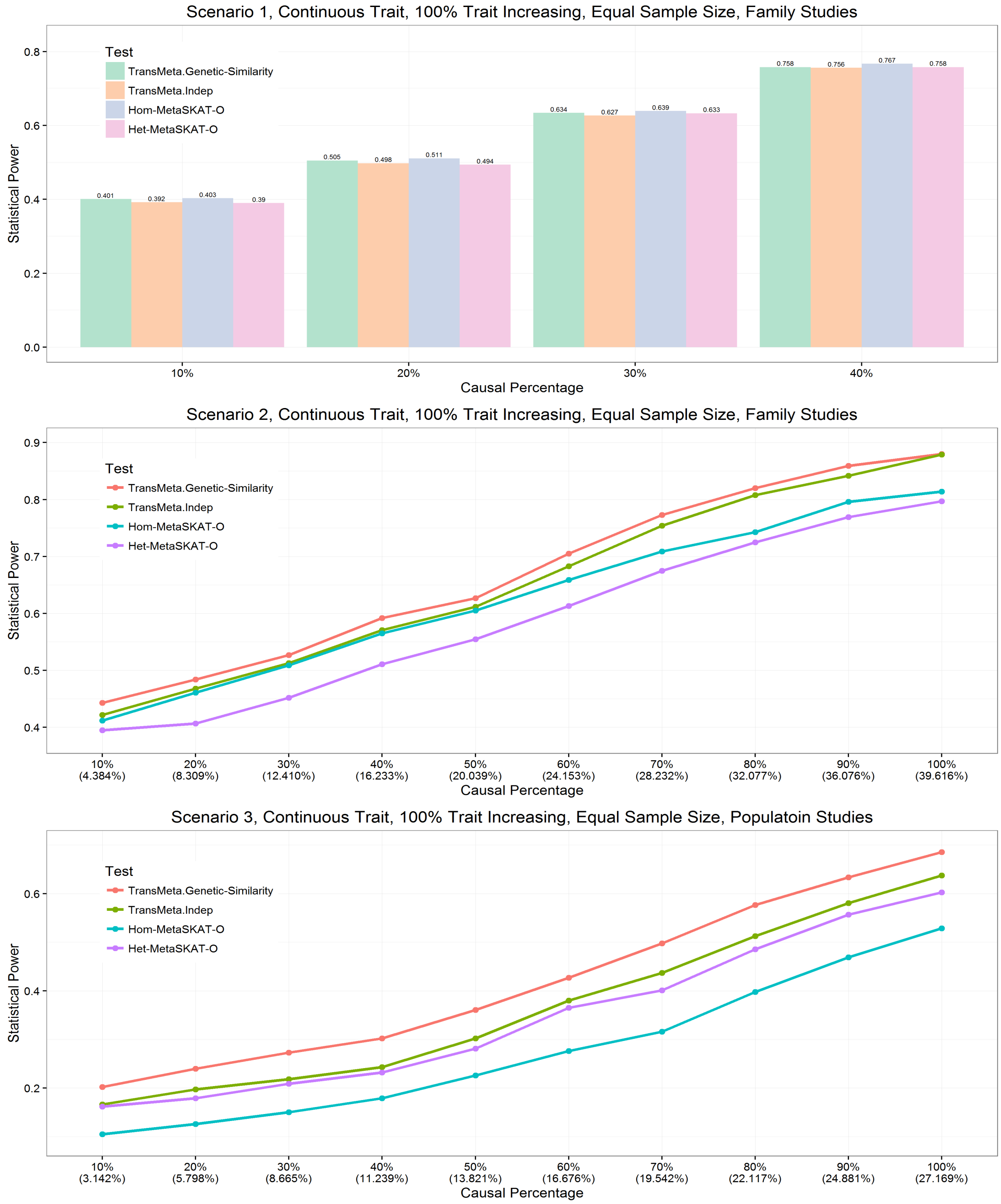


Figure 3.14: **Power comparison results for the continuous traits under different heterogeneity configurations, with family-based study design, equal sample size and all the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

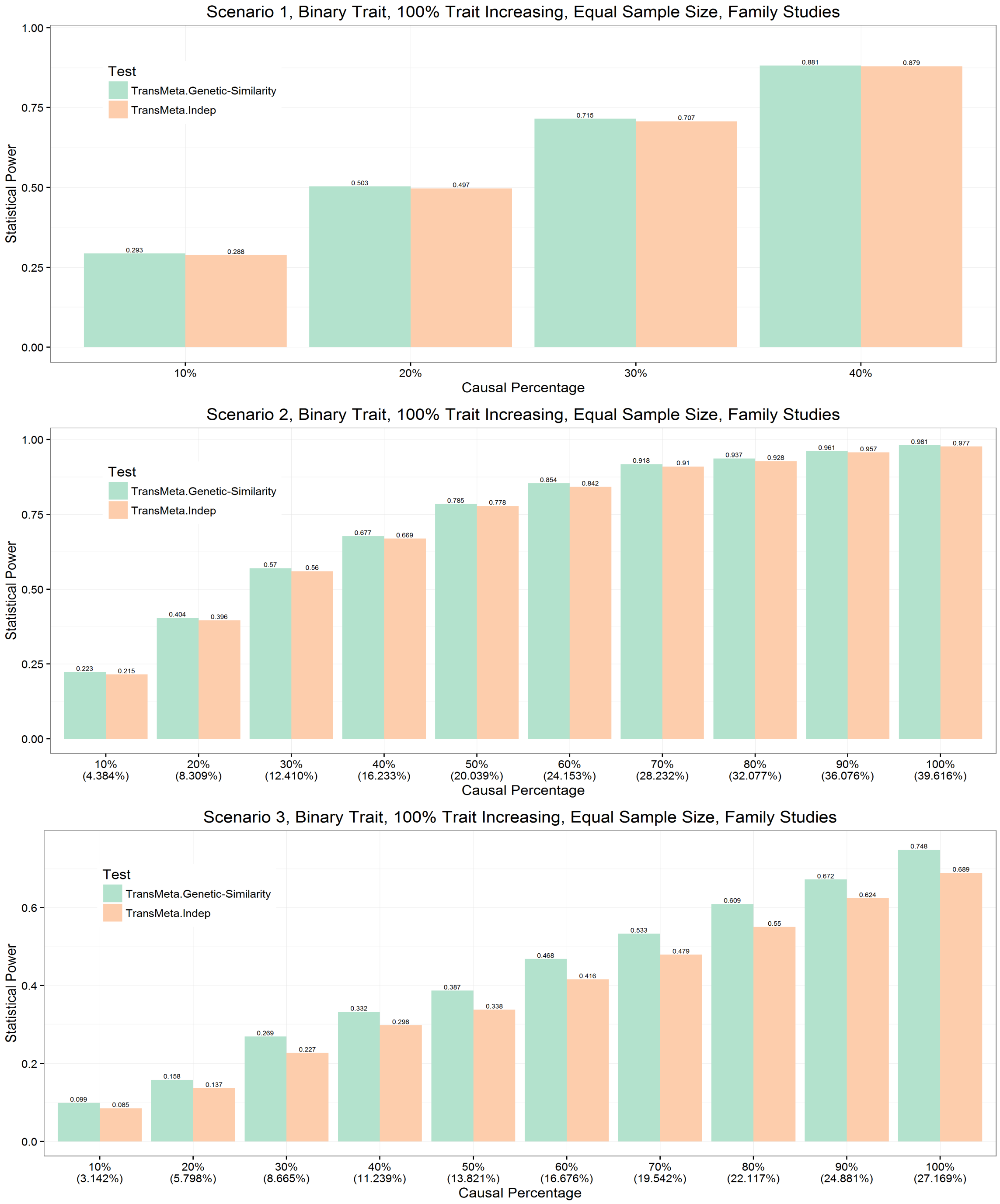


Figure 3.15: **Power comparison results for the binary traits under different heterogeneity configurations, with family-based study design, equal sample size and all the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ. For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

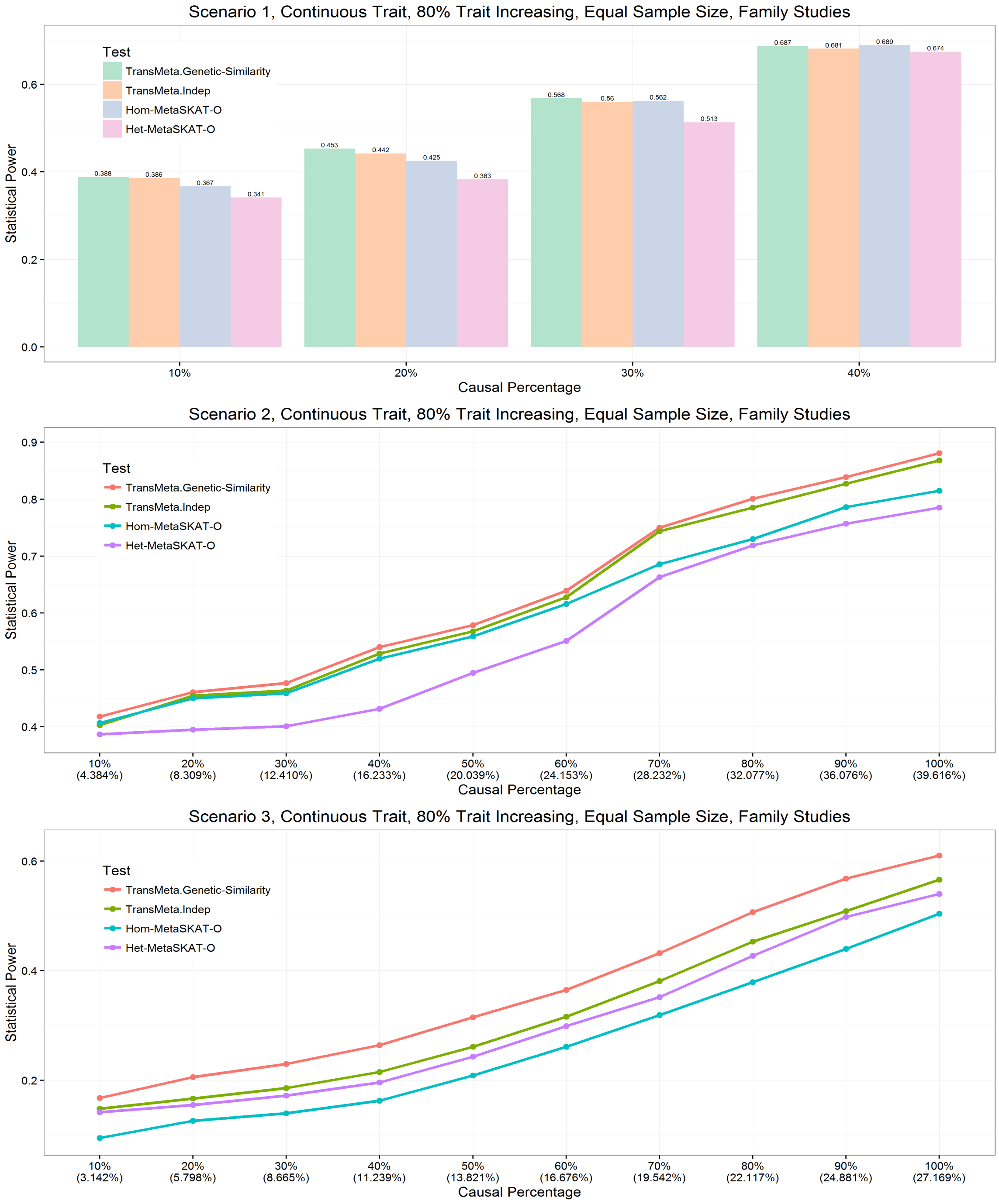


Figure 3.16: **Power comparison results for the continuous traits under different heterogeneity configurations, with family-based study design, equal sample size and 80% of the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

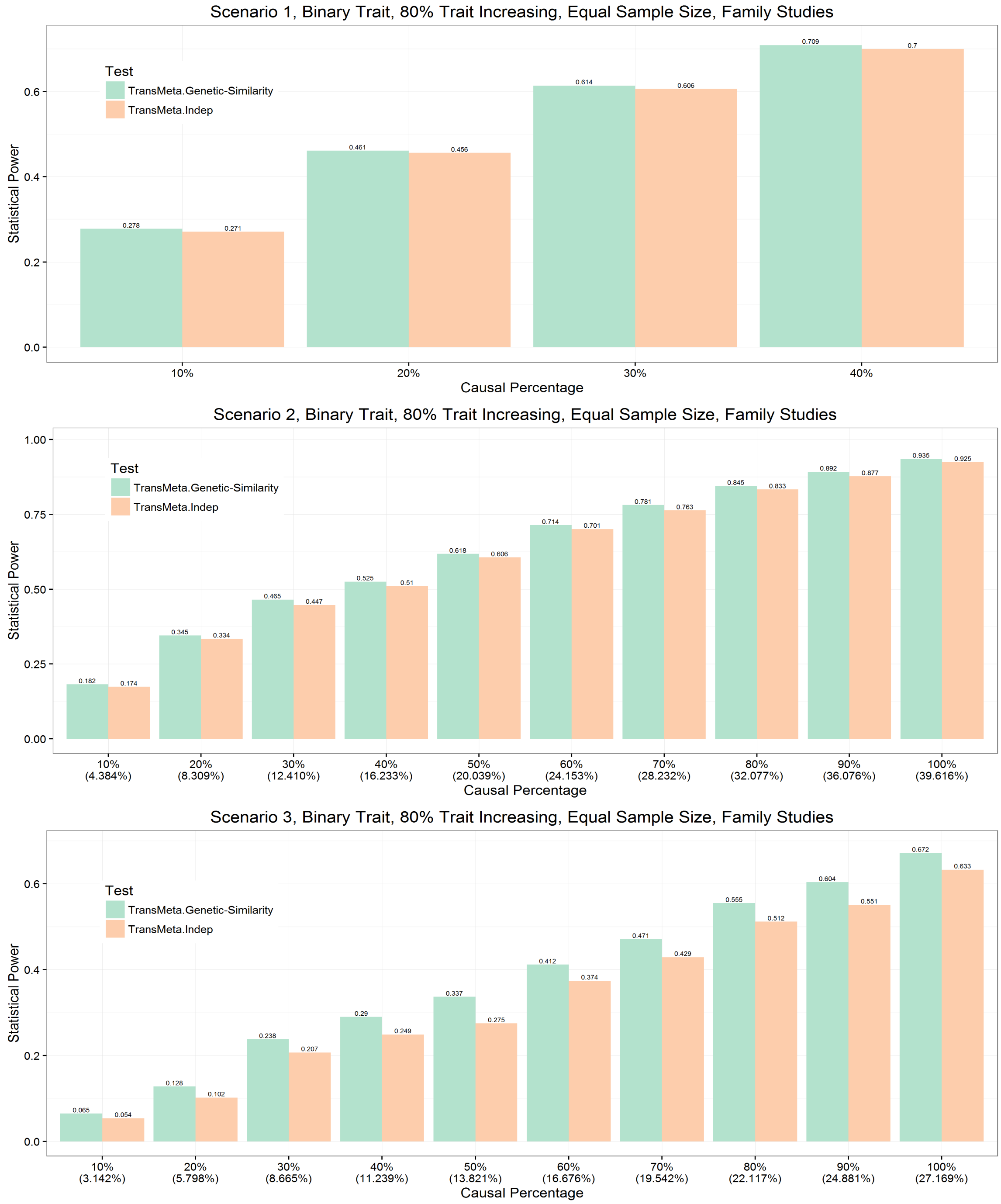


Figure 3.17: **Power comparison results for the binary traits under different heterogeneity configurations, with family-based study design, equal sample size and 80% of the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ. For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

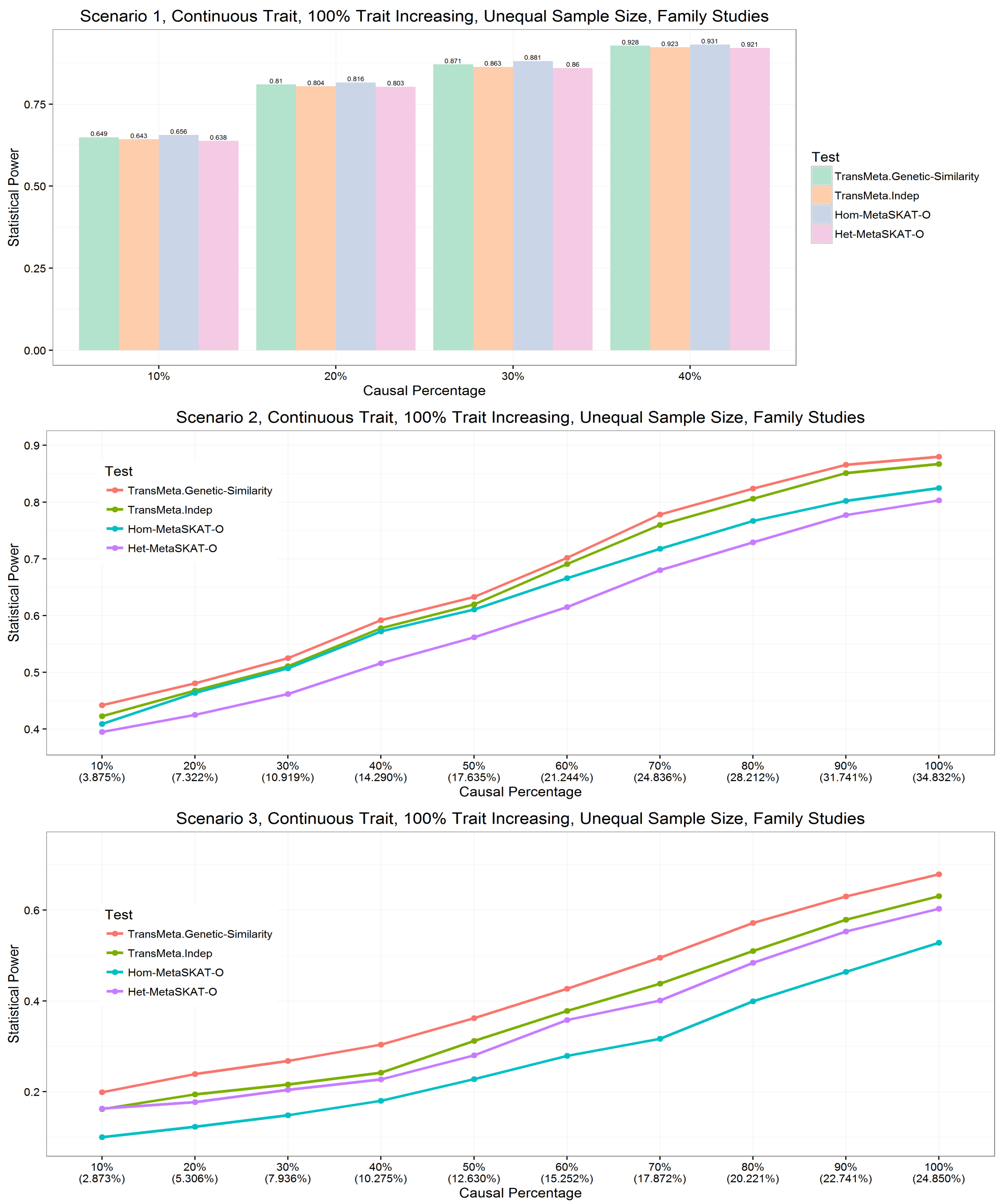


Figure 3.18: **Power comparison results for the continuous traits under different heterogeneity configurations, with family-based study design, unequal sample size and all the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

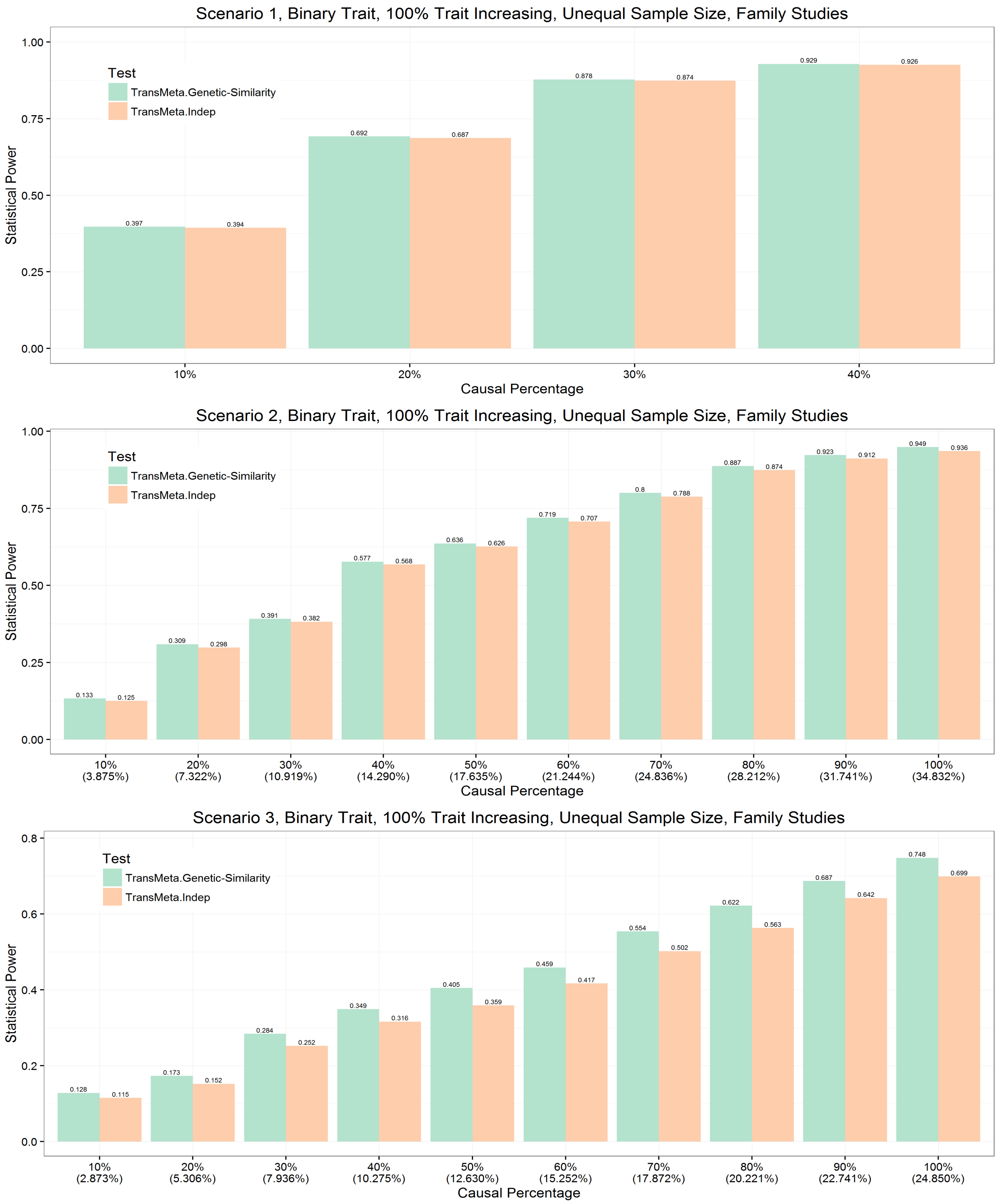


Figure 3.19: **Power comparison results for the binary traits under different heterogeneity configurations, with family-based study design, unequal sample size and all the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ. For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

