

**MULTI-MARKER GENETIC ASSOCIATION AND INTERACTION TESTS
FOR SURVIVAL OUTCOMES**

By

Di Wu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biostatistics – Doctor of Philosophy

2022

ABSTRACT

MULTI-MARKER GENETIC ASSOCIATION AND INTERACTION TESTS FOR SURVIVAL OUTCOMES

By

Di Wu

With advancements in high-throughput technologies, studies have been conducted to investigate the role of massive genetic variants in human diseases. While multi-marker tests have been developed for binary and continuous disease outcomes, there are few such tests available for time-to-event outcomes. The existing tests have various drawbacks, including slow computation speed, being conservative in small samples, incapability of dealing with confounding, etc. To facilitate the genetic association and interaction analyses of time-to-event outcomes, we develop four suites of novel multi-marker survival tests for genetic association and interaction. The new tests address all the drawbacks of the existing tests. Furthermore, they can account for potential genetic heterogeneity to enhance power and deal with left truncation of survival data. Some of the new tests can handle competing risks, and some apply to interval-censored data. Simulation studies show that the new tests perform very well in finite samples of various sizes. When the genetic effect is heterogeneous across individuals/subpopulations, the new association tests considering genetic heterogeneity are more powerful than the existing tests, which do not account for genetic heterogeneity. Using the new methods, we performed genome-wide association analyses of 1) age to Alzheimer's disease data from the Religious Orders Study and the Rush Memory and Aging Project (ROSMAP) and 2) age to early childhood caries data from a dbGaP study, Dental Caries: Whole Genome Association and Gene x Environment Studies.

ACKNOWLEDGMENTS

I am sincerely grateful to my dissertation advisors, Dr. Chenxi Li and Dr. Qing Lu for their valuable guidance and inspiring comments. Without their supervision and support, it would be impossible for me to make great progress in my academic research and complete the dissertation. Also, I want to express my appreciation to my committee members, Dr. Yuehua Cui and Dr. Honglei Chen for their precious time and comments to improve this dissertation.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	xii
CHAPTER 1 Introduction	1
1.1 Introduction and literature review	1
1.2 Organization of this dissertation	4
CHAPTER 2 Multi-marker genetic association and interaction tests for survival out- comes based on weighted V statistics	6
2.1 Methods	6
2.1.1 Association tests	6
2.1.2 G-G/G-E interaction test	12
2.2 Simulations	13
2.2.1 Testing genetic association in the absence of genetic heterogeneity	14
2.2.2 Testing genetic association in the presence of genetic heterogeneity	17
2.2.3 Testing G-G/G-E interaction effect	23
2.2.4 Empirical size under stringent p-value thresholds	25
2.3 A Real Application	26
2.4 Discussion	29
CHAPTER 3 Multi-marker genetic association and interaction tests with interval-censored survival outcomes	33
3.1 Methods	33
3.1.1 Association test	33
3.1.2 G-G/G-E interaction test	37
3.2 Simulations	38
3.2.1 Testing genetic association in the absence of genetic heterogeneity	39
3.2.2 Testing genetic association in the presence of genetic heterogeneity	41
3.2.3 Testing G-G/G-E interaction effect	46
3.2.4 Empirical size under stringent p-value thresholds	48
3.3 A Real Application	49
3.4 Discussion	51
CHAPTER 4 Multi-marker genetic association and interaction tests based on the ac- celerated failure time model	53
4.1 Methods	53
4.1.1 Association tests	53
4.1.2 G-G/G-E interaction test	55
4.1.3 Small sample adjustment	57