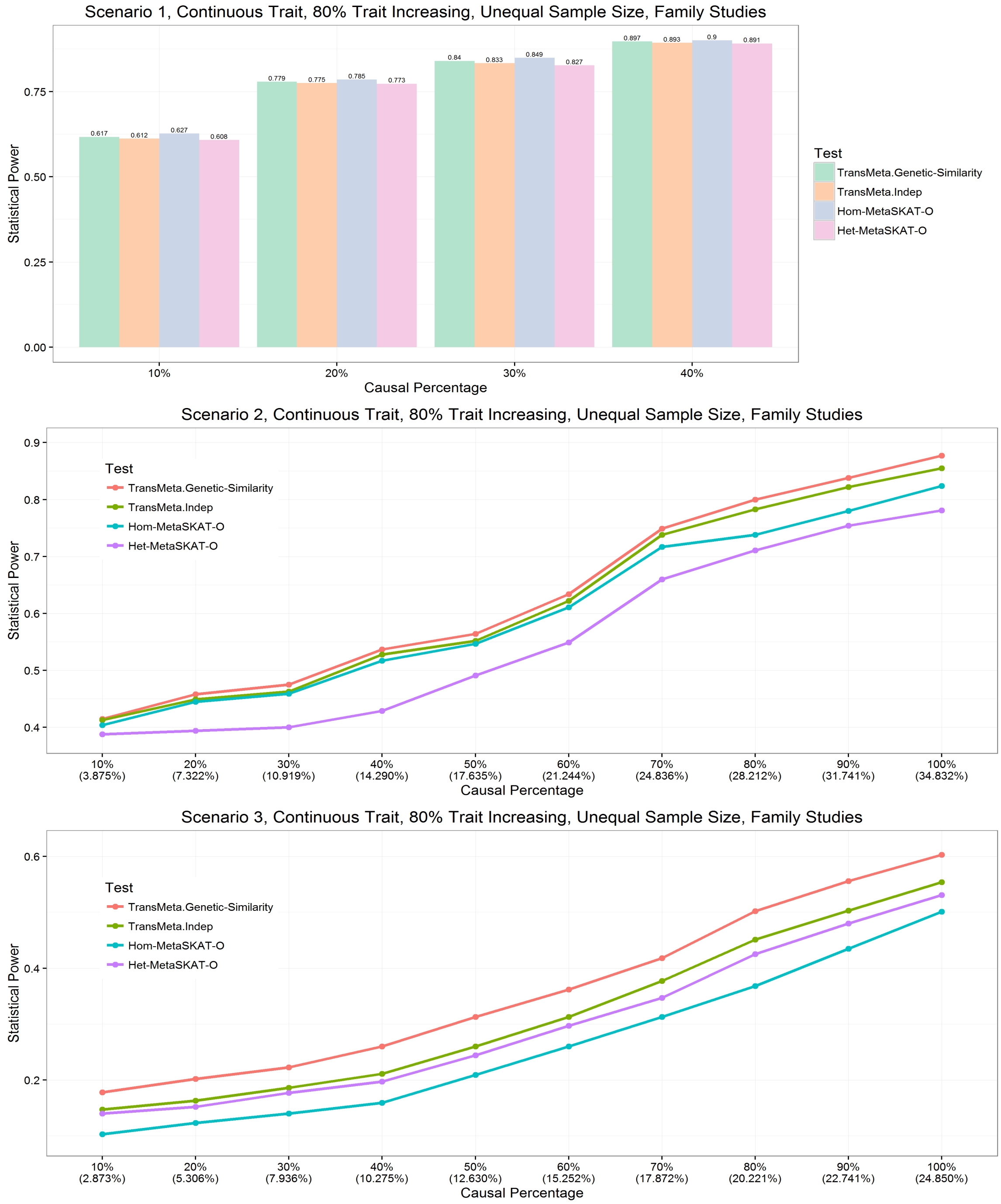


Figure 3.20: **Power comparison results for the continuous traits under different heterogeneity configurations, with Family-based study design, unequal sample size and 80% of the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ; Hom-MetaSKAT-O and Het-MetaSKAT-O refer to the methods proposed by Lee et al. (2013). For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

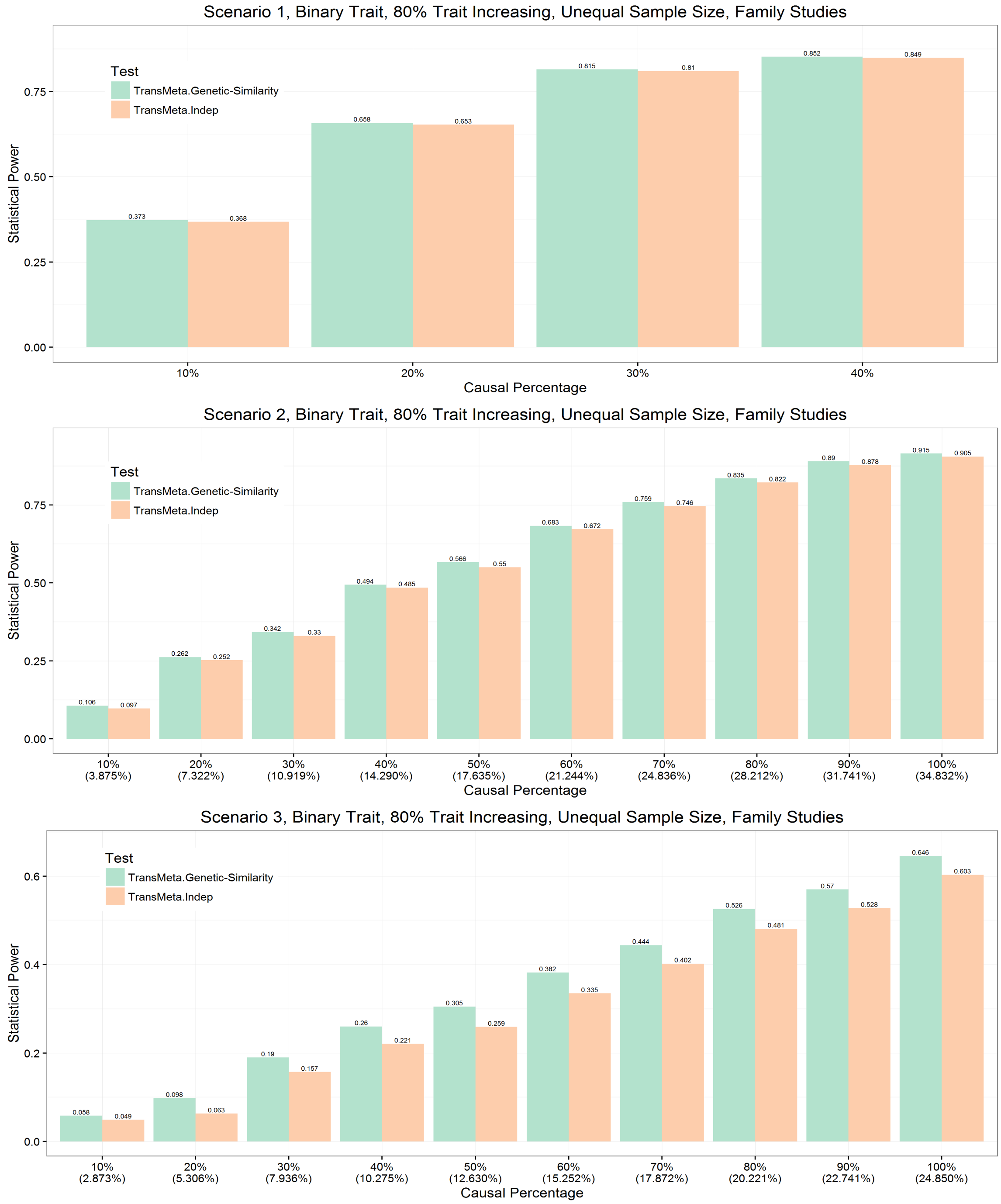


Figure 3.21: **Power comparison results for the binary traits under different heterogeneity configurations, with family-based study design, unequal sample size and 80% of the causal variants being trait increasing.** The empirical powers were evaluated at $\alpha = 2.5 \times 10^{-6}$ for the three scenarios. TransMeta.Genetic-Similarity and TransMeta.Indep refer to our proposed method TransMeta-Rare with the two kernel choices we provide for Ψ. For Scenario 2 and Scenario 3, the first line on the X-axis denotes the percentage of causal variants that are drawn from the designated populations as illustrated in Figure 3.2; while the second line on the X-axis denotes the corresponding average percentage of causal variants among the total number of variants from the four populations.

Table 3.9: **Meta-analysis p-value from TransMeta-Rare for each of the variant that is included in *PLCD1*.** The variants are sorted with an increasing order of the p-values. The ‘Genetic Similarity’ and ‘Indep’ categories refer to the two kernel choices we provide for Ψ in Section 3.2.2.1.

SNP ID	MAF ($\times 10^{-3}$)					p-value	
	EA	SA	EUR	HIS	AA	Genetic Similarity	Indep
rs116413867	0	0	0	0	2.97	4.83×10^{-3}	4.83×10^{-3}
var_3_38052863	0	0.23	0	0	0	1.19×10^{-2}	1.19×10^{-2}
var_3_38049748	0.69	0	0	0	0	1.35×10^{-2}	1.35×10^{-2}
rs141932732	1.62	2.05	1.78	1.09	10.20	1.41×10^{-2}	1.39×10^{-2}
var_3_38050599	0	0	0.77	0	0	3.44×10^{-2}	3.44×10^{-2}
var_3_38051432	0	0	0.22	0	0	5.90×10^{-2}	5.90×10^{-2}
var_3_38050895	0	0.23	0	0	0	7.63×10^{-2}	7.63×10^{-2}
rs142059541	0	0	0	0.27	0.25	7.63×10^{-2}	7.73×10^{-2}
var_3_38052047	0	0	0	3.26	0	0.12	0.12
var_3_38061760	0	0.23	0	0	0	0.12	0.12
rs78426951	0.23	0	2.89	1.63	0.99	0.15	0.15
rs115366708	0	0	0	0	1.98	0.15	0.15
var_3_38051233	0	0	0	0	0.74	0.17	0.17
var_3_38052764	0	0	0.11	0	0	0.17	0.17
var_3_38051263	0	0.23	0.11	0	0	0.17	0.17
var_3_38051630	0	0	0.11	0	0	0.22	0.22
var_3_338052849	0.23	0	0	0	0	0.22	0.22
var_3_38051475	0	0	0	0.54	0.25	0.24	0.24
var_3_38051293	0	4.10	0	0	0	0.24	0.24
var_3_38049848	2.32	0	0	0	0	0.24	0.24
var_3_38053121	0	0.23	0	0	0	0.27	0.27
var_3_38050885	0	0	0	0.54	0	0.30	0.30
var_3_38051537	0	0	0.11	0.27	0	0.32	0.32
var_3_38051657	0	0	0	0	0.25	0.32	0.32
var_3_38049598	0	0	0.11	0	0	0.33	0.33
var_3_38053082	0.23	0	0	0	0	0.37	0.37
var_3_38051543	0	0.23	0	0	0	0.37	0.37
var_3_38053165	0.23	0	0	0	0	0.37	0.37
var_3_38049533	0.46	1.37	0	0	0	0.38	0.38
var_3_38049337	0	0	0	0	0.25	0.39	0.39

Table 3.10: **Table 3.9 Continued. Meta-analysis p-value from TransMeta-Rare on each of the variant that is included in *PLCD1*.** The variants are sorted with an increasing order of the p-values. The 'Genetic Similarity' and 'Indep' categories refer to the two kernel choices we provide for Ψ in Section 3.2.2.1.

SNP ID	MAF ($\times 10^{-3}$)					p-value	
	EA	SA	EUR	HIS	AA	Genetic Similarity	Indep
var_3_38058126	0	0.68	0.44	0.54	0.49	0.40	0.38
rs139342994	0	0	0.88	0	0	0.44	0.44
var_3_38050620	0.46	0	0	0	0	0.45	0.45
var_3_38049798	0	0.23	0	0	0	0.48	0.48
var_3_38049588	0	0	0.11	0	0	0.49	0.49
var_3_38050635	0	0.23	0	0.27	0.25	0.51	0.51
var_3_38050789	0.23	0	0	0	0	0.52	0.52
var_3_38058127	0	0.23	0.44	0.54	0.49	0.53	0.53
rs143961610	0	0	0	0	2.72	0.59	0.59
var_3_38057997	0	0.23	0	0	0	0.61	0.61
var_3_38058097	0	0.23	0	0	0.25	0.63	0.62
var_3_38049974	0	0	0.11	0	0	0.63	0.63
var_3_38049550	0	0	0	0	0.25	0.65	0.65
var_3_38049969	0.46	0	0	0	0	0.70	0.70
var_3_38052906	0	0	0	0	0.25	0.73	0.73
rs150106099	0	0	0.11	0.27	0	0.73	0.73
var_3_38050628	0	0	0.23	0	0	0.73	0.73
var_3_38052809	0	0	0	0	0.25	0.73	0.73
var_3_38052791	0	0	0	0	0.25	0.73	0.73
var_3_38053061	0	0	0.56	0	0	0.74	0.74
var_3_38049998	0	0	0.11	0	0	0.77	0.77
var_3_38049604	0	0.23	0	0	0	0.77	0.77
var_3_38061700	0.23	0	0	0	0	0.77	0.77
var_3_38050089	0	0	0.11	0	0	0.78	0.78
var_3_38050002	0	0	0	0.27	0	0.78	0.78
var_3_38051526	0	0	0.11	0	0	0.78	0.78
var_3_38065835	0.23	0	0	0	0	0.83	0.83
rs146357368	0	0.11	0	0	0.25	0.83	0.83
var_3_38051447	0.69	0	0.11	0	0	0.83	0.83
var_3_38052933	0	0.910	0	0	0	0.85	0.85

Table 3.11: **Table 3.9 Continued. Meta-analysis p-value from TransMeta-Rare on each of the variant that is included in *PLCD1*.** The variants are sorted with an increasing order of the p-values. The 'Genetic Similarity' and 'Indep' categories refer to the two kernel choices we provide for Ψ in Section 3.2.2.1.

SNP ID	MAF ($\times 10^{-3}$)					p-value	
	EA	SA	EUR	HIS	AA	Genetic Similarity	Indep
var_3_38049534	0	0.23	0	0	0	0.87	0.87
var_3_38049999	0	0.23	0	0	0	0.87	0.87
var_3_38051675	0	0	0	0.28	0	0.89	0.89
var_3_38051746	0	0	0.11	0	0	0.89	0.89
var_3_38049574	0	0	0	0	0.25	0.89	0.89
var_3_38049589	0.23	0	0	0	0	0.880	0.90
rs150791261	0	0	0	0.28	0.25	0.92	0.90
var_3_38052830	0.23	0	0	0.27	0	0.93	0.93
var_3_38053066	0	0	0	0.27	0.25	0.94	0.94
var_3_38053120	0	0	0	0	0.25	0.96	0.96
var_3_38052711	0.23	0	0	0	0.25	0.96	0.96
rs139755577	0	0.23	0	0	0	0.96	0.96

CHAPTER IV

A Score Test for Jointly Testing the Fixed and Random Effects in Generalized Linear Mixed Models

Abstract

The framework of jointly testing the fixed and random effects has many applications in biomedical studies. One example is to use such tests for ascertaining associations when there exists heterogeneity in meta-analyzing genome-wide association studies (GWAS); another example is the nonparametric test of spline curves. Although extensive research has been conducted on testing random effect terms only, little work has been done for the joint test of fixed and random effects, especially for non-Gaussian outcomes. Here, we propose a score test for the joint test in Generalized Linear Mixed Models (GLMMs). Our method first re-parameterizes the fixed effects terms as a product of a scale parameter and a vector of nuisance parameters. With such re-parameterization, the joint test is equivalent to testing whether the scale parameter is zero. Since the nuisance parameters are hidden under the null hypothesis, we propose using the supremum of score test statistics over the nuisance parameters. We employ a resampling-based copula method to approximate the asymptotic null distribution of the proposed score test statistic. We investigate performances of our method through simulation studies and demonstrate its application to the Michigan Genomics Initiative (MGI) data.

Keywords: Joint testing; Fixed and Random effects; Generalized Linear Mixed Model (GLMM); Score test.

4.1 Introduction

We consider the problem of testing fixed and random effects jointly in a generalized linear mixed model (GLMM). Such a unified testing framework is applicable in many different scenarios, for example assessing the significance of a functional form in a semi-parametric additive mixed model (SAMM); testing an unspecified varying coefficient in a varying-coefficient model; and meta-analyzing heterogeneous effects in genetic association studies. Testing the random effects involves constraints on the variance component parameters, in which classical inference with a standard null distribution no longer holds, because those parameters under the null hypothesis lie on the boundary of the maintained hypothesis (Lin, 1997; Andrews, 2001). As a result, appropriate test statistics need to be developed, with carefully derived corresponding null distributions.

For the linear mixed model (LMM), Self and Liang (1987), Liang and Self (1996) and Stram and Lee (1994) showed that when the data can be divided into independent and identically distributed (i.i.d) subvectors, the asymptotic null distribution for the likelihood ratio test (LRT) of the one-sided variance component is a 50:50 mixture of χ^2 distributions. Crainiceanu and Ruppert (2004) and Crainiceanu et al. (2005) considered the LRT for testing both the fixed and random effects jointly under a more general situation, in which the data cannot be divided into i.i.d subvectors. They showed that if the conventional 50:50 mixture of χ^2 is used as the asymptotic null distribution under such a situation, the LRT can yield very conservative results.

In response, they derived the exact null distribution of the (restricted) LRT through spectral decomposition.

For inference of the joint testing in an LMM with Gaussian outcomes and (multiple) nuisance variance components, an exact distribution of the (R)LRT through spectral decomposition cannot be easily obtained. To address this problem, Greven et al. (2008) and Scheipl et al. (2008) employed a pseudo-likelihood ratio test approach to approximate the null distribution of the (R)LRT. First, they substituted the nuisance parameters by their consistent estimators to obtain the corresponding best linear unbiased predictors (BLUPs). Next, they constructed the pseudo-outcomes by subtracting the BLUPs from the responses, and applied the theories developed by Crainiceanu and Ruppert (2004) to the reduced model to derive the null distribution of the (R)LRT. Through extensive simulation studies, Scheipl et al. (2008) demonstrated that the pseudo-likelihood approach yields controlled type I error rates and equivalent power to bootstrap-based tests, but can be overly conservative when the nuisance variance component is very small and the covariates of the random effects are highly correlated. Wang and Chen (2012) proposed a generalized F-test to conduct the joint test under the setting of an LMM for Gaussian responses with multiple nuisance variance components. Through spectral decomposition of the residual sum of squares, a computationally efficient algorithm was derived to compute the null distribution of the proposed test statistic.

Although the statistical literature thus far offers an array of methods for testing the fixed and random effects jointly for Gaussian responses, corresponding methods for non-Gaussian outcomes remain limited. In fact, to the best of our knowledge, no systematic research exists addressing the joint testing problem with respect to both the Gaussian and non-Gaussian outcomes under the presence of nuisance variance components. To bridge this methodological gap, we propose a score test for the joint testing problem in the GLMM with or without the presence of nuisance random effects. Our score test approach can handle both the Gaussian and non-Gaussian response types, and is asymptotically equivalent to the corresponding LRT.

The rest of this chapter is organized as follows: In Section 4.2, we introduce several motivating examples of the joint testing problem followed by the general modeling framework. In Section 4.3, we propose a score test for the joint testing problem. In Section 4.4, we evaluate the performance of our proposed method and report results from simulation studies under diverse scenarios. We apply our score test to the Michigan Genomics Initiative (MGI) data in Section 4.5 and conclude this chapter with a discussion in Section 4.6. Supplementary texts, tables and figures are presented in Section 4.7.

4.2 Motivating Examples and Statistical Model

In this section, we first introduce several examples to illustrate the motivations for our joint testing problem, and then we present the model for our proposed score test.

4.2.1 Motivating Examples

Example 1 Testing the significance of a functional form in a SAMM.

Consider modeling clustered data in a SAMM:

$$g(\mu_{ij}) = f(t_{ij}) + \mathbf{s}_{ij}^T\boldsymbol{\alpha} + \mathbf{z}_{ij}^T\mathbf{b}_i, \quad i \in \{1, \ldots, m\},\ j \in \{1, \ldots, n_i\}, \tag{4.2.1}$$

where $g(\cdot)$ is a link function; μ_{ij} is the conditional mean of the outcome with given $\mathbf{b}_i$ for the jth observation in the ith cluster; $f(t_{ij})$ is a smooth function relating the scalar covariate t_{ij} to the outcome; $\mathbf{s}_{ij}$s are vectors of fixed effects covariates with coefficients $\boldsymbol{\alpha}$; $\mathbf{z}_{ij}$s are vectors of random effects covariates with cluster specific coefficients $\mathbf{b}_i$, $\mathbf{b}_i \sim N(\mathbf{0}, D_0(\boldsymbol{\theta}))$ with nuisance variance component vector $\boldsymbol{\theta}$. Suppose we are interested in testing the significance of the regression function $f(t)$; that is,

$H_0: f(t) = 0$.

Zhang and Lin (2003) proposed a mixed model representation of the smoothing spline estimator for $f(t)$. Under such a representation, the hth order smoothing spline estimator for $f(t)$ can be written as $\mathbf{f} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\mathbf{a}$, where $\mathbf{f}$ is the vector of $f(t)$ evaluated at all observed t_{ij} values, $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n_i\}$; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_h)^T$ is a vector of coefficients associated with the polynomial bases; and $\mathbf{a} \sim N(\mathbf{0}, \tau\boldsymbol{\Sigma})$ with a non-negative scalar τ and a given scaled covariance matrix $\boldsymbol{\Sigma}$. Specifications of $\mathbf{X}, \boldsymbol{\beta}, \mathbf{U}, \boldsymbol{\Sigma}$ and τ can be found in Supplementary Materials Section 4.7.1. Now, the SAMM (4.2.1) can be written as the following matrix form:

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\alpha} + \mathbf{U}\mathbf{a} + \mathbf{Z}\mathbf{b}, \tag{4.2.2}$$

where $\boldsymbol{\mu} = (\mu_{11}, \ldots, \mu_{1n_1}, \ldots, \mu_{m1}, \ldots, \mu_{mn_m})$; $\mathbf{S}$ is the fixed effects design matrix with the ith row $\mathbf{s}_i = (\mathbf{s}_{i1}, \ldots, \mathbf{s}_{in_i})^T$; $\mathbf{Z} = diag(\mathbf{z}_1, \ldots, \mathbf{z}_m)$ is the random effects design matrix with $\mathbf{z}_i = (z_{i1}, \ldots, z_{in_i})^T$; and $\mathbf{b} = (\mathbf{b}_1^T, \ldots, \mathbf{b}_m^T)^T \sim N(\mathbf{0}, \mathbf{D}(\boldsymbol{\theta}))$ with $\mathbf{D}(\boldsymbol{\theta}) = diag(D_0(\boldsymbol{\theta}), \ldots, D_0(\boldsymbol{\theta}))$.

Under the working GLMM (4.2.2), the significance of $f(t)$ can be assessed through:

$$\begin{aligned} H_0: \boldsymbol{\beta} &= (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_h)^T = \mathbf{0},\ \mathbf{a} = \mathbf{0}, \\ \text{or equivalently,}\quad H_0: \boldsymbol{\beta} &= (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_h)^T = \mathbf{0},\ \tau = 0. \end{aligned}$$

Example 2 Testing an unspecified varying coefficient in a varying-coefficient model.

The varying coefficient model is a natural extension to the classical parametric regression model, allowing the coefficient to vary smoothly over the covariates. Consider the following model with a varying coefficient:

$$g(\mu_{ij}) = \beta(t_{ij})g_{ij} + \mathbf{s}_{ij}^T\boldsymbol{\alpha} + \mathbf{z}_{ij}^T\mathbf{b}_i, \quad i \in \{1, \ldots, m\},\ j \in \{1, \ldots, n_i\}, \tag{4.2.3}$$

where $\beta(t_{ij})$ is a smooth functional coefficient describing the relationship between the covariate g_{ij} and the outcome, and the remaining of the model specifications is as defined in Equation (4.2.1). As in the parametric regression model, it is often of interest to assess the effect of g_{ij} on y_{ij}, that is, to evaluate whether the coefficient

$\beta(\cdot) = 0$. Applying the working GLMM representation by Zhang and Lin (2003) again to the functional coefficient $\beta(\cdot)$, we can reformulate the varying coefficient model (4.2.3) to a mixed model as represented in Equation (4.2.2), in which we effectively transform the question of evaluating the effect of $\beta(\cdot)$ into the joint testing of fixed effects $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_h)^T = \mathbf{0}$ and random effects $\mathbf{a} = \mathbf{0}$ (or equivalently $\tau = 0$).

Example 3 Meta-analysis of heterogeneous effects in association studies.

Meta-analysis is a practical approach to aggregate studies that have already been conducted in order to boost power to identify association signals. Let $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_n)^T$ be the effect-size estimates from n independent studies. Assume that $\widehat{\boldsymbol{\beta}}|\boldsymbol{\beta} \sim MVN(\boldsymbol{\beta}, \Sigma)$, with the true effect size vector denoted as $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^T$ and covariance matrix Σ. To account for the situation where effect sizes among the studies are significantly different from each other, one modeling strategy is to allow the true effects across studies to vary around the overall mean, i.e. $\beta_i = \mu + \eta_i,\ \eta_i \sim N(0, \tau)$, where the fixed scalar μ represents the overall mean, and the random effect η_i denotes the study-specific deviation. Consequently, testing for the association signal is equivalent to jointly testing the fixed effect μ and random effects η_is; that is, the null hypothesis now becomes $H_0 : \mu = 0,\ \tau = 0$ (Han and Eskin, 2011). Han and Eskin (2011) proposed using the likelihood ratio framework to conduct such hypothesis testing and provided pre-tabulated p-values to accurately approximate the statistical significance.

Example 4 Testing rare-variants associations in sequencing studies.

The recent advance in sequencing technologies have prompted significant research on developing statistical methods for testing associations between rare variants and complex traits. Lee et al. (2012) proposed a unified score test approach, named SKAT-O, to optimally combine the burden test and the sequence kernel association test (SKAT). The regression model for SKAT-O is $g(\mu_i) = \mathbf{X}_i^T \boldsymbol{\alpha} + \mathbf{G}_i^T \boldsymbol{\beta}$, where μ_i denotes the mean of phenotype y_i of the ith subject, $\mathbf{X}_i$ is a vector of adjusting covariates with coefficients $\boldsymbol{\alpha}$, and $\mathbf{G}_i$ is a genotype matrix with regression coefficients $\boldsymbol{\beta}$ for the genetic variants. SKAT-O was originally constructed as a weighted average of SKAT and the burden test, but an alternative view for deriving SKAT-O is to assume that the regression coefficient β_j of each genetic variant j independently follows $\beta_j = \omega_j \beta_c, \beta_c \sim N(\mu, \tau)$. Under such an alternative modeling strategy, the SKAT-O test statistic can be derived for testing the null hypothesis $H_0 : \mu = 0,\ \tau = 0$.

4.2.2 Statistical Model

The above motivating examples can all be viewed as applications of jointly testing fixed and random effects in a GLMM with or without the presence of nuisance random effects. In summary, the goal is to test $H_0 : \boldsymbol{\beta} = \mathbf{0},\ \tau = 0$ in the GLMM

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\alpha} + \mathbf{U}\mathbf{a} + \mathbf{Z}\mathbf{b}, \quad \mathbf{a} \sim N(\mathbf{0}, \tau\boldsymbol{\Sigma}),\ \mathbf{b} \sim N(\mathbf{0}, \mathbf{D}(\boldsymbol{\theta})); \tag{4.2.4}$$

where $g(\cdot)$ is a known link function; $\boldsymbol{\mu} = (\mu_{11}, \ldots, \mu_{1n_1}, \ldots, \mu_{m1}, \ldots, \mu_{mn_m})$ is the vectorized conditional mean μ_{ij} of the outcome variable y_{ij} for the jth observation in the ith cluster ($i \in \{1, \ldots, m\}, j \in \{1, \ldots, n_i\}$); $\mathbf{S}$ and $\mathbf{Z}$ are the design matrices for the fixed and random effects, respectively, to be adjusted for; $\mathbf{X}$ and $\mathbf{U}$ are the design matrices of interest for the fixed and random effects, respectively, to be tested for; and $\tau, \boldsymbol{\Sigma}$, and $\mathbf{D}(\boldsymbol{\theta})$ have the same specifications as in Section 4.2 Example 1. We assume that conditional on the unobserved random effects vectors $\mathbf{a}$ and $\mathbf{b}$, the outcome y_{ij}s are independent with means $E(y_{ij}|\mathbf{a}, \mathbf{b}) = \mu_{ij}$ and variances $var(y_{ij}|\mathbf{a}, \mathbf{b}) = V(\mu_{ij}) = \phi\zeta_{ij}^{-1}\nu(\mu_{ij})$, where ϕ is a scale parameter, ζ_{ij} is a prior weight, and $\nu(\cdot)$ is a variance function.

Although we construct our framework with respect to the longitudinal/clustered data structure, the setting is also adaptive to cross sectional data under the null hypothesis, in which case $j = 1$ for every $i \in \{1, \ldots, m\}$ and $\mathbf{Zb}$ will be eliminated from model (4.2.4).

4.3 Methods

4.3.1 The Score Test for Gaussian Responses

We first consider the scenario where the responses y_{ij}s follow a Gaussian distribution with the identity link function. In this case, the model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\alpha} + \mathbf{U}\mathbf{a} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \tag{4.3.1}$$

with $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$ for $n = \sum_{i=1}^{m} n_i$. Here we propose a score test for our goal $H_0 : \boldsymbol{\beta} = \mathbf{0},\ \mathbf{a} = \mathbf{0}$. First we re-parameterize the fixed effects parameters of interest with an unknown nuisance parameter vector $\boldsymbol{\gamma}$ scaled by the variance component τ, i.e. let $\boldsymbol{\beta} = \tau\boldsymbol{\gamma}$. Under this re-parameterization, jointly testing $\boldsymbol{\beta} = \mathbf{0}$ and $\mathbf{a} = \mathbf{0}$ is equivalent to testing $\tau = 0$ alone, and the mixed model becomes

$$\mathbf{y} = \tau\mathbf{X}\boldsymbol{\gamma} + \mathbf{S}\boldsymbol{\alpha} + \mathbf{U}\mathbf{a} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}. \tag{4.3.2}$$

Let $\boldsymbol{\nu} = (\boldsymbol{\alpha}^T, \boldsymbol{\theta}^T, \sigma^2)^T$ and $\mathbf{V} = \mathbf{Z}\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T + \sigma^2\mathbf{I}_n$. Denote $l(\tau, \boldsymbol{\nu},\boldsymbol{\gamma}; \mathbf{y})$ the log-likelihood function of model parameters $(\tau, \boldsymbol{\nu},\boldsymbol{\gamma})$ under the LMM in Equation (4.3.2). It can be easily shown that given $\boldsymbol{\gamma}$, the score statistic for testing $H_0 : \tau = 0$ is

$$U_\tau(\hat{\boldsymbol{\nu}}; \boldsymbol{\gamma}) = \frac{\partial l(\tau,\boldsymbol{\nu},\boldsymbol{\gamma};\mathbf{y})}{\partial \tau}\bigg|_{\tau=0,\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} = -\frac{1}{2}tr(\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T) + \frac{1}{2}\mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\hat{\mathbf{V}}^{-1}\mathbf{P}\mathbf{y} + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\boldsymbol{\gamma},$$

where $\mathbf{P} = \mathbf{I}_n - \mathbf{S}(\mathbf{S}^T\hat{\mathbf{V}}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\hat{\mathbf{V}}^{-1}$, and $\hat{\mathbf{V}} = \mathbf{Z}\mathbf{D}(\hat{\boldsymbol{\theta}})\mathbf{Z}^T + \hat{\sigma}^2\mathbf{I}_n$. $U_\tau(\hat{\boldsymbol{\nu}}; \boldsymbol{\gamma})$ has the nuisance parameter $\boldsymbol{\gamma}$ that is hidden under the null hypothesis. Several researchers have proposed methods for removing the unidentifiable nuisance parameters under the null. For example, Davies (1987) suggested using the supremum of a test statistic over all possible values of the nuisance parameter space. Similarly, Zhu et al. (2006) used the supremum of the score statistic to assess the homogeneity in a GLMM. Inspired by these works, we propose using the following test statistic T, which is the supremum of the standardized score $U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma})$:

$$\begin{aligned} T &= \sup_{\boldsymbol{\gamma}} \frac{U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma}) - \mathrm{E}(U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma}))}{\sqrt{\mathrm{VAR}(U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma}))}}, \qquad (4.3.3) \\ \text{where } \mathrm{E}(U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma})) &= \frac{1}{2}tr(\widetilde{\mathbf{P}}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\widetilde{\mathbf{P}}\hat{\mathbf{V}}) - \frac{1}{2}tr(\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T) \text{ for a given value of } \boldsymbol{\gamma}, \\ \text{and } \mathrm{VAR}(U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma})) &= \frac{1}{2}tr[(\widetilde{\mathbf{P}}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\widetilde{\mathbf{P}}\hat{\mathbf{V}})^2] + \boldsymbol{\gamma}^T\mathbf{X}^T\widetilde{\mathbf{P}}\hat{\mathbf{V}}\widetilde{\mathbf{P}}\mathbf{X}\boldsymbol{\gamma}, \text{with } \widetilde{\mathbf{P}} = \hat{\mathbf{V}}^{-1}\mathbf{P}. \end{aligned}$$

In Supplementary Materials Section 4.7.4, we show that the proposed test statistic in Equation (4.3.3) is asymptotically equivalent to the LRT.

Since it is difficult to explore the entire space of $\boldsymbol{\gamma}$ to obtain the supremum in Equation (4.3.3), an alternative way to solve this is to view the test statistic T as an optimization problem:

Maximize $U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma}) - \mathrm{E}(U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma}))$ subject to $\mathrm{VAR}(U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma})) = c,$ where c is a constant.

From this perspective, Lagrange multipliers can be applied to solve for $\boldsymbol{\gamma}$, which results in

$$\widehat{\boldsymbol{\gamma}} = \eta(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{y}, \qquad (4.3.4)$$

where η is a non-negative scalar representing the strength of $\boldsymbol{\gamma}$ (see Supplementary Materials Section 4.7.2 for derivation details). Notice that all quantities in Equation (4.3.4) are known except for η, thus, determining the vector $\boldsymbol{\gamma}$ is essentially reduced to determining the scalar η. Plug in (4.3.4) to the score statistic $U_\tau(\hat{\nu};\boldsymbol{\gamma})$, we now obtain

$$\begin{aligned} U_\tau(\hat{\boldsymbol{\nu}};\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}) &= \eta Q_1 + Q_2 - \frac{1}{2}tr(\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T) \qquad (4.3.5) \\ \text{where } Q_1 &= \mathbf{y}^T\widetilde{\mathbf{P}}^T\mathbf{X}(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{y}, \quad Q_2 = \frac{1}{2}\mathbf{y}^T\widetilde{\mathbf{P}}^T\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\widetilde{\mathbf{P}}\mathbf{y}. \end{aligned}$$

Plugging in (4.3.5) to (4.3.3), we obtain $T = \sup_{\eta \geq 0} \frac{\eta Q_1 + Q_2 - \mathrm{E}\{\eta Q_1 + Q_2\}}{\sqrt{\mathrm{VAR}(\{\eta Q_1 + Q_2\})}}$, where the supremum is reduced from searching the vector space of $\boldsymbol{\gamma}$ to the scalar space of η.

Re-parameterize η as $\eta = \rho(1-\rho)^{-1}, 0 \leq \rho \leq 1$, we then obtain:

$$\begin{aligned} T &= \sup_{0\leq\rho\leq 1} \frac{\rho Q_1 + (1-\rho)Q_2 - E_\rho}{\sqrt{V_\rho}}, \text{ with } E_\rho = \rho \cdot rank(\mathbf{X}) + \frac{1-\rho}{2} tr(\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\widetilde{\mathbf{P}}), \quad (4.3.6) \\ \text{and } V_\rho &= 2\rho^2 \cdot rank(\mathbf{X}) + \frac{(1-\rho)^2}{2} tr[(\mathbf{U}\Sigma\mathbf{U}^T\widetilde{\mathbf{P}})^2] + 2\rho(1-\rho)tr[\mathbf{X}(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{U}\Sigma\mathbf{U}^T\widetilde{\mathbf{P}}]. \end{aligned}$$

Equation (4.3.6) is the final form of our proposed score test for assessing $H_0 : \boldsymbol{\beta} = \mathbf{0},\ \mathbf{a} = \mathbf{0}$. It shows that T is a weighted sum of two quadratic forms, in which Q_1 represents the fixed effects and Q_2 represents the random effects. In fact, Q_1 is a score test statistic for $\boldsymbol{\beta}$ when $\tau = 0$, and Q_2 is a score test statistic for τ when $\boldsymbol{\beta} = \mathbf{0}$. The value of T can now be obtained by a simple grid search across a range of ρs: set a grid of B points as $0 = \rho_1 \leq \cdots \leq \rho_B = 1$, then an approximated value of the test statistic can be computed as

$$\hat{T} = \max_{\rho=\rho_1,\dots\rho_B} \frac{S_\rho - E_\rho}{\sqrt{V_\rho}}, \quad \text{where } S_\rho = \rho Q_1 + (1-\rho)Q_2. \qquad (4.3.7)$$

We observe that a dense grid of ρ does not necessarily improve power over a coarse grid. Therefore, we suggest using $\rho = (0, 0.25, 0.5, 0.75, 1)$ for simulations and real data analysis.

4.3.2 The Score Test for Non-Gaussian Responses

In this section, we extend the score test proposed in Section 4.3.1 for non-Gaussian responses. For general response types, we assume that conditional on $(\mathbf{x}_{ij}, \mathbf{s}_{ij}, \mathbf{u}_{ij}, \mathbf{z}_{ij})$ and $(\mathbf{a}_i, \mathbf{b}_i)$, y_{ij} follows a distribution in the exponential family, and we consider any canonical link function. As in the Gaussian responses case, we apply the re-parameterization $\boldsymbol{\beta} = \tau\boldsymbol{\gamma}$. Let $\boldsymbol{\Delta}$ and $\mathbf{W}$ denote the $n \times n$ diagonal matrix which has elements

$$\delta_{ij} = g'(\mu_{ij}), \quad w_{ij} = \left[V(\mu_i)\left\{g'(\mu_{ij})\right\}^2\right]^{-1},$$

respectively, where $\mu_{ij} = E(y_{ij}|\mathbf{a}_i, \mathbf{b}_i)$ and $g(\mu_{ij}) = \tau\mathbf{x}_{ij}^T\boldsymbol{\gamma} + \mathbf{s}_{ij}^T\boldsymbol{\alpha} + \mathbf{u}_{ij}^T\mathbf{a}_i + \mathbf{z}_{ij}^T\mathbf{b}_i$. Following (Zhang and Lin, 2003), it can be shown that given the nuisance parameter vector $\boldsymbol{\gamma}$, the score statistic of τ evaluated at $\tau = 0$ is:

$$U_\tau \approx \frac{1}{2}E\{(\mathbf{y}-\boldsymbol{\mu}^b)^T\boldsymbol{\Delta}\mathbf{W}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\mathbf{W}\boldsymbol{\Delta}(\mathbf{y}-\boldsymbol{\mu}^b) - tr(\mathbf{W}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T) - (\mathbf{y}-\boldsymbol{\mu}^b)^T\boldsymbol{\Delta}\mathbf{W}\mathbf{X}\boldsymbol{\gamma}\} \tag{4.3.8}$$

where $\nu = (\boldsymbol{\alpha}^T, \boldsymbol{\theta}^T)^T$ and $\boldsymbol{\mu}^b$ satisfies the null GLMM

$$g(\boldsymbol{\mu}^b) = \mathbf{S}\boldsymbol{\alpha} + \mathbf{Z}\mathbf{b}. \tag{4.3.9}$$

Denote the working vector as $\boldsymbol{Y} = \mathbf{S}\boldsymbol{\alpha} + \boldsymbol{\Delta}(\boldsymbol{y} - \boldsymbol{\mu})$ under the null GLMM (4.3.9). One can show that using the Laplace method, the score statistic in (4.3.8) can be approximated as

$$U_\tau \approx -\frac{1}{2}tr(\widetilde{\mathbf{P}}\mathbf{U}\Sigma\mathbf{U}^T) + \frac{1}{2}(\mathbf{Y}-\mathbf{S}\widehat{\boldsymbol{\alpha}})^T\widehat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\widehat{\mathbf{V}}^{-1}(\mathbf{Y}-\mathbf{S}\widehat{\boldsymbol{\alpha}}) + (\mathbf{Y}-\mathbf{S}\widehat{\boldsymbol{\alpha}})^T\widehat{\mathbf{V}}^{-1}\mathbf{X}\boldsymbol{\gamma} \tag{4.3.10}$$

where $\widehat{\boldsymbol{\alpha}}$ is the BLUP-type estimate of $\boldsymbol{\alpha}$; $\widehat{\boldsymbol{\theta}}$ is the REML estimate of $\boldsymbol{\theta}$; $\widehat{\mathbf{V}} = \mathbf{Z}\mathbf{D}(\widehat{\boldsymbol{\theta}})\mathbf{Z}^T + \widehat{\mathbf{W}}^{-1}$; and $\widetilde{\mathbf{P}} = \widehat{\mathbf{V}}^{-1}\mathbf{P} = \widehat{\mathbf{V}}^{-1} - \widehat{\mathbf{V}}^{-1}\mathbf{S}(\mathbf{S}^T\widehat{\mathbf{V}}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\widehat{\mathbf{V}}^{-1}$, as defined in the Gaussian responses case.

Notice that equation (4.3.10) corresponds to the score statistic of τ evaluated at $\tau = 0$ under the working linear mixed model $\mathbf{Y} = \tau\mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\alpha} + \mathbf{U}\mathbf{a} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}$, with $\mathbf{a} \sim N(\mathbf{0}, \tau\boldsymbol{\Sigma}), \mathbf{b} \sim \mathbf{N}(\mathbf{0}, \mathbf{D}(\boldsymbol{\theta}))$, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{W}^{-1})$. Consequently, by replacing $\mathbf{y}$ in Section 4.3.1 with the working vector $\mathbf{Y}$, we can apply the results in Section 4.3.1 to carry out the score test for non-Gaussian responses.