4.2	Simulations	58
4.2.1	Association test in the absence of genetic heterogeneity	59
4.2.2	Association test in the presence of genetic heterogeneity	61
4.2.3	Testing G-G/G-E interaction effect	67
4.2.4	Empirical size under stringent p-value thresholds	70
4.2.5	Small-sample correction	70
4.3	A Real Application	73
4.4	Discussion	76
CHAPTER 5 Multi-marker genetic association and interaction tests based on the additive hazards model		79
5.1	Methods	79
5.1.1	Association test	79
5.1.2	G-G/G-E interaction test	81
5.1.3	Small sample adjustment	83
5.2	Simulations	84
5.2.1	Association test in the absence of genetic heterogeneity	85
5.2.2	Association test in the presence of genetic heterogeneity	87
5.2.3	Testing G-G/G-E interaction effect	93
5.2.4	Empirical size under stringent p-value thresholds	94
5.2.5	Small-sample correction	95
5.3	A Real Application	98
5.4	Discussion	101
CHAPTER 6 Concluding remarks		104
6.1	Remarks regarding selection of survival model	104
6.2	Future work	106
APPENDICES		107
APPENDIX A: Proofs for Chapter 2		108
APPENDIX B: Proofs for Chapter 4		111
APPENDIX C: Proofs for Chapter 5		116
BIBLIOGRAPHY		121

LIST OF TABLES

Table 2.1	Computation time (seconds) of WV, coxKM, gt and coxSKATs under the null hypothesis of no genetic association and $p = 5$	15
Table 2.2	Empirical size and power comparison between WV and coxKM in testing association of a SNP set with right-censored survival time in the absence of genetic heterogeneity, adjusting for covariates.	16
Table 2.3	Empirical size and power comparison between HWV and coxKM in the presence of genetic heterogeneity across four observable sub-populations, adjusting for covariates.	18
Table 2.4	Empirical size comparison between HWV and coxKM in testing genetic association in the presence of genetic heterogeneity across two latent sub-populations, adjusting for covariates.	20
Table 2.5	Power comparison between HWV and coxKM in the presence of genetic heterogeneity across two latent sub-populations, adjusting for baseline covariates.	21
Table 2.6	Empirical size comparison between HWV and coxKM in the presence of genetic heterogeneity across 20 latent sub-populations, adjusting for covariates.	22
Table 2.7	Empirical size comparison of HWV and coxKM considering genetic heterogeneity due to genomic profile, adjusting for covariates.	23
Table 2.8	Empirical sizes and powers of WVI and LRT in testing G-G interaction without considering left truncation.	26
Table 2.9	Empirical sizes and powers of HWV and LRT in testing G-E interaction without considering left truncation ($p = 10, q = 2, n = 1000$).	26
Table 2.10	Empirical size for WV and HWV that considers heterogeneity across individual genome profiles and adjusts for covariates, under several stringent p -value thresholds.	26
Table 2.11	Top five significant genes discovered by WV/HWV from a genome-wide association analysis of ROSMAP dataset by assuming age-to-onset of AD follows Cox PH model.	29

Table 3.1	Empirical size and power of WV-IC with $r = 0$ in testing genetic effects in the absence of left truncation.	40
Table 3.2	Empirical size and power of WV-IC with $r = 0$ in testing genetic effects under linear confounding and no left truncation.	41
Table 3.3	Comparison of performance of WV-IC and HWV-IC ($r = 0$) in testing genetic association under genetic heterogeneity across four observable subpopulations and no left truncation.	43
Table 3.4	Empirical size of HWV-IC with $r = 0$ in testing genetic effects under genetic heterogeneity across two latent subpopulations and no left truncation.	44
Table 3.5	Power of HWV-IC with $r = 0$ in testing genetic effects under genetic heterogeneity across two latent subpopulations and no left truncation. Various heterogeneity scenarios were considered, determined by the values of β_{1k} and β_{2k} , including the same effect size and the same effect direction (T1), identical sizes but opposite directions (T2), no effect in one subpopulation while positive effect in the other (T3), different sizes and opposite directions (T4), and different sizes but the same direction (T5).	44
Table 3.6	Empirical size of HWV-IC with $r = 0$ in testing genetic effects under genetic heterogeneity across twenty latent subpopulations and no left truncation.	45
Table 3.7	Power of HWV-IC with $r = 0$ under genetic heterogeneity across twenty latent subpopulations and no left truncation.	45
Table 3.8	Empirical size of HWV-IC with $r = 0$ in testing genetic effects under genetic heterogeneity across individual genome profiles and no left truncation.	46
Table 3.9	Power of HWV-IC with $r = 0$ under genetic heterogeneity across individual genome profiles and no left truncation.	47
Table 3.10	Empirical sizes and powers of WVI-IC and LRT in testing G-G interaction under no left truncation.	48
Table 3.11	Empirical sizes and powers of WVI-IC and LRT in detecting G-E interaction under no left truncation.	48
Table 3.12	Empirical sizes of WV-IC and HWV-IC under stringent p -value thresholds and no left truncation.	49