4.3.3 Assess Statistical Significance of the Test

In this section, we outline a resampling-based copula method (Xianglin Li, Xianglin Li) to approximate the asymptotic null distribution of the proposed score test. Denote t as the observed value of the score test $\hat{T}$, and the statistical significance for t can be obtained from

$$\begin{aligned} P(\hat{T} \geq t) &= P(\max_{\rho=\rho_1,\ldots\rho_B} \frac{S_\rho - E_\rho}{\sqrt{V_\rho}} \geq t) \\ &= 1 - P(F_{\rho_1}(S_{\rho_1}) \leq F_{\rho_1}(\sqrt{V_{\rho_1}} \cdot t + E_{\rho_1}), \ldots, F_{\rho_B}(S_{\rho_B}) \leq F_{\rho_B}(\sqrt{V_{\rho_B}} \cdot t + E_{\rho_B})), \end{aligned}$$

where $F_{\rho_i}(\cdot)$ is the cumulative distribution function of S_{ρ_i}. For any given ρ, it can be shown that the marginal distribution of S_ρ asymptotically follows a mixture of chi-square distribution $\sum \lambda_l \chi^2_{l,1}$, where λ_ls are the non-zero eigenvalues of

$$\rho\hat{\mathbf{P}}\mathbf{V}^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}} + \frac{1}{2}(1-\rho)\hat{\mathbf{P}}^T\mathbf{V}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}},$$

with $\hat{\mathbf{P}} = \mathbf{I} - \mathbf{V}^{-\frac{1}{2}}\mathbf{S}(\mathbf{S}^T\mathbf{V}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{V}^{-\frac{1}{2}}$; and $\chi^2_{l,1}$s are chi-square distributions with one degree of freedom (see details in Supplementary Materials Section 4.7.3). As a result, the value of $F_\rho(\sqrt{V_\rho}\cdot t + E_\rho)$ can be easily obtained for any given ρ value by using Davies' method (Davies, 1980) to invert a characteristic function. We also notice that under the null hypothesis, $\mathbf{V}^{-\frac{1}{2}}\mathbf{y}$ approximately follows a multivariate normal distribution $N(\mathbf{0}, \mathbf{I}_n)$, where $\mathbf{I}_n$ is an identity matrix with dimension $n \times n$. Therefore, we propose the following resampling-based copula algorithm to approximate the null joint distribution of $(F_{\rho_1}(S_{\rho_1}), F_{\rho_2}(S_{\rho_2}), \ldots, F_{\rho_B}(S_{\rho_B}))$:

Step 1: Generate n_0 samples, say $\mathbf{u}$, from the standard normal distribution $N(\mathbf{0}, \mathbf{I}_n)$. We use $n_0 = 500$ in our simulation studies and data application.

Step 2: For each $\rho \in \{\rho_1, \ldots, \rho_B\}$, calculate the null scores as

$$S^0_\rho = \mathbf{u}^T \cdot [\rho\hat{\mathbf{P}}\mathbf{V}^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}} + \frac{1}{2}(1-\rho)\hat{\mathbf{P}}^T\mathbf{V}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}}] \cdot \mathbf{u}.$$

Step 3: Calculate the correlation matrix $\Sigma_{B\times B}$ of the generated null score S^0_ρs among any pair of ρ values, where B is the length of the $(\rho_1, \ldots, \rho_B)$ grid.

Step 4: With the estimated null correlation structure $\Sigma_{B\times B}$, we next use the Gaussian copula to approximate the null joint distribution of $(F_{\rho_1}(S_{\rho_1}), F_{\rho_2}(S_{\rho_2}), \ldots, F_{\rho_B}(S_{\rho_B}))$, which yields

$$P(\hat{T} \geq t) = 1 - \Phi_{\Sigma_{B\times B}}(\Phi^{-1}(1 - F_{\rho_1}(\sqrt{V_{\rho_1}}\cdot t + E_{\rho_1})), \ldots, \Phi^{-1}(1 - F_{\rho_B}(\sqrt{V_{\rho_B}}\cdot t + E_{\rho_B}))),$$

where $1 - F_\rho(\sqrt{V_\rho}\cdot t + E_\rho)$ is the p-value of S_ρ obtained from Davies' method, for any given ρ value; Φ^{-1} is the inverse cumulative distribution function of a standard normal (consequently, $\Phi^{-1}(1 - F_\rho(\sqrt{V_\rho}\cdot t + E_\rho))$ is the normal z value from the p-value of S_ρ); and $\Phi_{\Sigma_{B\times B}}$ is the joint cumulative distribution of a multivariate normal with zero mean vector and correlation matrix equal to $\Sigma_{B\times B}$.

It should be noted that when calculating the correlation matrix $\Sigma_{B\times B}$, Pearson's correlation coefficient can yield unreliable estimates due to its strong dependent on the normality and homoscedasticity assumptions (Hauke and Kossowski, 2011). Instead, we use Spearman's correlation, a non-parametric version of the Pearson's correlation based on ranks of the random variables.

4.4 Simulation Studies

In this section, we conducted a series of simulation studies to evaluate the size and power of our method with different types of responses under diverse scenarios. We compared the power of our score test with other existing approaches for Gaussian outcomes only, since to our knowledge, no solutions other than our method are available for non-Gaussian outcomes.

4.4.1 Scenario 1 - Non-parametric Regression Tests in Cross Sectional Studies

One key assumption in linear models is that the conditional mean of the response variable depends on the covariates parametrically. For data with complex covariate effects, such strong parametric assumption may not be appropriate. Thus, it is of substantial interest to test whether a predictor is related to the outcome in a flexible nonparametric fashion. We first performed such tests in cross sectional studies with no nuisance random effects under the null. Consider the regression equation $g(\mu_i) = f_d(t_i)$, where the link function $g(\cdot)$ is the identity function for Gaussian responses and logit function for binary/binomial responses, $f_d(\cdot)$ is an arbitrary smooth function. Our goal was to test whether $f_d(\cdot)$ is a constant value, i.e. whether $H_0 : f_d(t)$ is a constant function in t. We considered two different forms of $f_d(\cdot)$:

$$\begin{aligned} (a) \quad f_d(t) &= 0.5 + 0.25 \cdot d \cdot t \cdot exp(2 - 2 \cdot t),\ d \in \{0, 0.75, 1, 1.25\} \quad (4.4.1) \\ \text{and } (b) \quad f_d(t) &= 1 + 0.05 \cdot d \cdot t + 0.1 \cdot \cos(d \cdot \pi \cdot t),\ d \in \{0, 2, 2.5, 3\}. \end{aligned}$$

The functions $f_d(t)$ against t for different values of d were plotted in Figure 4.1, in which the parameter d controls the level of the effect size: $d = 0$ corresponds to the null hypothesis that $f_d(t)$ is a constant function in t; when $d \neq 0$, the larger d becomes, the further away $f_d(t)$ deviates from being constant in t.

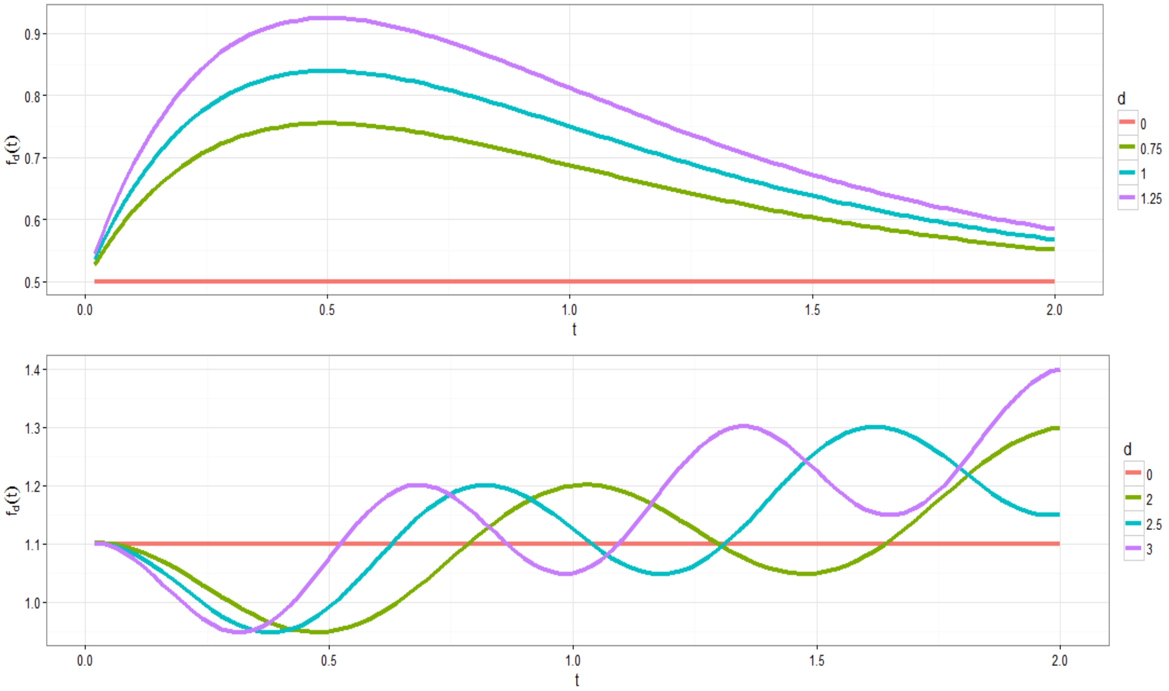


Figure 4.1: **Functions $f_d(t)$ used in the simulation studies for the nonparametric regression tests.** The upper and lower panel correspond to the first and second smooth functions defined in Equation (4.4.1) respectively.

Each simulated dataset was composed of sample size $n = 500$, with one hundred equally spaced points in the quintuple for t_i in [0,2]. For Gaussian responses, the residual term ϵ_i was generated from a $N(0, 0.65^2)$. We adapted Zhang and Lin (2003)'s mixed-model formulation of the natural smoothing spline estimator for the nonparametric function $f_d(\cdot)$. To ensure the desired flexibility, we considered different orders for the smoothing spline, namely, the linear, quadratic and cubic smoothing splines. As discussed in Example 1 of Section 4.2.1, this smoothing spline estimator of $f_d(\cdot)$ reformulates the regression model into a GLMM. Consequently, testing whether $f_d(\cdot)$ is a constant value is equivalent to jointly testing the fixed and random effects in the mixed model.

We first evaluated the empirical type I error rate of our proposed score test at the 5% nominal level. Results of the obtained empirical type I error rates based on 5000 simulations were presented in Table 4.1. Across various smoothing spline representations for $f_d(\cdot)$, our proposed score test behaved satisfactorily for both the Gaussian and non-Gaussian responses. The likelihood ratio test with the exact finite sample distribution developed by Crainiceanu and Ruppert (2004) also had controlled nominal levels for Gaussian outcomes. As expected, approximating the asymptotic distribution of the LRT using the 50:50 mixture χ^2 distribution resulted in conservative type I error rates.

Table 4.1: **Empirical sizes of the constancy test for Gaussian and non-Gaussian responses based on 5000 simulation runs for non-parametric regression test in cross sectional studies.** LRT refers to the likelihood ratio test with exact finite sample distribution developed by Crainiceanu and Ruppert (2004); Mixture χ^2 refers to the likelihood ratio test with an approximated 50:50 mixture χ^2 asymptotic distribution.

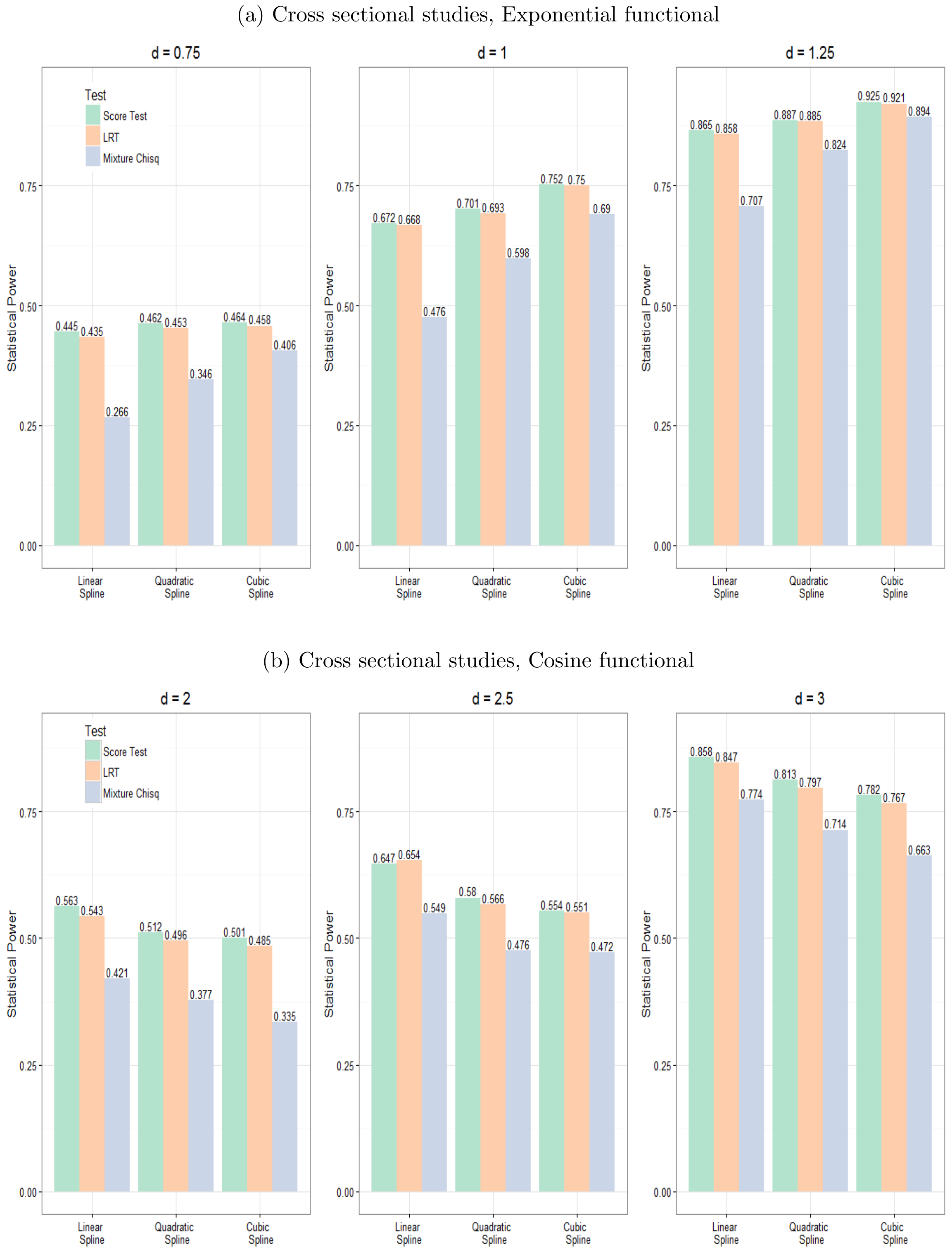
Data Type	Test	Size		
		Linear	Quadratic	Cubic
	ScoreTest	0.052	0.051	0.050
Gaussian	LRT	0.049	0.052	0.051
	Mixture χ^2	0.032	0.034	0.036
Binary	ScoreTest	0.051	0.049	0.051
Binomial (N=6)	ScoreTest	0.052	0.048	0.051

To compare the power of the various tests, we then simulated data under the alternative hypothesis with positive values for d. The barplots in Figure 4.2 summarized the power comparisons for Gaussian responses. From the barplot, it can be seen that for both smooth functions in (4.4.1), the power of our score test was very similar or slightly better than that of the LRT by Crainiceanu and Ruppert (2004). As expected, the power of the 50:50 mixture χ^2 approach had the lowest power. The power of all the tests increased as the effect size measure, d, increases. When fixing d, we observed different trends of the power across different orders of the spline estimator of $f_d(\cdot)$. With a fixed d value for the smooth function (a) in Equation (4.4.1), we observed that the power for all tests tends to increase as the order of splines increases. In contrast, for the functional form (b) in Equation (4.4.1), we observed that the power tends to decrease as the order of splines increases. The diverging results may look contradictory at first glance, but are justifiable. For functional (a) (Figure 4.1), the addition of quadratic and cubic spline terms to the linear spline helped better approximate the shape of the functional; as a result, inclusion of higher orders of spline bases contributed to power gain. In contrast, for functional (b), although the linear spline basis was capable of capturing the linear trend in the functional, the quadratic and cubic spline bases did not contribute to approximating the periodic cosine component in the functional; in this case, incorporation of the higher order bases resulted in lower power than using the linear spline basis alone.

In Figure 4.3, we display the results of the power simulation for our score test with respect to binary and binomial responses.

4.4.2 Scenario 2 - Non-parametric Regression Tests in Longitudinal Studies

The next scenario we considered is testing the significance of a regression function in longitudinal studies with a nuisance random effect under the null hypothesis. The regression equation was then modeled as $g(\mu_{ij}) = f_d(t_{ij}) + b_i$, where the link function


(a) Cross sectional studies, Exponential functional

(b) Cross sectional studies, Cosine functional

Figure 4.2: **Power of the constancy test for Gaussian responses based on 5000 simulations in cross sectional studies.** The upper and lower panel correspond to the functional form (a), (b) defined in Equation (4.4.1), respectively. LRT refers to the likelihood ratio test with the exact finite sample distribution developed by Crainiceanu and Ruppert (2004); Mixture χ^2 refers to the likelihood ratio test with an approximated 50:50 mixture χ^2 asymptotic distribution.

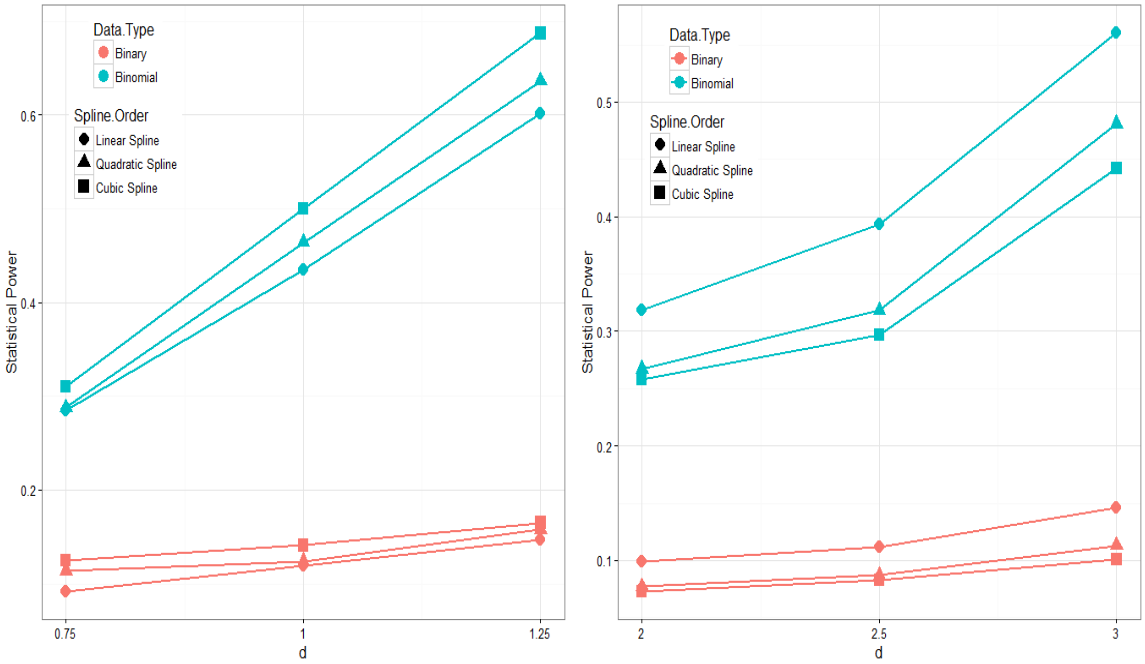


Figure 4.3: **Power of the constancy test for binary and binomial responses based on 5000 simulations in cross sectional studies.** The left and right panel correspond to the functional form (a), (b) defined in Equation (4.4.1) respectively.

$g(\cdot)$ and regression function $f_d(\cdot)$ were the same as defined in Section 4.4.1. Each simulated dataset still contained 500 observations, but with 100 clusters of size $n_i = 5$. We generated equally spaced values for t in [0,2] through $t_{ij} = [trun\{(i+4)/5\}/50] + 0.4(j-1)$ ($i = 1, \ldots, 100$ and $j = 1, \ldots, 5$), where $trun(\cdot)$ denoted a truncation operator with $trunc(x) = \lfloor x \rfloor$ for $x \in R_+$, and the random intercept b_i followed $N(0, 0.5^2)$.

To evaluate the empirical type I error rate and power of our proposed score test, we simulated 5000 datasets. Results of the empirical type I error rate at the 5% nominal level were summarized in Table 4.2. Similar to the cross sectional setting in Scenario 1, our proposed score test yielded well controlled type I error rates for both Gaussian and non-Gaussian responses under this longitudinal setting. The pseudo likelihood ratio test approach developed by Greven et al. (2008) and generalized F-test proposed by Wang and Chen (2012) also behaved satisfactorily for Gaussian outcomes. Unsurprisingly, the approximated 50:50 mixture χ^2 distribution for the LRT still yielded a conservative type I error rate.

The barplots in Figure 4.4 summarized the power comparisons for Gaussian responses. Similar to the cross sectional setting in Scenario 1, the power of our score test was very similar or slightly better than the pseudo-LRT by Greven et al. (2008) and the generalized F-test by Wang and Chen (2012), with the power of the 50:50 mixture χ^2 approach yielding the least powerful results among all. As the effect size measure d increased, the power of all tests increased as expected. For a fixed d value, regardless of the test used, the power improved as the spline representation more closely approximates the underlying functional form. Results of the power simulation for our score test with respect to binary and binomial responses were presented in Figure 4.5. The power of the score test increased quickly as we increase the binomial denominator from 1 (i.e. binary responses) to 6.

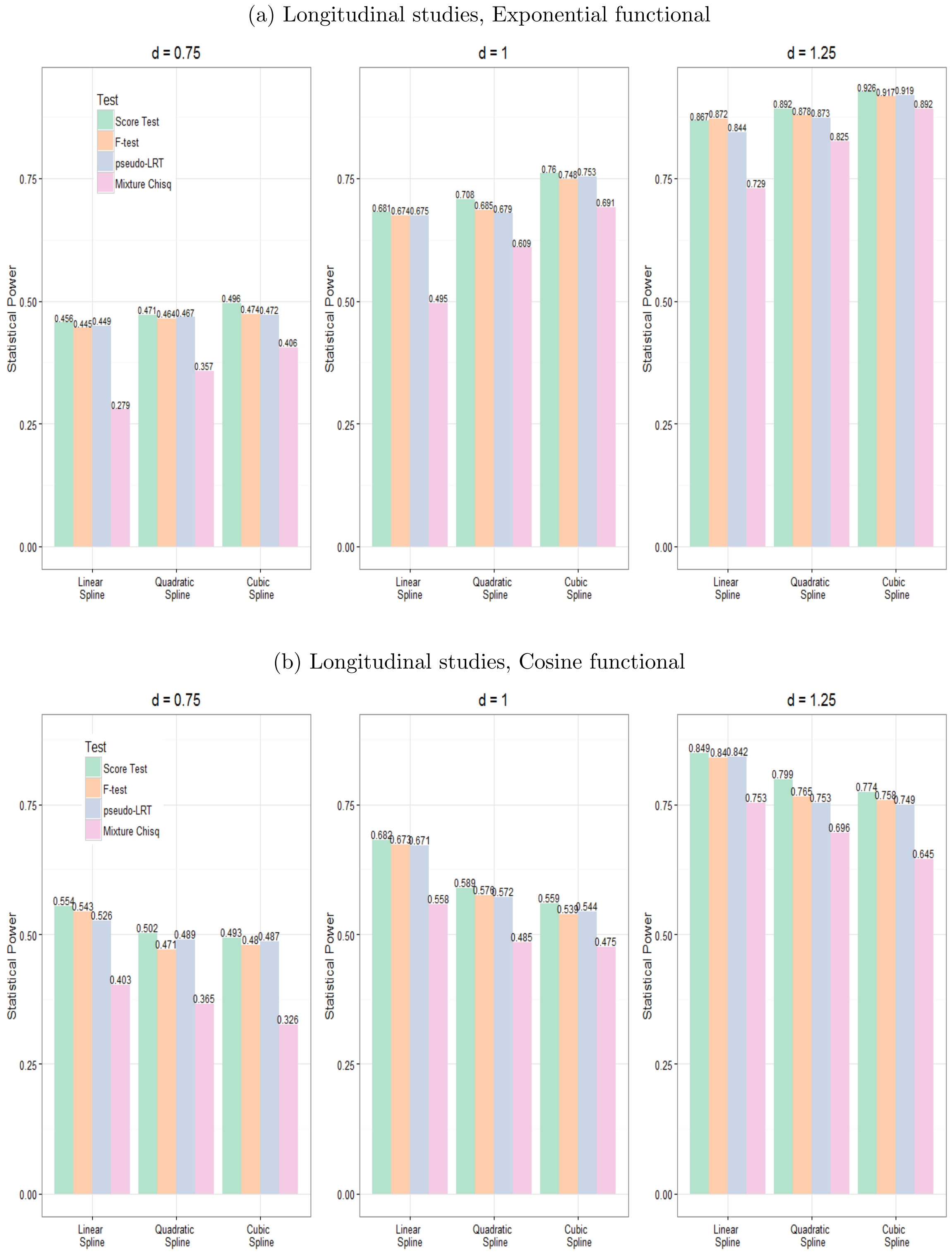


Figure 4.4: **Power of the constancy test for Gaussian responses based on 5000 simulations in longitudinal studies.** The upper and lower panel correspond to the functional form (a), (b) defined in Equation (4.4.1) respectively. F-test refers to the generalized F-test proposed by Wang and Chen (2012); Pseudo-LRT refers to the pseudo likelihood ratio test approach developed by Greven et al. (2008); Mixture χ^2refers to the likelihood ratio test with an approximated 50:50 mixture χ^2 asymptotic distribution.

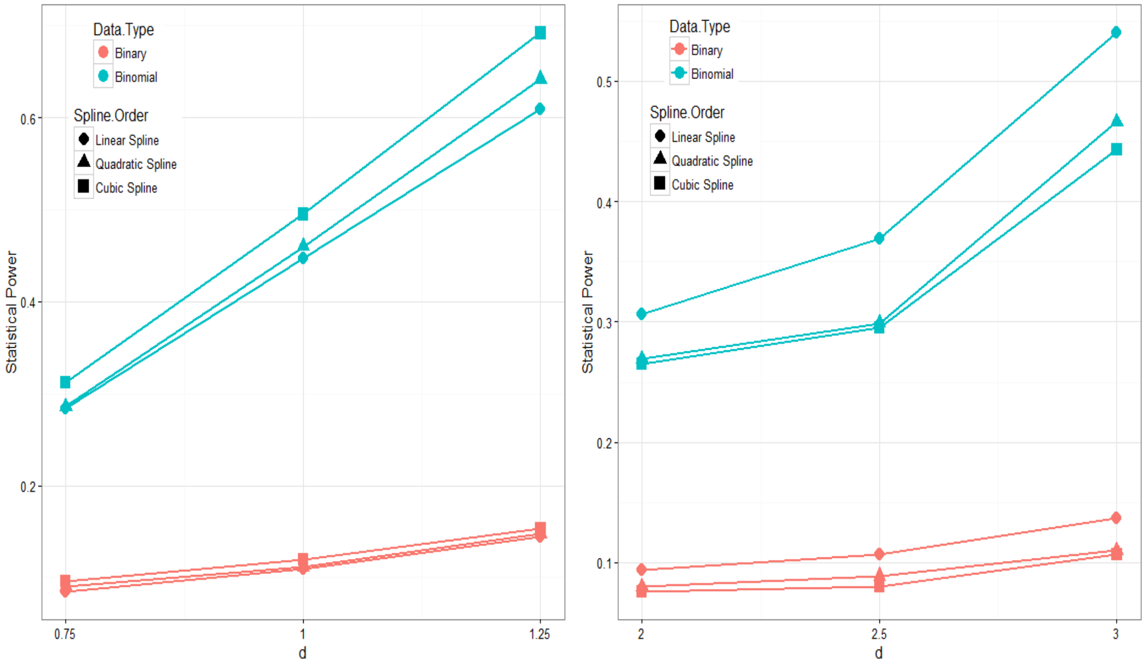


Figure 4.5: **Power of the constancy test for Binary and Binomial responses based on 5000 simulations in longitudinal studies.** The left and right panel correspond to the functional form (a), (b) defined in Equation (4.4.1) respectively.

Table 4.2: **Empirical sizes of the constancy test for Gaussian and non-Gaussian responses based on 5000 simulation runs for non-parametric regression tests in longitudinal studies.** Pseudo-LRT refers to the pseudo likelihood ratio test approach developed by Greven et al. (2008); F-test refers to the generalized F-test proposed by Wang and Chen (2012); Mixture χ^2refers to the likelihood ratio test with an approximated 50:50 mixture χ^2 asymptotic distribution.

Data Type	Test	Size		
		Linear	Quadratic	Cubic
	ScoreTest	0.049	0.051	0.050
Gaussian	F-test	0.048	0.051	0.052
	pseudo-LRT	0.049	0.051	0.051
	Mixture χ^2	0.033	0.036	0.038
Binary	ScoreTest	0.047	0.046	0.048
Binomial (N=6)	ScoreTest	0.051	0.052	0.051

4.4.3 Scenario 3 - Meta-analysis on Heterogeneous Effect-sizes

As outlined in Example 3 of Section 4.2.1, another application of the joint testing framework is to assess the heterogeneous genetic effects in meta-analysis. We carried out simulations to evaluate the performance of our proposed score test to existing meta-analysis approaches. The power comparison showed that under the existence of between-study heterogeneity, our method achieved very similar power to the LRT-based approach, which was developed to explicitly model the heterogeneity (Han and Eskin, 2011). Details of the simulation and comparison result can be found in Supplementary Materials Section 4.7.5.1.

4.5 Data Application

Genome-wide association studies (GWAS) have identified more than 200 genetic variants that influence blood lipid levels, such as the Low Density Lipoprotein Cholesterol (LDL-C) (Teslovich et al., 2010; Waterworth et al., 2010; Willer et al., 2013; Surakka et al., 2015; Spracklen et al., 2017). For these variants, an important question is whether their genetic effects vary with age. Knowing whether the variants act during very early life, childhood or in adulthood would improve our understanding of the genetics of lipid markers of cardiovascular disease. In addition, it would help explain the proportion of lipid heritability that are attributable to gene-age interaction. Previous evidence has shown that certain polymorphisms on LDL-C, a quantitative blood lipid marker, is age-dependent (Giolo et al., 2010; Shirts et al., 2011; Dumitrescu et al., 2011; Simino et al., 2014).

To illustrate performance of the proposed method in real data, we use the Michigan Genomics Initiative (MGI) data to investigate whether genetic effects to LDL-C can be modified by age. Launched in 2012, MGI was a biorepository effort to create a longitudinal cohort of participants in Michigan Medicine. It enrolled participants undergoing anesthesia prior to a surgery or diagnostic procedure, creating a patient community with genome-wide data, electronic health information, and permission for follow-up and re-contact in future studies. To date, more than 50,000 participants have been recruited through Michigan Medicine health system while awaiting diagnostic or interventional procedures either during a preoperative visit prior to the procedure or on the day of procedure that required anesthesia. Opt-in written informed consent was obtained. In addition to coded biosamples and protected secure health information, participants understood that all Electronic Health Records (EHRs), claims, and national data sources linkable to the participant may be incorporated into the MGI databank. Each participant donated a blood sample for genetic analysis, underwent baseline vital signs and a comprehensive history and physical, and completed validated self-report measures of pain, mood and function, including NIH Patient Reported Outcomes Measurement Information System (PROMIS) measures. Data were collected according to Declaration of Helsinki principles. Study participants provided written informed consent, and protocols were reviewed and approved by local ethics committees (IRB ID HUM00099605) (Fritsche et al., 2018).

MGI genotyped the participants using the Illumina Human CoreExome v.12.1 array, which is a combined genotyping plus exome array of $>$ 500,000 targeted SNPs. The phased MGI genotypes (SHAPEIT2 on autosomal chromosomes and Eagle2 on chromosome X) were then imputed using Minimac3 with the Haplotype Reference Consortium (chromosomes 1-22: HRC release; chromosome X: HRC release 1.1). After excluding variants with low imputation quality ($R^2 < 0.3$), over 39 million quality-imputed genetic markers were obtained from the imputation. Since the imputation quality is low for very rare variants, we further filtered out the imputed variants with MAF $<$ 0.001, which resulted in around 13 million variants in our data analysis. For methodology illustration purpose, we retrieved the first LDL-C lab measurement from participants' longitudinal EHRs and used it as the response variable. Genotyped samples with missing LDL-C information are excluded from the analysis, which results in 11,016 subjects with complete LDL-C and genotype information.

For those 11,016 unrelated European participants, we first performed the main effect association analysis to relate the genetic variants to LDL-C. We assumed an additive genetic model for the variant and applied an inverse normal transformation to LDL-C, since the transformed variable more closely approximated the Gaussian distribution. Specifically, we used the following linear regression model to conduct the main effect association analysis:

$$y_i = \mathbf{S}_i^T \boldsymbol{\alpha} + G_i \beta_0 + \epsilon_i, \quad \text{for } i = 1, 2, \ldots, 11016, \tag{4.5.1}$$

where y_i is the inverse-normal-transformed LDL-C level for the ith subject; $\mathbf{S}_i$ is a vector of adjusting covariates including gender, age, age-squared and four principal components; G_i denotes the number of minor alleles ($G_i = 0, 1, 2$) of the variant to be tested. Among the 13 million variants tested for main genetic effect association, four reached the genome-wide significance level (i.e. p-value of β_0 is less than 5×10^{-8}), and three of those four SNPs have been identified by previous GWAS for their associations with LDL-C (Willer et al., 2013; Surakka et al., 2015; Spracklen et al., 2017). Table 4.3 listed the genetic information and association signals for those significant SNPs.

Table 4.3: **SNPs with significant main effect associations with LDL-C in the MGI data.** The main p-value refers to the hypothesis testing for $H_0 : \beta_0 = 0$ in Equation (4.5.1); the interaction p-value refers to the hypothesis testing for $H_0 : \beta(age) = 0$ in Equation (4.5.2); and the optimal ρ is the ρ value which yields the maximum score test value for our proposed test statistic in Section 4.3, Equation (4.3.6).

dbSNP ID	Position	Allele	MAF	Nearest Gene	Main P-Value	Interaction P-Value	Optimal ρ
rs646776	1: 109,818,530	C > T	0.19	*CELSR2*	7.823×10^{-12}	0.054	1
rs76681713	16: 72,333,346	T > C	0.13	*PMFBP1*	1.771×10^{-9}	0.401	0
rs6511720	19: 11,202,306	G > T	0.1	*LDLR*	2.124×10^{-9}	0.792	0
rs7412	19: 45,412,079	C > T	0.062	*APOE*	2.617×10^{-45}	3.642×10^{-4}	0.75

Among those 4 significant SNPs, we further incorporated age into a varying coefficient model (4.5.2) to investigate the possible age-dependent genetic effects on LDL-C. Specifically,

$$y_i = \mathbf{S}_i^T \alpha + G_i \beta_0 + \beta(age_i) G_i + \epsilon_i \quad \text{for} i = 1, 2, \ldots, 11016, \qquad (4.5.2)$$

where $\beta(age_i)$ is an unspecified varying coefficient with respect to age. We note that here the functional $\beta(age)$ should not include an intercept term, since the main genetic effect has already been accounted for in the $G_i\beta_0$ part of the model. Testing for the age-dependent genetic effect on LDL-C is equivalent to assessing $H_0 : \beta(age) = 0$ in Equation (4.5.2). We modeled the unspecified functional $\beta(age)$ using a quadratic smoothing spline estimator; the technique of mixed model representation for the smoothing spline estimator effectively transformed the hypothesis testing of $H_0 : \beta(age) = 0$ into the problem of joint testing the fixed and random effects as outlined in Section 4.2.1, Example 2.

Then we employed our proposed score test to assess for $H_0 : \beta(age) = 0$ under its LMM representation. After using Bonferroni correction for multiple testing, the resulting p-values suggested that the genetic effects for rs7412 appear to vary with respect to age (Table 4.3). For rs7412, we also illustrated in Figure 4.6 the effect of age on the LDL-C, with respect to different minor allele count values for this SNPs. Regression curves were imposed to summarize the effect of age on the trait. From Figure 4.6, it can observed that the variant has a stronger effect on LDL-C in adulthood (before age 55), and that the effect tends to attenuate in older adults (after age 55).

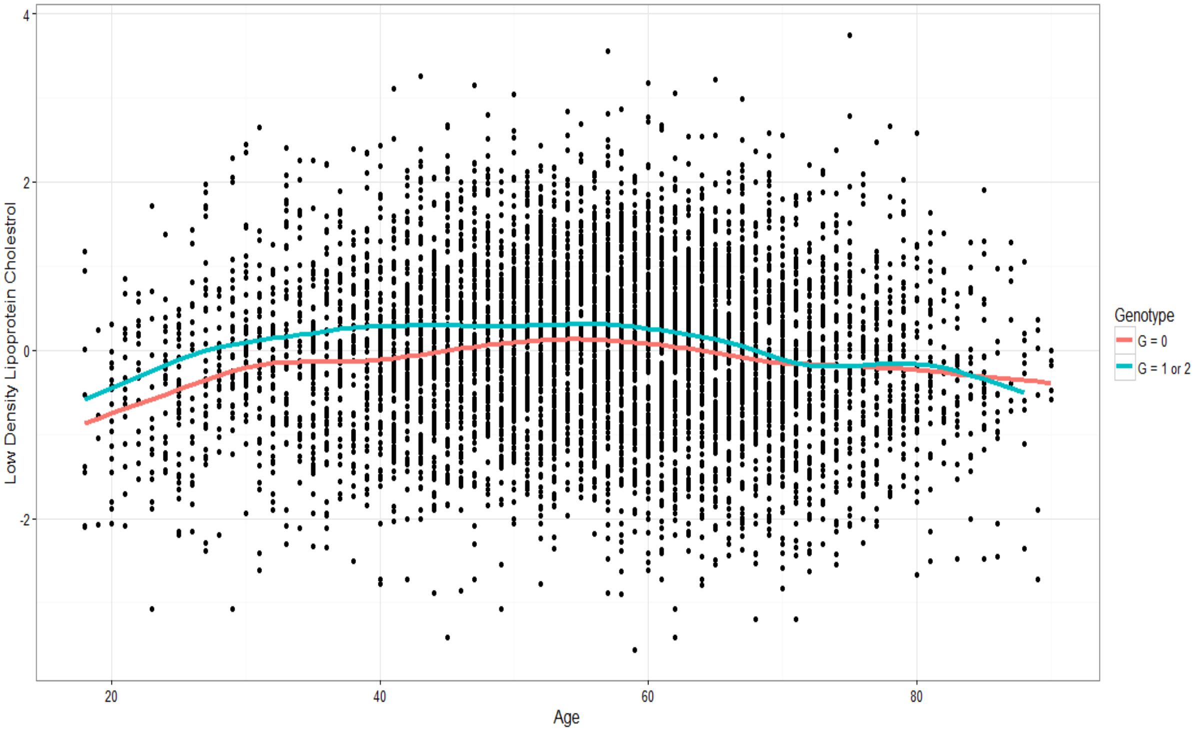


Figure 4.6: **Scatter plot of inverse-normal-transformed LDL-C v.s. age from the MGI data.** Values of inverse-normal-transformed LDL-C measurements plotted as a function of age in years. Separate regression curves are fitted to the data for the number of minor alleles (G = 0 in blue line v.s. G = 1 or 2 in red line), which suggests stronger genetic effect on LDL-C in adulthood (before age 55), and attenuated effect in older adults (after age 55).

This data application demonstrates that our proposed method can be a useful tool for obtaining empirical evidence of whether the genetic effect on a phenotype is being modified by risk factors such as age. In addition, it confirms earlier findings that the genotype-associated differences in LDL cholesterol can be age-dependent (Giolo et al., 2010; Shirts et al., 2011; Dumitrescu et al., 2011; Simino et al., 2014)

4.6 Discussions

In this chapter, we developed a score test for jointly testing the fixed and random effects in a GLMM. We address the following four issues in the construction of the score test statistic: (a) a convenient re-parameterization which reformulates the joint testing problem into testing the variance component alone; (b) the conversion of non-Gaussian outcomes into pseudo-Gaussian outcomes through Laplace approximation; (c) the asymptotic equivalence of the score test to the LRT for joint testing; (d) the p-value calculation from the asymptotic distribution of the score test. An R package 'JointScoreTest' has been developed to implement our proposed method and can be downloaded at the website https://sites.google.com/a/umich.edu/leeshawn/software.

The key idea is to re-parameterize the fixed effects parameter into the product of the random effects variance component and a nuisance parameter vector; such re-parameterization reformulates the null hypothesis into testing the variance component alone. The nuisance parameter is unidentifiable under the null, to remove the unknown nuisance parameter, we propose a test using the supremum of the standardized score statistic over the unknown parameter space. By applying the method of Lagrange multipliers, we reduce the test statistic from searching over the entire unconstrained vector space into a constrained scalar space, which makes the test statistic easily obtainable. For non-Gaussian responses, to avoid the high-dimensional integration, we employ the Laplace approximation to reform the outcomes into pseudo-Gaussian responses and re-apply the Gaussian responses derivations to construct the test statistic accordingly. Besides its capability of handling both Gaussian and non-Gaussian responses, we show that the asymptotic null distribution of our score test is equivalent to the LRT.

Unlike the LRT and F-test, which require calculating the maximum likelihood estimator under both the null and alternative hypothesis, the score test only requires calculating the maximum likelihood under the null, which makes it computationally less expensive given many tests. Our proposed score test does not have a standard asymptotic distribution, we hence approximate the asymptotic null distribution using a computationally efficient resampling-based copula method. The copula approximation provides scalable computing time, which makes it feasible to conduct the joint testing in large scale experiments. In our power simulations, to run 5000 iterations, our score test takes 1.11 hours on average on a Linux cluster node with 2.80 GHz CPU.

We suggest using a coarse grid $\rho = (0, 0.25, 0.5, 0.75, 1)$ for simulations and real data analysis. In fact, in the three simulation scenarios, we also use a dense grid of ρ with 50 equally spaced points from 0 to 1, and observe that the dense grid does not meaningfully increase power (data not shown).

In both the simulation studies and data application, we reformulate the problem into the joint testing framework by representing the SAMM and varying-coefficient model as a working GLMM. Although we focus on using the penalized natural smoothing spline estimator to represent the SAMM and varying-coefficient model, the results are not limited to such a representation. Other types of basis functions such as truncated polynomials and B-splines can also be employed. When assessing whether a functional form holds constant value or not, we observe some power differences depending on the order of the natural spline estimator for the nonparametric function. In practice, when the true underlying model is unknown, our method does not provide a guideline on how to choose the appropriate order of natural splines to avoid power loss. We have left this for future research.