Table 3.13	Top five genes discovered by WV-IC and HWV-IC from a gene-based genome-wide association analysis of the DC-WGAGE dataset. PH and PO stand for the Cox proportional hazard model and the proportional odds model respectively. CP and IBS stand for cross-product kernel and IBS kernel respectively. Various types of heterogeneity were considered, including no genetic heterogeneity (S1), heterogeneity across races (S2), heterogeneity between the two home water fluoride levels (S3), heterogeneity between sexes (S4), and heterogeneity across genetic backgrounds (S5).	52
Table 4.1	Empirical size and power comparison of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under covariate adjustment and left truncation.	60
Table 4.2	Empirical size and power of $V_{\mathbf{Z}}$ in testing genetic effects under quadratic confounding and left truncation.	61
Table 4.3	Empirical sizes of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ in testing genetic association under genetic heterogeneity across two observable subpopulations and left truncation.	63
Table 4.4	Powers of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ in testing genetic association under genetic heterogeneity across two observable subpopulations and left truncation.	64
Table 4.5	Empirical sizes of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across two latent subpopulations, covariate adjustment and left truncation.	65
Table 4.6	Powers of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across two latent subpopulations, covariate adjustment and left truncation. Various heterogeneity scenarios were considered, determined by the values of β_{1k} and β_{2k} , including the same effect size and the same effect direction (T1), identical sizes but opposite directions (T2), no effect in one subpopulation while positive effect in the other (T3), and different sizes but the same direction (T4).	65
Table 4.7	Empirical sizes of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across twenty latent subpopulations, covariate adjustment and left truncation.	66
Table 4.8	Power comparison of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ under genetic heterogeneity across twenty latent subpopulations, covariate adjustment and left truncation.	67
Table 4.9	Empirical sizes of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ in testing genetic effects under genetic heterogeneity across individual genome profiles, covariate adjustment and left truncation.	68

Table 4.10	Powers of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ under genetic heterogeneity across individual genome profiles, covariate adjustment and left truncation.	68
Table 4.11	Empirical sizes and powers of V_I and the Wald test in testing G-G interaction under left truncation.	69
Table 4.12	Empirical sizes and powers of V_I and the Wald test in testing G-E interaction under left truncation.	70
Table 4.13	Empirical sizes of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ under stringent p -value thresholds and left truncation.	71
Table 4.14	Empirical size and power of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^c$ considering left truncation, with $n = 100, p = 15$	71
Table 4.15	Empirical size and power of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}^{H,c}$ considering genetic heterogeneity across two observable subpopulations and left truncation, with $n = 200, p = 10$	72
Table 4.16	Empirical size and power of test V_I and V_I^c in testing G-G interaction effect, considering left truncation, with $n = 600, p = q = 2$	73
Table 4.17	Empirical size and power of test V_I and V_I^c in testing G-E interaction effect, considering left truncation, with $n = 500, p = 2$	73
Table 4.18	Top five genes discovered by $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^{H,c}$ from ROSMAP dataset by assuming the age-to-onset of Alzheimer’s disease follows AFT model. IBS, CP, Lap and Quad stand for IBS, cross-product, Laplacian and Quadratic kernel, respectively, that measures genetic similarity. Various types of heterogeneity were considered, including no genetic heterogeneity (S1), heterogeneity between sexes (S2), heterogeneity across education categories (S3), and heterogeneity across genetic backgrounds (S4).	77
Table 4.19	Function of IGSF23, the third most frequent top gene, according to Entrez.	77
Table 5.1	Empirical size and power comparison of test $V_{\mathbf{Z}}$ and test $V_{\mathbf{Z}}^H$ in testing genetic effects considering adjustment covariates and left truncation.	86
Table 5.2	Empirical size and power of test $V_{\mathbf{Z}}$ in testing genetic effects under quadratic confounding and left truncation.	87
Table 5.3	Empirical sizes of test $V_{\mathbf{Z}}$ and test $V_{\mathbf{Z}}^H$ in testing genetic association under genetic heterogeneity across two observable subpopulations, considering adjustment covariates and left truncation.	88