4.7 Supplementary Materials

4.7.1 Mixed-effect Representation of the Natural Spline Estimator for the Non-parametric Function

To model data with clustered structures, Zhang and Lin (2003) proposed using the semiparametric additive mixed model (SAMM), which allows a predictor to be associated with the outcome through a nonparametric function. In this section, we summarize how to construct the mixed-effect representation of the natural spline estimator for the nonparametric function.

Denote y_{ij} as the response variable for the jth observation in the ith cluster; $\mathbf{s}_{ij}$'s as the $p \times 1$ vectors of fixed effects covariates with coefficients $\boldsymbol{\alpha}$; $\mathbf{z}_{ij}$'s as the $q \times 1$ vectors of random effects covaraites with cluster specific coefficients $\mathbf{b}_i$; and $\mathbf{b}_i \sim N(\mathbf{0}, D_0(\boldsymbol{\theta}))$ with nuisance variance component vector $\boldsymbol{\theta}$. Assume that conditional on the unobserved random effects vectors $\mathbf{b}_i$, the outcome y_{ij}s are independent with means $E(y_{ij}|\mathbf{b}_i) = \mu_{ij}$ and variances $var(y_{ij}|\mathbf{b}_i) = V(\mu_{ij}) = \phi\zeta_{ij}^{-1}\nu(\mu_{ij})$, where ϕ is a scale parameter, ζ_{ij} is a prior weight and $\nu(\cdot)$ is a variance function. Under those assumptions, the conditional mean μ_{ij} in a SAMM takes the form

$$g(\mu_{ij}) = f(t_{ij}) + \mathbf{s}_{ij}^T\boldsymbol{\alpha} + \mathbf{z}_{ij}^T\mathbf{b}_i, \tag{4.7.1}$$

where $g(\cdot)$ is a known link function, and $f(t)$ is an arbitrary smooth function relating the scalar covariate t to the outcome. In addition, it is assumed that there exists a positive integer h such that $f(t)$ has absolutely continuous derivatives up to the order $h-1$, and the area under the absolute function of $f(t)$'s hth order derivative is bounded.

It can be shown that $f(t)$ can be estimated as a natural spline of order h. Here, we consider the smoothing spline representation of the natural spline estimator. Without lost of generality, we assume that t_{ij}s are all bounded in $[0,1]$, and there are r distinct values of t_{ij} with $0 < t_1^0 < \ldots < t_r^0 < 1$. Denote $\{\phi_k(t) = t^{k-1}/(k-1)!\}_{k=1}^h$ as the basis for the polynomial space of order $h-1$, and $R(t,t_l^0)=\int_0^1 (t_l^0-u)_+^{h-1}(t-u)_+^{h-1}du/[(h-1)!]^2$, where $(t-u)_+ = t-u$ if $t \geq u$ and 0 otherwise. The hth order smoothing spline estimator for $f(t)$ can be written as

$$f(t) = \sum_{k=1}^{h} \beta_k \phi_k(t) + \sum_{l=1}^{r} a_l R(t,t_l^0), \tag{4.7.2}$$

or equivalently in the matrix form

$$f(\mathbf{t}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\mathbf{a}, \tag{4.7.3}$$

where $\mathbf{t} = (t_{11},\ldots,t_{1n_1},\ldots,t_{m1},\ldots,t_{mn_m})^T$, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_h)^T$, $\mathbf{a} = (a_1,\ldots,a_r)^T \sim N(\mathbf{0},\tau\boldsymbol{\Sigma})$, and $\mathbf{X} = \mathbf{N}\mathbf{T}$, $\mathbf{U} = \mathbf{N}\boldsymbol{\Sigma}^{-1}$ with

$$\mathbf{N} = \begin{pmatrix} I(t_{11}=t_1^0) & I(t_{11}=t_2^0) & \cdots & I(t_{11}=t_r^0) \\ \vdots & \vdots & \vdots & \vdots \\ I(t_{1n_1}=t_1^0) & I(t_{1n_1}=t_2^0) & \cdots & I(t_{1n_1}=t_r^0) \\ \vdots & \vdots & \vdots & \vdots \\ I(t_{mn_1}=t_1^0) & I(t_{mn_1}=t_2^0) & \cdots & I(t_{mn_1}=t_r^0) \\ \vdots & \vdots & \vdots & \vdots \\ I(t_{mn_m}=t_1^0) & I(t_{mn_m}=t_2^0) & \cdots & I(t_{mn_m}=t_r^0) \end{pmatrix}_{n\times r},$$

$$\mathbf{T} = \begin{pmatrix} \phi_1(t_1^0) & \phi_2(t_1^0) & \cdots & \phi_h(t_1^0) \\ \phi_1(t_2^0) & \phi_2(t_2^0) & \cdots & \phi_h(t_2^0) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_1(t_r^0) & \phi_2(t_r^0) & \cdots & \phi_h(t_r^0) \end{pmatrix}_{r\times r}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} R(t_1^0,t_1^0) & R(t_1^0,t_2^0) & \cdots & R(t_1^0,t_r^0) \\ R(t_2^0,t_1^0) & R(t_2^0,t_2^0) & \cdots & R(t_2^0,t_r^0) \\ \vdots & \vdots & \vdots & \vdots \\ R(t_r^0,t_1^0) & R(t_r^0,t_2^0) & \cdots & R(t_r^0,t_r^0) \end{pmatrix}_{r\times r}.$$

Under the mixed-effect representation of $f(\mathbf{t})$, the SAMM in (4.7.1) can be represented in the following working generalized linear mixed model (GLMM):

$$\begin{aligned} g(\boldsymbol{\mu}) &= \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\alpha} + \mathbf{U}\mathbf{a} + \mathbf{Z}\mathbf{b}, \\ \mathbf{a} &\sim N(\mathbf{0},\tau\boldsymbol{\Sigma}),\ \mathbf{b} \sim N(\mathbf{0},\mathbf{D}(\boldsymbol{\theta})), \end{aligned} \tag{4.7.4}$$

where $\boldsymbol{\mu} = (\mu_{11},\ldots,\mu_{1n_1},\ldots,\mu_{m1},\ldots,\mu_{mn_m})$ is the vectorized conditional mean values; in which μ_{ij} represents the conditional mean value for the outcome variable y_{ij} $(i \in \{1,\ldots,m\}, j \in \{1,\ldots,n_i\}, n = \sum_{i=1}^m n_i)$; $\mathbf{S}$ is the fixed effects design matrix with the ith row block being $\mathbf{s}_i = (\mathbf{s}_{i1}, \ldots, \mathbf{s}_{in_i})^T$; and $\mathbf{Z} = diag\{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$ is the random effects design matrix with $\mathbf{z}_i = (z_{i1}, \ldots, z_{in_i})^T$, $\mathbf{b} = (\mathbf{b}_1^T, \ldots, \mathbf{b}_m^T)^T \sim N(\mathbf{0}, \mathbf{D}(\boldsymbol{\theta}))$, and $\mathbf{D}(\boldsymbol{\theta}) = diag(D_0(\boldsymbol{\theta}), \ldots, D_0(\boldsymbol{\theta}))$.

4.7.2 Using Lagrange Multipliers to Determine the Form of $\boldsymbol{\gamma}$

The proposed test statistic,

$$\begin{aligned} T &= \sup_{\boldsymbol{\gamma}} \frac{U_\tau(\hat{\boldsymbol{\nu}}) - E(U_\tau(\hat{\boldsymbol{\nu}}))}{\sqrt{VAR(U_\tau(\hat{\boldsymbol{\nu}}))}} \\ &= \sup_{\boldsymbol{\gamma}} \frac{-\frac{1}{2}tr[\hat{\mathbf{V}}^{-1}\mathbf{PU\Sigma U}^T] + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{U\Sigma U}^T + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\boldsymbol{\gamma}}{\sqrt{\frac{1}{2}tr[(\hat{\mathbf{V}}^{-1}\mathbf{PU\Sigma U})^2] + \boldsymbol{\gamma}^T\mathbf{X}^T\tilde{\mathbf{P}}\hat{\mathbf{V}}\tilde{\mathbf{P}}\mathbf{X}\boldsymbol{\gamma}}}, \qquad (4.7.5) \end{aligned}$$

requires searching over the entire space of $\boldsymbol{\gamma}$ in order to obtain the supremum. The searching would be very challenging when the dimension of $\boldsymbol{\gamma}$ is high. An alternative way to solve this problem is to view the test as follows:

Maximize $U_\tau(\hat{\boldsymbol{\nu}}) - E(U_\tau(\hat{\boldsymbol{\nu}}))$ subject to $VAR(U_\tau(\hat{\boldsymbol{\nu}})) = c,$ where c is a constant.

From this point of view, Lagrange multipliers can be applied to find the form of $\boldsymbol{\gamma}$:

$$\begin{aligned} \text{Let } \Delta(\boldsymbol{\gamma}, \lambda) &= -\frac{1}{2}tr[\hat{\mathbf{V}}^{-1}\mathbf{PU\Sigma U}^T] + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{U\Sigma U}^T + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\boldsymbol{\gamma} \\ &- \lambda\{\frac{1}{2}tr[(\hat{\mathbf{V}}^{-1}\mathbf{PU\Sigma U})^2] + \boldsymbol{\gamma}^T\mathbf{X}^T\tilde{\mathbf{P}}\hat{\mathbf{V}}\tilde{\mathbf{P}}\mathbf{X}\boldsymbol{\gamma} - c\} \\ \Rightarrow \frac{\partial\Delta}{\partial\boldsymbol{\gamma}} &= \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{X} - 2\lambda\boldsymbol{\gamma}^T\mathbf{X}^T\tilde{\mathbf{P}}\hat{\mathbf{V}}\tilde{\mathbf{P}}\mathbf{X} = 0 \\ \text{and } \frac{\partial\Delta}{\partial\boldsymbol{\gamma}} &= -\{\frac{1}{2}tr[(\hat{\mathbf{V}}^{-1}\mathbf{PU\Sigma U})^2] + \boldsymbol{\gamma}^T\mathbf{X}^T\tilde{\mathbf{P}}\hat{\mathbf{V}}\tilde{\mathbf{P}}\mathbf{X}\boldsymbol{\gamma} - c\} = 0 \\ \Rightarrow \boldsymbol{\gamma} &= \eta(\mathbf{X}^T\tilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T\tilde{\mathbf{P}}\mathbf{y}, \end{aligned}$$

where η is a non-negative scalar representing the strength of $\boldsymbol{\gamma}$.

4.7.3 Asymptotic Distribution of S_ρ

Recall, for a given ρ, S_ρ can be written as:

$$\begin{aligned}
S_\rho &= \rho\mathbf{y}^T\widetilde{\mathbf{P}}^T\mathbf{X}(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{y} + \frac{1-\rho}{2}\mathbf{y}^T\widetilde{\mathbf{P}}^T\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\widetilde{\mathbf{P}}\mathbf{y} \\
&= \mathbf{y}^T\widetilde{\mathbf{P}}^T\{\rho\mathbf{X}(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T + \frac{1-\rho}{2}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\}\widetilde{\mathbf{P}}\mathbf{y}, \\
\text{where } \widetilde{\mathbf{P}} &= \mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}}\mathbf{V}^{-\frac{1}{2}} \text{ with } \hat{\mathbf{P}} = \mathbf{I} - \mathbf{V}^{-\frac{1}{2}}\mathbf{S}(\mathbf{S}^T\mathbf{V}^{-1}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{V}^{-\frac{1}{2}}. \\
\text{Thus, } S_\rho &= \mathbf{y}^T(\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}}\mathbf{V}^{-\frac{1}{2}})^T\{\rho\mathbf{X}(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T + \frac{1-\rho}{2}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\}(\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}}\mathbf{V}^{-\frac{1}{2}})\mathbf{y} \\
&= (\mathbf{V}^{-\frac{1}{2}}\mathbf{y})^T\{\rho\hat{\mathbf{P}}\mathbf{V}^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}} + \frac{1-\rho}{2}\hat{\mathbf{P}}\mathbf{V}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}}\}(\mathbf{V}^{-\frac{1}{2}}\mathbf{y}).
\end{aligned}$$

Under the null hypothesis, $\mathbf{V}^{-\frac{1}{2}}\mathbf{y}$ approximately follows a multivariate normal distribution $MVN(\mathbf{0}, \mathbf{I}_n)$, where $\mathbf{I}_n$ is an identity matrix with dimension n. Consequently, S_ρ asymptotically follows a mixture of chi-square distribution $\sum \lambda_l \chi^2_{l,1}$, where $\chi^2_{l,1}$s are chi-square distributions with one degree of freedom; and λ_ls are the non-zero eigenvalues of $\rho\hat{\mathbf{P}}\mathbf{V}^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}^T\widetilde{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}} + \frac{1}{2}(1-\rho)\hat{\mathbf{P}}^T\mathbf{V}^{-\frac{1}{2}}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{P}}$.

4.7.4 Asymptotic Equivalency of the Score Test to the Likelihood Ratio Test

In this section, we adopt the asymptotic properties from Zhu and Zhang (2006) to establish the asymptotic equivalence of our proposed score test statistic to the corresponding LRT for the joint testing of fixed and random effects in a GLMM. We demonstrate the equivalence for Gaussian outcomes, and the conclusions for non-Gaussian responses will follow naturally using the same working linear mixed model and Laplace approximation techniques as in the test statistic derivation for non-Gaussian outcomes.

Recall, for Gaussian outcomes, our mixed model can be written as

$$\begin{aligned}
\mathbf{y} &= \tau\mathbf{X}\boldsymbol{\gamma} + \mathbf{S}\boldsymbol{\alpha} + \mathbf{U}\mathbf{a} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \qquad (4.7.6) \\
\text{where } \mathbf{a} &\sim N(\mathbf{0}, \tau\boldsymbol{\Sigma}),\ \mathbf{b} \sim N(\mathbf{0}, \mathbf{D}(\boldsymbol{\theta})),\ \boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n),
\end{aligned}$$

and our goal is to test $H_0 : \tau = 0$. Denote $\boldsymbol{\nu} = (\boldsymbol{\alpha}^T, \boldsymbol{\theta}^T, \sigma^2)^T$ with true value $\boldsymbol{\nu}_* = (\boldsymbol{\alpha}_*^T, \boldsymbol{\theta}_*^T, \sigma_*^2)^T$. We assume that $\boldsymbol{\nu}_*$ is an interior point of the parameter space Ξ and the length of vector $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ is $q_1 \times 1$ and $q_2 \times 1$ respectively. Denote $\boldsymbol{\xi} = (\tau, \boldsymbol{\nu})$ with true value $\boldsymbol{\xi}_*$, and the corresponding log-likelihood function as $L_N(\boldsymbol{\xi}|\boldsymbol{\gamma})$. Define $\hat{\boldsymbol{\xi}}$ as the maximum likelihood estimate of $\boldsymbol{\xi}$ under H_0. One can easily see that $\hat{\boldsymbol{\xi}}$ does not depend on $\boldsymbol{\gamma}$ under H_0, since $L_N(\boldsymbol{\xi}|\boldsymbol{\gamma})$ is independent of $\boldsymbol{\gamma}$ when $\tau = 0$. Define $\tilde{\boldsymbol{\xi}}(\boldsymbol{\gamma})$ as the maximum likelihood estimate of $\boldsymbol{\xi}$ for any $\boldsymbol{\gamma} \in \Gamma$ under H_1. The log-likelihood function for $\boldsymbol{\xi}$ can be written as :

$$\begin{aligned}
\frac{\partial L_N(\boldsymbol{\xi}|\boldsymbol{\gamma})}{\partial \boldsymbol{\xi}} &= U_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}|\boldsymbol{\xi}) = \begin{pmatrix} U_{\tau,N}(\boldsymbol{\gamma}|\boldsymbol{\xi}) \\ U_{\boldsymbol{\nu},N}(\boldsymbol{\gamma}|\boldsymbol{\xi}) \end{pmatrix}, \\
\text{with score } U_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}) &= U_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}|\boldsymbol{\xi})|_{\boldsymbol{\xi}=\boldsymbol{\xi}_*} = \begin{pmatrix} U_{\tau,N}(\boldsymbol{\gamma}) \\ U_{\boldsymbol{\nu},N}(\boldsymbol{\gamma}) \end{pmatrix}, \\
\text{and information matrix } J_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}) &= -E\{\partial^2 L_N(\boldsymbol{\xi}|\boldsymbol{\gamma})/\partial\boldsymbol{\xi}\partial\boldsymbol{\xi}^T\}|_{\boldsymbol{\xi}=\boldsymbol{\xi}_*}.
\end{aligned}$$

In order to establish the asymptotic properties, Zhu and Zhang (2006) required that the following regularity conditions to be satisfied.

Assumption S.1 $\sup_{\boldsymbol{\gamma}\in\Gamma} \parallel \tilde{\xi}(\boldsymbol{\gamma}) - \xi_* \parallel \to 0$ and $\parallel \hat{\xi} - \xi_* \parallel \to 0$ in probability.

Assumption S.2 Assume that

$$O_p(1) = L_N(\boldsymbol{\xi}|\boldsymbol{\gamma}) = L_N(\boldsymbol{\xi}_*|\boldsymbol{\gamma}) + \sqrt{N}(\boldsymbol{\xi}-\boldsymbol{\xi}_*)^T U_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}) - \frac{N}{2}(\boldsymbol{\xi}-\boldsymbol{\xi}_*)^T J_{\boldsymbol{\xi},N}(\boldsymbol{\gamma})(\boldsymbol{\xi}-\boldsymbol{\xi}_*)^T + o_p(1)$$

holds uniformly for all $\sqrt{N} \parallel \boldsymbol{\xi} - \boldsymbol{\xi}_* \parallel \leq c_0$, where c_0 is any positive scalar. In addition, $\sup_{\boldsymbol{\gamma}\in\Gamma} \parallel U_{\boldsymbol{\xi},N}(\boldsymbol{\xi}|\boldsymbol{\gamma}) \parallel = O_p(1)$, and $c_2 \geq \sup_{\boldsymbol{\gamma}\in\Gamma} \mu_{min}[J_{\boldsymbol{\xi},N}(\boldsymbol{\gamma})] \geq \inf_{\boldsymbol{\gamma}\in\Gamma}[J_{\boldsymbol{\xi},N}(\boldsymbol{\gamma})] \geq 4c_1^2$ holds almost surely for some fixed c_1 and c_2, where μ_{max} and μ_{min} represents the minimum and maximum eigenvalue of a matrix.

Assumption S.3 $(U_{\boldsymbol{\xi},N}(\cdot), J_{\boldsymbol{\xi},N}(\cdot)) \Rightarrow (U_{\boldsymbol{\xi}}(\cdot), J_{\boldsymbol{\xi}}(\cdot))$, where $\Rightarrow$ denotes weak convergence of a stochastic process under the uniform metric, $(U_{\boldsymbol{\xi}}(\cdot), J_{\boldsymbol{\xi}}(\cdot))$ has bounded continuous sample paths with probability one. Moreover, the $(q_1+q_2+1)\times(q_1+q_2+1)$ matrix $J_{\boldsymbol{\xi}}(\cdot)$ is symmetric and $\infty > \sup_{\boldsymbol{\gamma}\in\Gamma} \mu_{max}[J_{\boldsymbol{\xi}}(\boldsymbol{\gamma})] \geq \inf_{\boldsymbol{\gamma}\in\Gamma} \mu_{min}[J_{\boldsymbol{\xi}}(\boldsymbol{\gamma})] > 0$

holds almost surely.

In addition to those assumptions, we define the following notations in order to facilitate the asymptotic properties:

$$\begin{aligned}
U_{\boldsymbol{\xi}}(\boldsymbol{\gamma}) &= \begin{pmatrix} U_\tau(\boldsymbol{\gamma}) \\ U_{\boldsymbol{\nu}}(\boldsymbol{\gamma}) \end{pmatrix}, \quad J_{\boldsymbol{\xi}}(\boldsymbol{\gamma}) = \begin{pmatrix} J_{\tau\tau}(\boldsymbol{\gamma}) & J_{\tau\boldsymbol{\nu}}(\boldsymbol{\gamma}) \\ J_{\boldsymbol{\nu}\tau}(\boldsymbol{\gamma}) & J_{\boldsymbol{\nu}\boldsymbol{\nu}}(\boldsymbol{\gamma}) \end{pmatrix}, \\
z_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}) &= J_{\boldsymbol{\xi},N}^{-1}(\boldsymbol{\gamma})U_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}) = \begin{pmatrix} z_{\tau,N}(\boldsymbol{\gamma}) \\ z_{\boldsymbol{\nu},N}(\boldsymbol{\gamma}) \end{pmatrix}, \quad z_{\boldsymbol{\xi}}(\boldsymbol{\gamma}) = J_{\boldsymbol{\xi}}^{-1}(\boldsymbol{\gamma})U_{\boldsymbol{\xi}}(\boldsymbol{\gamma}) = \begin{pmatrix} z_\tau(\boldsymbol{\gamma}) \\ z_{\boldsymbol{\nu}}(\boldsymbol{\gamma}) \end{pmatrix}, \\
\text{and } \mathbf{e}_1 &= (1, 0, \ldots, 0)^T \in \mathbf{R}^{q_1+q_2+1}.
\end{aligned}$$

Under the assumptions S.1 - S.3, Zhu and Zhang (2006) established the following theorems for the asymptotic null distribution of LRT as well as the score test statistic for testing $H_0 : \tau = 0$ against H_1.

Theorem S.1 Suppose assumptions S.1 - S.3 hold. Then, under the null hypothesis, the asymptotic null distribution of LRT is

$$\sup_{\boldsymbol{\gamma}\in\Gamma}[L_N(\tilde{\boldsymbol{\xi}}(\boldsymbol{\gamma}))|\boldsymbol{\gamma}] - L_N(\hat{\boldsymbol{\xi}}) \Rightarrow 0.5 \sup_{\boldsymbol{\gamma}\in\Gamma} z_\tau(\boldsymbol{\gamma})^2/(\mathbf{e}_1^T J_{\boldsymbol{\xi}}^{-1}(\boldsymbol{\gamma})\mathbf{e}_1).$$

Theorem S.2 Suppose assumptions S.1 - S.3 hold. Then the score statistic S_s for testing H_0 against H_1 has the following asymptotic null distribution:

$$S_s = \sup_{\boldsymbol{\gamma}\in\Gamma}\{U_{\boldsymbol{\xi},N}^T(\boldsymbol{\gamma}|\hat{\boldsymbol{\xi}})J_{\boldsymbol{\xi},N}^{-1}(\boldsymbol{\gamma}|\hat{\boldsymbol{\xi}})U_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}|\hat{\boldsymbol{\xi}})\} = \sup_{\boldsymbol{\gamma}\in\Gamma}\{\frac{z_\tau(\boldsymbol{\gamma})}{\sqrt{\mathbf{e}_1^T J_{\boldsymbol{\xi}}^{-1}(\boldsymbol{\gamma})\mathbf{e}_1}}\}^2 + o_p(1). \quad (4.7.7)$$

Together, Theorem S.1 and S.2 establish the asymptotic equivalence of LRT and score test statistic S_s. We now show that our score statistic T defined in Equation (4.3.3) has the same form as S_s in (4.7.7). Consequently, its asymptotic equivalence with the corresponding LRT would follow naturally from Theorem S.1 and S.2.

The log-likelihood under the LMM (4.7.6) is

$$\begin{aligned}
l(\boldsymbol{\xi}) &= -\frac{N}{2\pi} - \frac{1}{2}log|\tau\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T + \mathbf{Z}\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T + \sigma^2\mathbf{I}| \\
&- \frac{1}{2}(\mathbf{y} - \tau\mathbf{X}\boldsymbol{\gamma} - \mathbf{S}\boldsymbol{\alpha})^T(\tau\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T + \mathbf{Z}\mathbf{D}(\boldsymbol{\theta})\mathbf{Z}^T + \sigma^2\mathbf{I})^{-1}(\mathbf{y} - \tau\mathbf{X}\boldsymbol{\gamma} - \mathbf{S}\boldsymbol{\alpha}).
\end{aligned}$$

It can be shown that

$$U_{\boldsymbol{\xi},N}(\gamma|\hat{\boldsymbol{\xi}}) = \begin{pmatrix} \frac{\partial L_N(\boldsymbol{\xi}|\boldsymbol{\gamma})}{\partial \tau}|_{\tau=0,\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} \\ \\ \frac{\partial L_N(\boldsymbol{\xi}|\boldsymbol{\gamma})}{\partial \boldsymbol{\alpha}}|_{\tau=0,\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} \\ \\ \frac{\partial L_N(\boldsymbol{\xi}|\boldsymbol{\gamma})}{\partial \boldsymbol{\theta}}|_{\tau=0,\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} \\ \\ \frac{\partial L_N(\boldsymbol{\xi}|\boldsymbol{\gamma})}{\partial \sigma^2}|_{\tau=0,\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}} \end{pmatrix} = \begin{pmatrix} U_{\tau,N}(\boldsymbol{\gamma}|\tau=0,\boldsymbol{\nu}=\hat{\boldsymbol{\nu}}) \\ \\ \mathbf{0}_{p_1\times 1} \\ \\ \mathbf{0}_{p_2\times 1} \\ \\ 0 \end{pmatrix},$$

where $U_{\tau,N}(\boldsymbol{\gamma}|\tau = 0, \boldsymbol{\nu} = \hat{\boldsymbol{\nu}}) = -\frac{1}{2}tr[\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T] + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\hat{\mathbf{V}}^{-1}\mathbf{P}\mathbf{y} + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\boldsymbol{\gamma}$. And $J_{\tau\tau,N}(\boldsymbol{\gamma}|\tau = 0, \boldsymbol{\nu} = \hat{\boldsymbol{\nu}}) = \frac{1}{2}tr[(\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U})^2] + \boldsymbol{\gamma}^T\mathbf{X}^T\tilde{\mathbf{P}}\hat{\mathbf{V}}\tilde{\mathbf{P}}\mathbf{X}\boldsymbol{\gamma}$, with $\hat{\mathbf{V}} = \mathbf{Z}\mathbf{D}\mathbf{Z}^T + \hat{\sigma}^2\mathbf{I}$, $\mathbf{P} = \mathbf{I} - \mathbf{S}(\mathbf{S}^T\hat{\mathbf{V}}^{-1}\mathbf{S})\mathbf{S}^T\hat{\mathbf{V}}^{-1}$, and $\tilde{\mathbf{P}} = \hat{\mathbf{V}}^{-1}\mathbf{P}$. Plug in these values into equation (4.7.7), we obtain

$$\begin{aligned} S_s &= \sup_{\boldsymbol{\gamma}\in\Gamma} U_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}|\hat{\boldsymbol{\xi}})^T J_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}|\hat{\boldsymbol{\xi}})^{-1} U_{\boldsymbol{\xi},N}(\boldsymbol{\gamma}|\hat{\boldsymbol{\xi}}) \\ &= \sup_{\boldsymbol{\gamma}\in\Gamma}\{\sqrt{U_{\tau,N}(\boldsymbol{\gamma}|\tau=0,\boldsymbol{\nu}=\hat{\boldsymbol{\nu}})\cdot J^{-1}_{\tau\tau,N}(\boldsymbol{\gamma}|\tau=0,\boldsymbol{\nu}=\hat{\boldsymbol{\nu}})\cdot U_{\tau,N}(\boldsymbol{\gamma}|\tau=0,\boldsymbol{\nu}=\hat{\boldsymbol{\nu}})}\}^2 \\ &= \sup_{\boldsymbol{\gamma}\in\Gamma}\{\frac{-\frac{1}{2}tr[\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T] + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T\hat{\mathbf{V}}^{-1}\mathbf{P}\mathbf{y} + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\boldsymbol{\gamma}}{\sqrt{\frac{1}{2}tr[(\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U})^2] + \boldsymbol{\gamma}^T\mathbf{X}^T\tilde{\mathbf{P}}\hat{\mathbf{V}}\tilde{\mathbf{P}}\mathbf{X}\boldsymbol{\gamma}}}\}^2. \quad (4.7.8) \end{aligned}$$

To account for the fact that $\boldsymbol{\nu} = (\boldsymbol{\alpha}^T, \boldsymbol{\theta}^T, \sigma^2)^T$ is estimated by its MLE $\hat{\boldsymbol{\nu}} = (\hat{\boldsymbol{\alpha}}^T, \hat{\boldsymbol{\theta}}^T, \hat{\sigma}^2)^T$, Zhang and Lin (2003) proposed to use a bias-corrected score statistic S_R for S_s, in which inverse of the marginal covariance matrix $\hat{V}^{-1} = (\mathbf{Z}\mathbf{D}\mathbf{Z}^T + \hat{\sigma}^2\mathbf{I})^{-1}$ for the response $\mathbf{y}$ is replaced by the projection matrix $\tilde{\mathbf{P}} = \hat{\mathbf{V}}^{-1}\mathbf{P}$. Applying this bias-corrected version to the score statistic in (4.7.8), we obtain

$$S_R = \sup_{\boldsymbol{\gamma}\in\Gamma}\{\frac{-\frac{1}{2}tr[\hat{\mathbf{V}}^{-1}\mathbf{P}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T] + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T + \mathbf{y}^T\mathbf{P}^T\hat{\mathbf{V}}^{-1}\mathbf{X}\boldsymbol{\gamma}}{\sqrt{\frac{1}{2}tr[(\hat{\mathbf{V}}^{-1}\mathbf{P}\mathbf{U}\boldsymbol{\Sigma}\mathbf{U})^2] + \boldsymbol{\gamma}^T\mathbf{X}^T\tilde{\mathbf{P}}\hat{\mathbf{V}}\tilde{\mathbf{P}}\mathbf{X}\boldsymbol{\gamma}}}\}^2,$$

which is just the square of our proposed score statistic T in Equation (4.3.3).

4.7.5 Additional Simulations

4.7.5.1 Scenario 3 - Meta-analysis on heterogeneous effect-sizes

Meta-analysis is a practical and effective tool for combining multiple association studies into a single aggregate analysis in order to identify association signals with small effect sizes. To account for the effect size heterogeneity across studies, several approaches have been proposed under the framework of jointly testing the fixed and random effects in an LMM. For example, the random-effects meta-analysis method (RE) is developed to explicitly model the heterogeneity; however, it implicitly assumes heterogeneity under the null hypothesis, which causes power loss. To relax the conservative assumption of RE, Han and Eskin (2011) proposed a new random-effects method (RE-HE) under the LMM to appropriately model the expected heterogeneity between different studies. Specification of the model set-up is as outlined in Example 3 in Section 4.2.1. RE-HE tests the null hypothesis $H_0 : \mu = 0$ and $\tau = 0$ using the likelihood ratio test approach and assesses the strength of association signals based on pre-tabulated p-values. When heterogeneity exists, Han and Eskin (2011) demonstrated in their simulation studies that RE-HE achieves higher statistical power than RE and the traditional fixed-effects meta-analysis method (FE), which assumes constant effect-size across all studies.

In order to generate a realistic spectrum of the effect-size estimates and the corresponding standard errors, we adapted the neoadjuvant chemotherapy meta-analysis data from the Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration (for Locally Advanced Cervical Cancer Meta-analysis Collaboration et al., 2003). The meta-analysis contained data from 11 trials to compare neoadjuvant chemotherapy followed by radical radiotherapy versus the same radiotherapy alone. The forest plot in the left panel of Supplementary Figure 4.7 summarized the log hazard ratios of the neoadjuvant chemotherapy and the associated 95% CIs from the

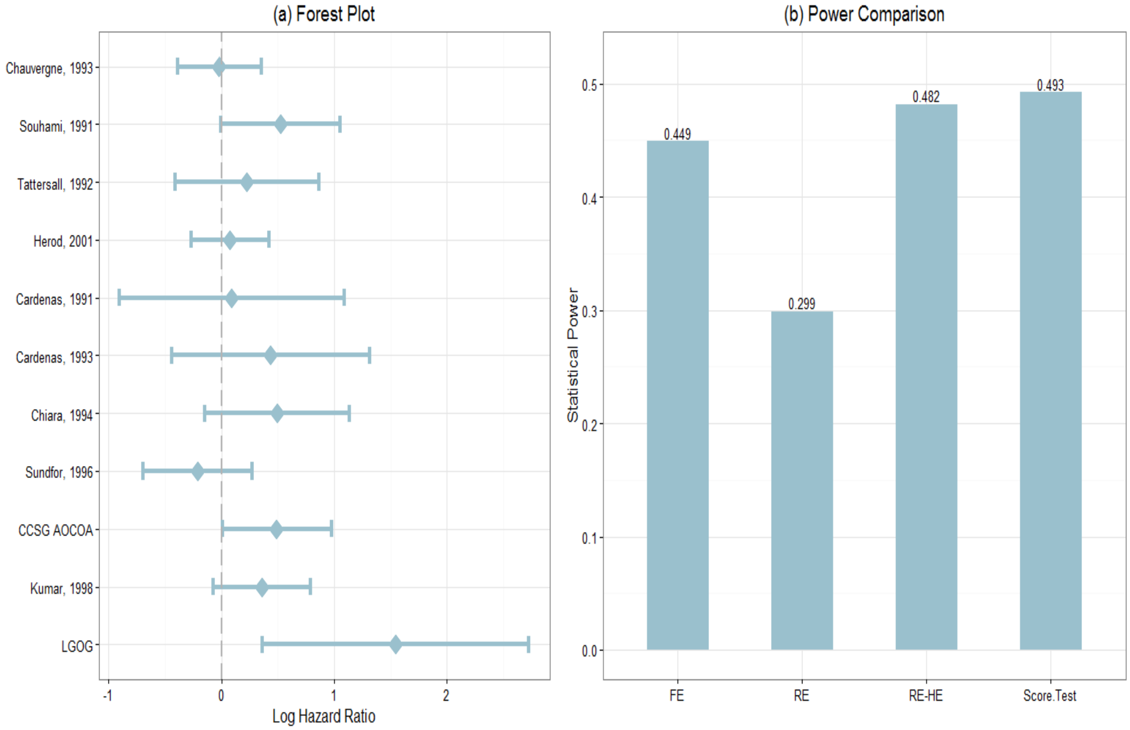


Figure 4.7: **Meta-analysis simulation results.** Left panel: Forest plot of the log hazard ratios and 95% CIs of the neoadjuvant chemotherapy from the 11 trials in the Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration (2003). Right panel: Empirical power for score test and existing methods for the meta-analysis simulations.

11 trials. The forest plot showed varying effect sizes among the trials, which indicates the possible existence of between-study heterogeneity.

Under the Cox proportional hazard model, the time-independent hazard rate for subject k with explanatory variable Z_k has the form $\lambda(Z_k) = \lambda_0 \exp\{\beta Z_k\}$, where λ_0 denotes the hazard function of the radical radiotherapy; β denotes the effect size for neoadjuvant chemotherapy; Z_k is an indicator variable with value 1 if subject k receives neoadjuvant chemotherapy followed by radical radiotherapy, and value 0 if the subject only receives the radiotherapy. Re-arranging the Cox proportional hazard model, we have:

$$\begin{aligned}\log(\text{hazard ratio of the neoadjuvant therapy}) &= \log\{\frac{\lambda(Z_k = 1)}{\lambda_0}\} \\ &= \beta \cdot (Z_k = 1),\end{aligned}$$

which suggests that the log hazard ratio can be linked to the neoadjuvant chemotherapy assignment via a linear model. Thus, we propose to simulate $\hat{\beta}$, the effect-size estimates of the log hazard ratio through the following approach:

Step 1: For the ith trial, generate $\epsilon_{ij} \sim N(0, \sigma_i^2)$, where $i = 1, \ldots, 11$; $j = 1, \ldots, n_i$; σ_i^2 is the listed variance value in the left panel of Figure 1 (on Page 2478) in the Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration (2003) ; and n_i is the total sample size of the ith trial.

Step 2: For the ith trial, denote n_{i1} and n_{i2} as the number of participants with and without the neoadjuvant therapy respectively, and Z_{ij} as the treatment group indicator for each of the jth participant in the ith trial. Under those notations, we have $Z_{ij} = 1$ for $j = 1, \ldots, n_{i1}$ and $Z_{ij} = 0$ for $j = 1 + n_{i1}, \ldots, n_{i1} + n_{i2}$. Obtain $y_{ij} = \beta_i Z_{ij} + \epsilon_{ij}$ for $j = 1, \ldots, n_{i1}$, where value of β_i is equal to the log hazard ratio as summarized in the forest plot of Supplementary Figure 4.7. Here, we consider y_{ij} as an observed log hazard ratio.

Step 3: Regress y_{ij} on Z_{ij} to obtain the simulated log hazard ratio $\hat{\beta}_i$ and the associated standard error for the ith trial.

Step 4: Repeat Steps 1 to 3 for each $i \in \{1, \ldots, 11\}$.

Following the above outlined steps, we carried out 5000 simulations and perform meta-analysis using FE, RE, RE-HE and our score test . The right panel in Supplementary Figure 4.7 summarized the power comparison results. The barplot showed that our score test yields very similar but slightly higher power to the LRT-based RE-HE; unsurprisingly, FE and RE were less powerful at identifying the association signal due to the expected between-study heterogeneity.

CHAPTER V

Summary and Discussion

5.1 Summary

The first two projects in this dissertation ware motivated by the observation that the efficiency of genetic association studies – whether by genotyping or by sequencing – depends critically on sample size (Chatterjee et al., 2013). Trans-ethnic meta-analyses are increasingly being used to boost study power by enlarging the total sample size. In addition to the improved power of disease/trait locus discovery in trans-ethnic studies, differences in LD levels across genetically diverse populations are potentially a powerful tool for fine mapping the rare or causal variants that underlie disease associations (Kichaev and Pasaniuc, 2015). Despite the promising potential, however, the between-study genetic effect heterogeneity among different ethnic groups presents challenges in performing trans-ethnic meta-analysis. Since traditional GWAS meta-analysis approaches do not appropriately account for the expected between-study heterogeneity, in this dissertation, I have proposed two novel statistical methods for modeling the effect-size heterogeneity in genetic association studies.

In Chapter II, I developed a score test to detect the common variant associations in trans-ethnic meta-analysis. To account for the expected genetic effect heterogeneity across diverse populations, I adapted a modified random effects model from the kernel regression framework. I specifically constructed the correlation structure for the genetic effect coefficients to reflect the level of genetic effect similarities across ancestry groups. Through analytical approximation of the asymptotic distribution of the proposed test, I achieved efficient computing time for genome-wide datasets.

In Chapter III, I extended the score test in Chapter II to the gene- or region-based rare variant trans-ethnic meta-analysis in sequencing association studies. The proposed method again uses the kernel regression framework to construct the modified random effects model, and incorporates the genetic similarities across ancestry groups into modeling the heterogeneity structure of the genetic effect coefficients. To enable efficient estimation of p-values, I employed a resampling-based copula method to estimate the asymptotic null distribution of the proposed test.

The last project in this dissertation was motivated by the problem of testing an unspecified varying-coefficient for assessing the possible modification of non-genetic factors on the effect of genetic variants. I generalized the problem into a framework of jointly testing the fixed and random effects in a GLMM, so that such a unified framework is applicable in many different scenarios in biomedical studies. To the best of my knowledge, it is the only systematic research which addresses the joint testing problem with respect to both the Gaussian and non-Gaussian outcomes under the presence of nuisance variance components. More specifically, in Chapter IV, I developed a supremum score test for jointly testing the fixed and random effects in a generalized linear mixed model (GLMM) for both Gaussian and non-Gaussian outcomes. The modified random effect model (RE-HE) by Han and Eskin (2011) and the gene-based rare variant association test (SKAT-O) by Lee et al. (2012) can both be viewed as special cases of this general framework of jointly testing the fixed and random effects. The proposed method can also be employed in the nonparametric test of spline curves as well as in assessing the significance of the varying-coefficient component in the varying-coefficient model. In terms of its application in genetic association studies, one can use the test to investigate whether the genetics effects to any disease/trait of interest can be modified by confounders such as age.

5.2 Future Plans

The parallel advancement of DNA array platform, high-throughput sequencing technology and genotyping imputation accuracy has opened a new era of genomics and molecular biology. Availability of the high throughput data at relatively low cost enables large-scale biobanks to genotype or sequence hundreds of thousands of participants. For example, the HUNT study includes large total population-based cohorts, covering 125,000 Norwegian participants (Krokstad et al., 2012); the UK Biobank project is a large prospective cohort study of 500,000 participants from across the United Kingdom (Bycroft et al., 2017); the National Heart, Lung and Blood Institute (NHLBI) launched their TOPMed (Trans-Omics for Precision Medicine) program has collected over 120,000 individual genomes for its WGS project (Lung et al., 2016). In these biobank data, in addition to genotyping/sequencing the participants' genomes, either carefully designed surveys or electronic health records (EHRs) together with the International Classification of Disease (ICD) billing codes are employed to obtain participants' phenotypic data and health-related information.

Despite the available rich variety of genotypic, phenotypic and health-related information in those biobanks, computing summary statistics from each of the enormously large-scale GWAS to conduct trans-ethnic meta-analysis poses new statistical and computational challenges. For example, given that most of the biobank data are based on cohort study designs, one challenge of using EHR-derived phenotypes is that most of them are dichotomized with imbalanced ($<$ 1:10) or extremely imbalanced ($<$1:100) case-control ratios. Standard asymptotic tests (such as the score test) for assessing the rare variant associations typically approximate the asymptotic null distribution using a Gaussian density. However, under the imbalanced case-control ratio, the normal approximation will yield highly inflated type-I error rates, since the underlying distribution is highly skewed and thus cannot be well-captured by the symmetric Gaussian density. In addition, those large-scale cohort studies tend to have a diverse mixture of family structures and/or contain samples with both familial and unrelated individuals. Consequently, methods that can accommodate familial and cryptic relatedness are needed. Unfortunately, there exist few approaches which can handle the sample relatedness and/or type-I error inflation due to imbalanced case-control ratio. Efficient algorithms are needed to retain the computing time for hundreds of millions of variants on hundreds of thousands of samples in a scalable fashion.

The SAIGE method, developed by Zhou et al. (2017), is currently the only available mixed model approach which is practical for large-scale phenome-wide association studies (PheWAS) while controlling for case-control imbalance and correcting for sample relatedness. It uses the saddlepoint approximation (SPA) to calibrate the distribution of score statistics, and utilizes optimization strategies such as the preconditioned conjugate gradient (PCG) approach to reduce the computational burden and memory cost. As a next step, I will seek to incorporate the SPA and optimization strategies used in the SAIGE method into my trans-ethnic meta-analysis approaches, to further boost the study power by taking advantages of data from the ever larger cohorts, additional phenotypes and wider ethnic groups.

5.3 Closing Remarks and Perspective

Over the last decade, multi-ethnic studies have proved instrumental to unraveling the genetic complexities of disease risks In particular, trans-ethnic meta-analysis are increasingly being used for locus replication and discovery, as well as fine-mapping of causal variants associated with complex diseases. One key advantage of using trans-ethnic meta-analysis is to boost study power by leveraging LD structure and the underlying differential genetic architecture across disparate ancestral human genomes. Based upon this concept, central to my thesis research is the goal of maximize study power for locus discoveries when there is significant inter-study heterogeneity.

In the past few years, in response to the criticisms of the limited utility of GWAS-findings, the genomics community has gradually shifted its focus to causal or functional variant identifications. The widely available and economically feasible re-sequencing technologies have made it possible to conduct locus fine-mapping through trans-ethnic GWASs. By including populations with more diverse haplotypes, such as the African population, trans-ethnic GWASs can help pinpoint the causal or functional variants of interest and identify candidate gene mutations. Findings from several global genomics consortia have demonstrated that trans-ethnic fine-mapping studies can identify functional gene mutations as well as increase the total variance explained by the identified loci (Galarneau et al., 2010; Sanna et al., 2011; Franceschini et al., 2012; Wu et al., 2013; Liu et al., 2014; Saunders et al., 2014).