Table 5.4	Power of test $V_{\mathbf{Z}}$ and test $V_{\mathbf{Z}}^H$ in testing genetic association under genetic heterogeneity across two observable subpopulations, considering adjustment covariates and left truncation.	89
Table 5.5	Empirical size of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across two latent subpopulations, with covariates adjustment and left truncation.	90
Table 5.6	Power of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across two latent subpopulations, considering adjustment covariates and left truncation. Various heterogeneity scenarios were considered, determined by the values of β_{1k} and β_{2k}	91
Table 5.7	Empirical size of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across twenty latent subpopulations considering adjustment covariates and left truncation.	91
Table 5.8	Power of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ under genetic heterogeneity across twenty latent subpopulations, considering adjustment covariates and left truncation.	92
Table 5.9	Empirical size of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ in testing genetic effects under genetic heterogeneity across individual genome profiles, considering adjustment covariates and left truncation.	93
Table 5.10	Power of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ under genetic heterogeneity across individual genome profiles, considering adjustment covariates and left truncation.	93
Table 5.11	Empirical sizes and powers of test V_I and Wald test in testing G-G interaction considering left truncation.	95
Table 5.12	Empirical sizes and powers of test V_I and Wald test in detecting G-E interaction considering left truncation.	95
Table 5.13	Empirical sizes of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ under stringent p -value thresholds and left truncation.	96
Table 5.14	Empirical size and power of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^c$ considering left truncation, with $n = 100, p = 15$	97
Table 5.15	Empirical size and power of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}^{H,c}$ considering genetic heterogeneity across two observable subpopulations and left truncation, with $n = 200, p = 10$	97

Table 5.16 Empirical size and power of test V_I and V_I^c in testing G-G interaction effect, considering left truncation, with $n = 300$, $p = q = 10$ 98

Table 5.17 Empirical size and power of test V_I and V_I^c in testing G-E interaction effect, considering left truncation, with $n = 200$, $p = 10$ 99

Table 5.18 Top five genes discovered by $V_{\mathbf{Z}}^c$ and $V_{\mathbf{Z}}^{H,c}$ from ROSMAP dataset by assuming age-to-onset of AD follows additive hazards model. CP, IBS, Lap and Quad are IBS, cross-product, Laplacian and Quadratic kernel, respectively, which measure genetic similarity. Various types of heterogeneity were considered, including no genetic heterogeneity (S1), heterogeneity across education categories (S2), heterogeneity between sexes (S3), and heterogeneity across genetic backgrounds (S4). 102

LIST OF FIGURES

Figure 2.1	<i>Uniform</i> (0, 1) Q-Q plots of p-values for WV, coxKM, gt and coxSKATs under $n = 150$, $p = 5$ and the null hypothesis of no genetic association.	16
Figure 2.2	<i>Uniform</i> (0, 1) Q-Q plots of p-values for WV and coxKM under linear confounding and the null hypothesis of no genetic association.	18
Figure 2.3	Power comparison between HWV and coxKM in the presence of genetic heterogeneity across twenty latent subpopulations when $p = 15$ and $n = 1000$. 24	24
Figure 2.4	Power comparison between HWV and coxKM in the presence of genetic heterogeneity across individual genome profiles when $p = 10$ and $n = 1000$	25
Figure 4.1	Q-Q plot of the null p-value of $V_{\mathbf{Z}}$ under quadratic confounding with $n = 400$ and $p = 3$ or 5.	62
Figure 4.2	Comparison of uniform QQ plot of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^c$ in detecting genetic association without considering genetic heterogeneity effect, with $n=100$ and $p=15$	72
Figure 4.3	Uniform QQ plot of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}^{H,c}$, considering genetic heterogeneity across two observable subpopulations, with $n=200$ and $p=10$	74
Figure 4.4	Comparison of uniform QQ plot of test V_I and V_I^c in testing G-G interaction effect, with $n=600$, $p=q=2$	75
Figure 4.5	Comparison of uniform QQ plot of test V_I and V_I^c in testing G-E interaction effect, with $n=500$, $p=2$, $q=1$	76
Figure 5.1	Comparison of uniform QQ plot of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^c$ in detecting genetic association without considering genetic heterogeneity effect, with $n=100$ and $p=15$	97
Figure 5.2	Comparison of uniform QQ plot of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}^{H,c}$, considering genetic heterogeneity across two observable subpopulations, with $n = 200$ and $p = 10$. 98	98
Figure 5.3	Comparison of uniform QQ plot of test V_I and V_I^c in testing G-G interaction effect, with $n=300$, $p=q=10$	100
Figure 5.4	Comparison of uniform QQ plot of test V_I and V_I^c in testing G-E interaction effect, with $n=200$, $p=10$, $q=1$	100

CHAPTER 1 Introduction

1.1 Introduction and literature review

Advances in high throughput sequencing (HTS) technologies have been used to create large datasets, which enables researchers to comprehensively investigate the insights of a wide-ranging catalog of variants (e.g., single nucleotide polymorphisms (SNPs), gene expression, and copy number variations). Also, with HTS technologies, genome-wide association studies (GWAS), and whole-genome/exome sequencing association studies can be used to detect the association between novel genetic variants and complex human diseases. A case-control study has been a popular experiment designed in genetic association studies, by comparing the frequency of SNP alleles in two groups of individuals: 1) cases who have been diagnosed with the target disease and 2) controls who are randomly selected and disease-free. Although case-control is one of the primary tools in identifying genetic variants that are associated with disease susceptibility, there is increased interest in studying the genetic association with survival outcomes such as age-to-onset of cancer diagnosis (Nagy, Munkácsy, and Gyórfy, 2021; Huang et al., 2017, 2009; Azzato et al., 2010), progressive supranuclear palsy (PSP) (Jabbari et al., 2021), and alcohol dependence (Kapoor et al., 2014). The reason behind this increasing interest is that, with the continued growth of longitudinal data in healthcare research, a cohort study is becoming more and more popular in experiment design. In cohort studies, working with survival outcomes has been shown to be more powerful in detecting the association between SNPs and the onset of a wide range of human phenotypes (van der Net et al., 2008; Staley et al., 2017; Hughey et al., 2019), because in addition to binary outcomes (i.e. disease occurrence), survival outcomes also provided information about age/time of the disease onset. Numerous studies in cancer research, for example, lung cancer (Beer et al.,

2002; Chen et al., 2007; Yu et al., 2008) and breast cancer (Shu et al., 2012; Azzato et al., 2010), have demonstrated the prediction of survival outcomes using genetic markers allow the identification of high-risk individuals at an early stage of disease, therefore additional treatment can be given after primary treatment to prevent the recurrence of the disease.

HTS technologies are able to sequence millions of DNA molecules at a time, producing 120Gb data in less than 30 hours. Despite the comprehensive insights about DNA molecules, one challenge is the high dimensionality of the sequencing dataset. Single variant tests identify important genetic variants by assessing their marginal effects individually. However, this suffers from 1) low efficiency when there exists a complex relationship between multiple genetic variants and the phenotype, 2) high computation burden due to a massive number of genetic variants, 3) low statistical power on rare variants unless sample size or effect size is large (Lee et al., 2014), 4) fewer SNPs meet the threshold after correcting for multiple testing issue due to a large number of genetic variants. To address these issues, a biologically relevant region (e.g., genes, genetic pathways) is tested instead of a single variant. Such region-based test enables the aggregation of association signals from rare variants to boost statistical testing power. Moreover, the results of region-based genetic association analyses are more interpretable and reproducible, since the effect of single variants in a biological region is integrated, therefore the test is robust for the effect of poorly annotated single variant (Subramanian et al., 2005; Beyene et al., 2009; Goeman et al., 2005).