New findings from trans-ethnic studies will enrich our understanding of the genetic basis of complex diseases/traits. Although it is not a simple task to interpret GWAS findings, given that most GWAS signals are either in the intronic or intergenic non-coding regions of the genome, integration of multiple "omics" resources, such as epigenetic features, eQTLs, tissue-specific transcript expressions, chromatin conformation can help advance the identification of functional or mechanistic variations in the post-GWAS era. The continued expansion of GWAS, and its integration with other efforts capturing the molecular function of the human genome, will be a critical asset for the study of gene coding and regulatory mechanisms and how they contribute to complex diseases/traits. Disentangling the mechanism by which genotype influences phenotype will ultimately lead to the identification of important biological pathways and presentation of suitable targets for drug development and repositioning of known therapeutics. Continuing steps toward filling the knowledge gap will bring us closer to elucidating disease etiology and offer opportunities of innovative preventative and therapeutic strategies in precision medicine.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abecasis, G. R., Cherny, S. S., Cookson, W. O., and Cardon, L. R. (2002). Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nature genetics* **30,** 97.

Amberger, J. S., Bocchini, C. A., Schiettecatte, F., Scott, A. F., and Hamosh, A. (2014). Omim. org: Online mendelian inheritance in man (omim®), an online catalog of human genes and genetic disorders. *Nucleic acids research* **43,** D789–D798.

Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* **69,** 683–734.

Bansal, V., Libiger, O., Torkamani, A., and Schork, N. J. (2010). Statistical analysis strategies for association studies involving rare variants. *Nature Reviews Genetics* **11,** 773.

Bustamante, C. D., Francisco, M., and Burchard, E. G. (2011). Genomics for the world. *Nature* **475,** 163.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2017). Genome-wide genetic data on˜ 500,000 uk biobank participants. *bioRxiv* page 166298.

Campbell, C. D., Ogburn, E. L., Lunetta, K. L., Lyon, H. N., Freedman, M. L., Groop, L. C., Altshuler, D., Ardlie, K. G., and Hirschhorn, J. N. (2005). Demonstrating stratification in a european american population. *Nature genetics* **37,** 868.

Cantor, R. M., Lange, K., and Sinsheimer, J. S. (2010). Prioritizing gwas results: a review of statistical methods and recommendations for their application. *The American Journal of Human Genetics* **86,** 6–22.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Lane, C. R., Lim, E. P., Kalyanaraman, N., Nemesh, J., et al. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature genetics* **22,** 231.

Carlson, C. S., Matise, T. C., North, K. E., Haiman, C. A., Fesinmeyer, M. D., Buyske, S., Schumacher, F. R., Peters, U., Franceschini, N., Ritchie, M. D., et al. (2013). Generalization and dilution of association results from european gwas in populations of non-european ancestry: the page study. *PLoS biology* **11,** e1001661.

Chakravarti, A. (1999). Population genetics—making sense out of sequence. *Nature genetics* **21,** 56–60.

Chatterjee, N., Wheeler, B., Sampson, J., Hartge, P., Chanock, S. J., and Park, J.-H. (2013). Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nature genetics* **45,** 400.

Chen, H., Meigs, J. B., and Dupuis, J. (2013). Sequence kernel association test for quantitative traits in family samples. *Genetic epidemiology* **37,** 196–204.

Chen, H., Wang, C., Conomos, M. P., Stilp, A. M., Li, Z., Sofer, T., Szpiro, A. A., Chen, W., Brehm, J. M., Celedón, J. C., et al. (2016). Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. *The American Journal of Human Genetics* **98,** 653–666.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics* **10,** 101–129.

Consortium, I. H. . et al. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* **467,** 52.

Cooper, R. S., Tayo, B., and Zhu, X. (2008). Genome-wide association studies: implications for multiethnic samples. *Human molecular genetics* **17,** R151–R155.

Coram, M. A., Duan, Q., Hoffmann, T. J., Thornton, T., Knowles, J. W., Johnson, N. A., Ochs-Balcom, H. M., Donlon, T. A., Martin, L. W., Eaton, C. B., et al. (2013). Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. *The American Journal of Human Genetics* **92,** 904–916.

Crainiceanu, C., Ruppert, D., Claeskens, G., and Wand, M. P. (2005). Exact likelihood ratio tests for penalised splines. *Biometrika* **92,** 91–103.

Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66,** 165–185.

Cruchaga, C., Karch, C. M., Jin, S. C., Benitez, B. A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., et al. (2014). Rare coding variants in the phospholipase d3 gene confer risk for alzheimer's disease. *Nature* **505,** 550.

Davies, R. B. (1980). Algorithm as 155: The distribution of a linear combination of χ 2 random variables. *Applied Statistics* pages 323–333.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74,** 33–43.

Dichgans, M., Malik, R., König, I. R., Rosand, J., Clarke, R., Gretarsdottir, S., Thorleifsson, G., Mitchell, B. D., Assimes, T. L., Levi, C., et al. (2014). Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants. *Stroke* **45,** 24–36.

Dumitrescu, L., Brown-Gentry, K., Goodloe, R., Glenn, K., Yang, W., Kornegay, N., Pui, C.-H., Relling, M. V., and Crawford, D. C. (2011). Evidence for age as a modifier of genetic associations for lipid levels. *Annals of human genetics* **75,** 589–597.

Dumitrescu, L., Carty, C. L., Taylor, K., Schumacher, F. R., Hindorff, L. A., Ambite, J. L., Anderson, G., Best, L. G., Brown-Gentry, K., Buzkova, P., et al. (2011). Genetic determinants of lipid traits in diverse populations from the population architecture using genomics and epidemiology (page) study. *PLOS genetics* **7,** e1002138.

Eichler, E. E., Flint, J., Gibson, G., Kong, A., Leal, S. M., Moore, J. H., and Nadeau, J. H. (2010). Missing heritability and strategies for finding the underlying causes of complex disease. *Nature Reviews Genetics* **11,** 446.

Evangelou, E. and Ioannidis, J. P. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics* **14,** 379–389.

Farrer, L. A., Cupples, L. A., Haines, J. L., Hyman, B., Kukull, W. A., Mayeux, R., Myers, R. H., Pericak-Vance, M. A., Risch, N., and Van Duijn, C. M. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein e genotype and alzheimer disease: a meta-analysis. *Jama* **278,** 1349–1356.

for Locally Advanced Cervical Cancer Meta-analysis Collaboration, N. C. et al. (2003). Neoadjuvant chemotherapy for locally advanced cervical cancer: a systematic review and meta-analysis of individual patient data from 21 randomised trials. *European journal of cancer (Oxford, England: 1990)* **39,** 2470.

Franceschini, N., Fox, E., Zhang, Z., Edwards, T. L., Nalls, M. A., Sung, Y. J., Tayo, B. O., Sun, Y. V., Gottesman, O., Adeyemo, A., et al. (2013). Genome-wide association analysis of blood-pressure traits in african-ancestry individuals reveals common associated genes in african and non-african populations. *The American Journal of Human Genetics* **93,** 545–554.

Franceschini, N., Van Rooij, F. J., Prins, B. P., Feitosa, M. F., Karakas, M., Eckfeldt, J. H., Folsom, A. R., Kopp, J., Vaez, A., Andrews, J. S., et al. (2012). Discovery and fine mapping of serum protein loci through transethnic meta-analysis. *The American Journal of Human Genetics* **91,** 744–753.

Frazer, K. A., Murray, S. S., Schork, N. J., and Topol, E. J. (2009). Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics* **10,** 241.

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N., Gabriel, S. B., Topol, E. J., Smoller, J. W., Pato, C. N., et al. (2004). Assessing the impact of population stratification on genetic association studies. *Nature genetics* **36,** 388–393.

Fritsche, L. G., Gruber, S. B., Wu, Z., Schmidt, E. M., Zawistowski, M., Moser, S. E., Blanc, V. M., Brummett, C. M., Kheterpal, S., Abecasis, G. R., and Mukherjee, B. (2018). Association of polygenic risk scores for multiple cancers in a phenome-wide study: Results from the michigan genomics initiative. *bioRxiv* .

Fuchsberger, C., Flannick, J., Teslovich, T. M., Mahajan, A., Agarwala, V., Gaulton, K. J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D. J., et al. (2016). The genetic architecture of type 2 diabetes. *Nature* .

Galarneau, G., Palmer, C. D., Sankaran, V. G., Orkin, S. H., Hirschhorn, J. N., and Lettre, G. (2010). Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. *Nature genetics* **42,** 1049.

Giolo, S. R., Pereira, A. C., De Andrade, M., Krieger, J. E., and Soler, J. P. (2010). Evaluating gene by sex and age interactions on cardiovascular risk factors in brazilian families. *BMC medical genetics* **11,** 132.

Greven, S., Crainiceanu, C. M., Küchenhoff, H., and Peters, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics* **17,** 870–891.

Gudmundsson, J., Sulem, P., Gudbjartsson, D. F., Masson, G., Agnarsson, B. A., Benediktsdottir, K. R., Sigurdsson, A., Magnusson, O. T., Gudjonsson, S. A., Magnusdottir, D. N., et al. (2012). A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer. *Nature genetics* **44,** 1326.

Han, B. and Eskin, E. (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *The American Journal of Human Genetics* **88,** 586–598.

Hauke, J. and Kossowski, T. (2011). Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data. *Quaestiones geographicae* **30,** 87–93.

Hedges, L. V. and Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological methods* **3,** 486.

Higgins, J. P., Thompson, S. G., et al. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in medicine* **21,** 1539–1558.

Hirata, M., Suzuki, M., Ishii, R., Satow, R., Uchida, T., Kitazumi, T., Sasaki, T., Kitamura, T., Yamaguchi, H., Nakamura, Y., et al. (2011). Genetic defect in phospholipase cδ1 protects mice from obesity by regulating thermogenesis and adipogenesis. *Diabetes* **60,** 1926–1937.

Hu, X. (2011). Plcd1 (phospholipase c, delta 1).

Hu, Y.-J., Berndt, S. I., Gustafsson, S., Ganna, A., Hirschhorn, J., North, K. E., Ingelsson, E., Lin, D.-Y., of ANthropometric Traits (GIANT) Consortium, G. I., et al. (2013). Meta-analysis of gene-level associations for rare variants based on single-variant statistics. *The American Journal of Human Genetics* **93,** 236–248.

Ionita-Laza, I., Capanu, M., De Rubeis, S., McCallum, K., and Buxbaum, J. D. (2014). Identification of rare causal variants in sequence-based studies: methods and applications to vps13b, a gene involved in cohen syndrome and autism. *PLoS genetics* **10,** e1004729.

Kan, M., Zhou, D., Zhang, D., Zhang, Z., Chen, Z., Yang, Y., Guo, X., Xu, H., He, L., and Liu, Y. (2010). Two susceptible diabetogenic variants near/in mtnr1b are associated with fasting plasma glucose in a han chinese cohort. *Diabetic Medicine* **27,** 598–602.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association* **90,** 773–795.

Kichaev, G. and Pasaniuc, B. (2015). Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *The American Journal of Human Genetics* **97,** 260–271.

Kote-Jarai, Z., Al Olama, A. A., Giles, G. G., Severi, G., Schleutker, J., Weischer, M., Campa, D., Riboli, E., Key, T., Gronberg, H., et al. (2011). Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study. *Nature genetics* **43,** 785.

Krokstad, S., Langhammer, A., Hveem, K., Holmen, T., Midthjell, K., Stene, T., Bratberg, G., Heggland, J., and Holmen, J. (2012). Cohort profile: the hunt study, norway. *International journal of epidemiology* **42,** 968–977.

Lasky-Su, J., Himes, B., Raby, B., Klanderman, B., Sylvia, J. S., Lange, C., Melen, E., Martinez, F., Israel, E., Gauderman, J., et al. (2012). Hla-dq strikes again: genome-wide association study further confirms hla-dq in the diagnosis of asthma among adults. *Clinical & Experimental Allergy* **42,** 1724–1733.

Lee, S., Teslovich, T. M., Boehnke, M., and Lin, X. (2013). General framework for meta-analysis of rare variants in sequencing association studies. *The American Journal of Human Genetics* **93,** 42–53.

Lee, S., Wu, M. C., and Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13,** 762–775.

Li, Y. R. and Keating, B. J. (2014). Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome medicine* **6,** 91.

Liang, K.-Y. and Self, S. G. (1996). On the asymptotic behaviour of the pseudo-likelihood ratio test statistic. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 785–796.

Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84,** 309–326.

Liu, C.-T., Buchkovich, M. L., Winkler, T. W., Heid, I. M., Consortium, A. A. A. G., Consortium, G., Borecki, I. B., Fox, C. S., Mohlke, K. L., North, K. E., et al. (2014). Multi-ethnic fine-mapping of 14 central adiposity loci. *Human molecular genetics* **23,** 4738–4744.

Liu, D., Lin, X., and Ghosh, D. (2007). Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics* **63,** 1079–1088.

Liu, D. J., Peloso, G. M., Zhan, X., Holmen, O. L., Zawistowski, M., Feng, S., Nikpay, M., Auer, P. L., Goel, A., Zhang, H., et al. (2014). Meta-analysis of gene-level tests for rare variant association. *Nature genetics* **46,** 200–204.

Lung, N. H., Institute, B., et al. (2016). Trans-omics for precision medicine (topmed) program.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2016). The new nhgri-ebi catalog of published genome-wide association studies (gwas catalog). *Nucleic acids research* **45,** D896–D901.

Madsen, B. E. and Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet* **5,** e1000384.

Mahajan, A., Go, M. J., Zhang, W., Below, J. E., Gaulton, K. J., Ferreira, T., Horikoshi, M., Johnson, A. D., Ng, M. C., Prokopenko, I., et al. (2014). Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature genetics* **46,** 234.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. *Nature* **461,** 747.

Marigorta, U. M. and Navarro, A. (2013). High trans-ethnic replicability of gwas results implies common causal variants. *PLoS genetics* **9,** e1003566.

McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P., and Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics* **9,** 356–369.

Morris, A. P. (2011). Transethnic meta-analysis of genomewide association studies. *Genetic epidemiology* **35,** 809–822.

Morris, A. P. and Zeggini, E. (2010). An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genetic epidemiology* **34,** 188–193.

Morton, N. and Collins, A. (1998). Tests and estimates of allelic association in complex inheritance. *Proceedings of the National Academy of Sciences* **95,** 11389–11393.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506,** 376.

Peloso, G. M., Auer, P. L., Bis, J. C., Voorman, A., Morrison, A. C., Stitziel, N. O., Brody, J. A., Khetarpal, S. A., Crosby, J. R., Fornage, M., et al. (2014). Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *The American Journal of Human Genetics* **94,** 223–232.

Popejoy, A. B. and Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature* **538,** 161.

Price, A. L., Kryukov, G. V., de Bakker, P. I., Purcell, S. M., Staples, J., Wei, L.-J., and Sunyaev, S. R. (2010). Pooled association tests for rare variants in exon-resequencing studies. *The American Journal of Human Genetics* **86,** 832–838.

Pritchard, J. K. and Przeworski, M. (2001). Linkage disequilibrium in humans: models and data. *The American Journal of Human Genetics* **69,** 1–14.

Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science* **273,** 1516–1517.

Rönn, T., Wen, J., Yang, Z., Lu, B., Du, Y., Groop, L., Hu, R., and Ling, C. (2009). A common variant in mtnr1b, encoding melatonin receptor 1b, is associated with type 2 diabetes and fasting plasma glucose in han chinese individuals. *Diabetologia* **52,** 830–833.

Sanna, S., Li, B., Mulas, A., Sidore, C., Kang, H. M., Jackson, A. U., Piras, M. G., Usala, G., Maninchedda, G., Sassu, A., et al. (2011). Fine mapping of five loci associated with low-density lipoprotein cholesterol detects variants that double the explained heritability. *PLoS genetics* **7,** e1002198.

Saunders, E. J., Dadaev, T., Leongamornlert, D. A., Jugurnauth-Little, S., Tymrakiewicz, M., Wiklund, F., Al Olama, A. A., Benlloch, S., Neal, D. E., Hamdy, F. C., et al. (2014). Fine-mapping the hoxb region detects common variants tagging a rare coding allele: evidence for synthetic association in prostate cancer. *PLoS genetics* **10,** e1004129.

Scheipl, F., Greven, S., and Küchenhoff, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational statistics & data analysis* **52,** 3283–3299.

Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82,** 605–610.

Shi, J. and Lee, S. (2016). A novel random effect model for gwas meta-analysis and its application to trans-ethnic meta-analysis. *Biometrics* .

Shirts, B. H., Hasstedt, S. J., Hopkins, P. N., and Hunt, S. C. (2011). Evaluation of the gene–age interactions in hdl cholesterol, ldl cholesterol, and triglyceride levels: the impact of the sort1 polymorphism on ldl cholesterol levels is age dependent. *Atherosclerosis* **217,** 139–141.

Shlyakhter, I., Sabeti, P. C., and Schaffner, S. F. (2014). Cosi2: an efficient simulator of exact and approximate coalescent with selection. *Bioinformatics* **30,** 3427–3429.

Siddiq, A., Couch, F. J., Chen, G. K., Lindström, S., Eccles, D., Millikan, R. C., Michailidou, K., Stram, D. O., Beckmann, L., Rhie, S. K., et al. (2012). A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Human molecular genetics* **21,** 5373–5384.

Simino, J., Kume, R., Kraja, A. T., Turner, S. T., Hanis, C. L., Sheu, W. H.-H., Chen, Y.-D. I., Jaquish, C. E., Cooper, R. S., Chakravarti, A., et al. (2014). Linkage analysis incorporating gene–age interactions identifies seven novel lipid loci: The family blood pressure program. *Atherosclerosis* **235,** 84–93.

Sparsø, T., Bonnefond, A., Andersson, E., Bouatia-Naji, N., Holmkvist, J., Wegner, L., Grarup, N., Gjesing, A. P., Banasik, K., Cavalcanti-Proença, C., et al. (2009). G-allele of intronic rs10830963 in mtnr1b confers increased risk of impaired fasting glycemia and type 2 diabetes through an impaired glucose-stimulated insulin release studies involving 19,605 europeans. *Diabetes* **58,** 1450–1456.

Spracklen, C. N., Chen, P., Kim, Y. J., Wang, X., Cai, H., Li, S., Long, J., Wu, Y., Wang, Y. X., Takeuchi, F., et al. (2017). Association analyses of east asian individuals and trans-ancestry analyses with european individuals reveal new loci associated with cholesterol and triglyceride levels. *Human molecular genetics* **26,** 1770–1784.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* pages 1171–1177.

Stumpf, M. P. and Goldstein, D. B. (2003). Demography, recombination hotspot intensity, and the block structure of linkage disequilibrium. *Current Biology* **13,** 1–8.

Sun, J., Zheng, Y., and Hsu, L. (2013). A unified mixed-effects model for rare-variant association in sequencing studies. *Genetic epidemiology* **37,** 334–344.

Surakka, I., Horikoshi, M., Mägi, R., Sarin, A.-P., Mahajan, A., Lagou, V., Marullo, L., Ferreira, T., Miraglio, B., Timonen, S., et al. (2015). The impact of low-frequency and rare variants on lipid levels. *Nature genetics* **47,** 589.

Tang, Z.-Z. and Lin, D.-Y. (2014). Meta-analysis of sequencing studies with heterogeneous genetic associations. *Genetic epidemiology* **38,** 389–401.

Teo, Y.-Y., Small, K. S., and Kwiatkowski, D. P. (2010). Methodological challenges of genome-wide association analysis in africa. *Nature Reviews Genetics* **11,** 149.

Teslovich, T. M., Musunuru, K., Smith, A. V., Edmondson, A. C., Stylianou, I. M., Koseki, M., Pirruccello, J. P., Ripatti, S., Chasman, D. I., Willer, C. J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466,** 707–713.

Thomas, D. C. and Witte, J. S. (2002). Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiology and Prevention Biomarkers* **11,** 505–512.

Vlassi, M., Gazouli, M., Paltoglou, G., Christopoulos, P., Florentin, L., Kassi, G., and Mastorakos, G. (2012). The rs10830963 variant of melatonin receptor mtnr1b is associated with increased risk for gestational diabetes mellitus in a greek population. *Hormones (Athens)* **11,** 70–6.

Wang, X., Chua, H.-X., Chen, P., Ong, R. T.-H., Sim, X., Zhang, W., Takeuchi, F., Liu, X., Khor, C.-C., Tay, W.-T., et al. (2013). Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies. *Human molecular genetics* page ddt064.

Wang, Y. and Chen, H. (2012). On testing an unspecified function through a linear mixed effects model with multiple variance components. *Biometrics* **68,** 1113–1125.

Waterworth, D. M., Ricketts, S. L., Song, K., Chen, L., Zhao, J. H., Ripatti, S., Aulchenko, Y. S., Zhang, W., Yuan, X., Lim, N., et al. (2010). Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arteriosclerosis, thrombosis, and vascular biology* **30,** 2264–2276.

Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., Ganna, A., Chen, J., Buchkovich, M. L., Mora, S., et al. (2013). Discovery and refinement of loci associated with lipid levels. *Nature genetics* **45,** 1274.

Wright, S. (1949). The genetical structure of populations. *Annals of eugenics* **15,** 323–354.

Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics* **89,** 82–93.

Wu, Y., Waite, L. L., Jackson, A. U., Sheu, W. H., Buyske, S., Absher, D., Arnett, D. K., Boerwinkle, E., Bonnycastle, L. L., Carty, C. L., et al. (2013). Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS genetics* **9,** e1003379.

Xianglin Li, D. Gaussian copula model. *Encyclopedia of Quantitative Finance* .

Zhang, D. and Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics* **4,** 57–74.

Zhou, W., Nielsen, J. B., Fritsche, L. G., Dey, R., Elvestad, M. B., Wolford, B. N., LeFaive, J., VandeHaar, P., Gifford, A., Bastarache, L. A., et al. (2017). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *bioRxiv* page 212357.

Zhu, H., Zhang, H., et al. (2006). Generalized score test of homogeneity for mixed effects models. *The Annals of Statistics* **34,** 1545–1569.