Kernel method has been extensively used in machine learning applications (Hofmann, Schölkopf, and Smola, 2008). By applying feature mapping functions, the kernel method enables researchers to view data from a different perspective and capture complex relationships between predictive variables and response variables. Some popular applications of the kernel

method are support vector machine (SVM) classifier, kernel principal component analysis (kernel PCA) for dimension reduction, and kernel K-means for clustering. Kernel machine regression (KMR) is a more general and rigorous framework that applies the kernel method in regression analysis. In the past two decades, numerous set-based genetic association tests for survival outcomes were developed based on KMR (Cai, Tonini, and Lin, 2011; Goeman et al., 2005; Sinnott and Cai, 2013; Chen et al., 2014). These tests have been shown to detect novel genes or genetic pathways associated with age-to-onset of human diseases such as breast cancer and obesity.

However, the existing KMR-based tests for survival outcomes have several drawbacks: 1) the test is time-consuming for permutation-based globaltest: *gt* (Goeman et al., 2005), kernel machine Cox regression: *coxKM* (Cai, Tonini, and Lin, 2011), and kernel machine regression under accelerated failure time (AFT) model: *aftKM* (Sinnott and Cai, 2013), they rely on resampling procedures (1-10K permutations) to approximate the null distribution of test statistic, thereby compute p -value, which is time-consuming as shown in simulation results in Chapter 2. 2) asymptotic theory-based *globaltest* (Goeman et al., 2005), KMR for AFT model (Sinnott and Cai, 2013), and the score-based sequence kernel association test (SKAT) for Cox regression: CoxSKATs (Chen et al., 2014) are not reliable under moderate sample size ($n=150$). This is demonstrated by the uniform Quantile-Quantile (QQ) plots of p -value in simulation results in Chapter 2, it can be observed that there is an obvious deviation from the uniform distribution. 3) The major existing KMR-based association tests assume genetic homogeneous effect, where the genetic effects are identical across different individuals or subpopulations. All existing KMR-based association test are subject to power loss when the genetic variants in the target population have a heterogeneous effect, where genetic effects

vary across individuals and (un)observable subpopulations. Studies (McClellan and King, 2010; Galvan, Ioannidis, and Dragani, 2010) suggested that genetic heterogeneity is more suitable in capturing genetic effects in genetic research of complex human diseases, because through the evolutionary history of human genetics, the same genetic marker may possess numerous different rare mutations, and each rare mutations may lead to different phenotype in different individuals and environments. 4) *coxKM* (Cai, Tonini, and Lin, 2011) is not accurate when there exists observable confounder. As shown in simulation results in Chapter 2, when in the presence of adjustment covariates that are linearly correlated with genetic markers, *coxKM* becomes anti-conservative and the QQ plot deviates from $Uniform(0, 1)$. 5) the existing genetic association tests: *globaltest* (Goeman et al., 2005), *CoxSKATs* (Chen et al., 2014), *coxKM* (Cai, Tonini, and Lin, 2011), and *aftKM* (Sinnott and Cai, 2013) all studied right-censored data, but not applicable to survival times that subject to general interval-censoring, which arise in many important studies of complex human diseases: age-related macular degeneration (AMD) (Sun and Ding, 2019) and breast cancer (Kim, Williamson, and Lin, 2016). To address the drawbacks of existing KMR-based tests and to facilitate the genetic association and interaction analyses of interval-censored time-to-event outcomes, four suites of multi-marker survival tests were developed based on weighted V statistics.

1.2 Organization of this dissertation

In the rest of this dissertation, a series of novel association and interaction tests were developed based on different types of survival outcomes or survival models. They are organized as follows. In Chapter 2, we assume that survival times follow Cox proportional hazards model with possible covariate adjustment and left truncation. In Chapter 3, we assume that

survival times follow a semiparametric transformation model and are subject to interval censoring, with possible covariate adjustment and left truncation. In Chapters 4 and 5, we assume that survival times respectively follow the accelerated failure time (AFT) model and the additive hazards model subject to competing risk and left truncation. Finally, Chapter 6 describes concluding remarks and future works based on this dissertation.

CHAPTER 2 Multi-marker genetic association and interaction tests for survival outcomes based on weighted V statistics

2.1 Methods

A set of genetic association and interaction tests are developed for survival times that follow Cox proportional hazard (PH) model, with or without considering adjustment covariates and left truncation time.

2.1.1 Association tests

Consider a cohort study with n independent individuals, who are disease-free when entering the study. Let $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ denotes p -dimensional genetic marker we are interested in. $\mathbf{G}_j = \{G_{ij}\}_{i=1}^n$ is the j -th genetic marker that represents either a gene expression, a SNP (coded by 0, 1, or 2) or a genetic pathway. To mimic linkage disequilibrium (LD) structure in real world application, we assume that there is moderate correlation between genetic markers of the underlying population. Let T and C denote respectively survival time and right censoring time, then the observed survival outcome for each individual is denoted as (A, \tilde{T}, Δ) , where A is truncation time (i.e., individual's age at study entry), $\tilde{T} = \min(T, C)$, and $\Delta = I(T \leq C)$ is an indicator for event ($\Delta = 1$) or censoring ($\Delta = 0$). When adjust for covariates, a k -dim time-invariant baseline covariates $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_k)$ is considered to reduce confounding and/or increase test power. We developed two sets of association test by assuming the effect of \mathbf{G} is either homogeneous or heterogeneous across sub-population structure indicated explicitly by d -dim variable $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_d)$ or implicitly by observable/latent variables inferred by \mathbf{X} . Denote the observed data as $D = (A, \tilde{T}, \Delta, \mathbf{G}, \mathbf{Z}, \mathbf{X})$, where \mathbf{X} exists only when considering genetic heterogeneity. We assume non-informative censoring and left truncation given \mathbf{G} and \mathbf{Z} (and \mathbf{X} when considering

genetic heterogeneity).

In the absence of genetic heterogeneity, without covariate adjustment

Under the null hypothesis that \mathbf{G} is not associated with the time-to-onset of the interesting event when do not adjust for \mathbf{Z} , we assume that the survival time follows a Cox proportional hazard (PH) model with hazard function, $\lambda(t) = \lambda_0(t)$, where $\lambda_0(t)$ is unspecified baseline hazard function. The cumulative hazard function $\Lambda(t) = \int_0^t \lambda(s)ds$ will be estimated non-parametrically by using the Nelson-Aalen estimator. The proposed association test is based on the following weighted V statistic,

$$V = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \tilde{f}(\mathbf{G}_i, \mathbf{G}_j) M_i M_j, \quad (2.1)$$

where $M_i = N_i(\infty) - \int_0^\infty Y_i(s)\lambda(s)ds$ is martingale residual of subject i , where $N_i(t) = I(U_i \leq t, \Delta_i = 1)$ and $Y_i(t) = I(U_i \geq t)$ are corresponding counting process and at-risk process. The test statistic includes two parts 1) V statistic kernel $M_i M_j$ and 2) weight function $\tilde{f}(\mathbf{G}_i, \mathbf{G}_j)$, which is a centered genetic similarity between subject i and j ,

$$\tilde{f}(\mathbf{G}_i, \mathbf{G}_j) = f(\mathbf{G}_i, \mathbf{G}_j) - E[f(\mathbf{G}_i, \mathbf{G}_j)|\mathbf{G}_i] - E[f(\mathbf{G}_i, \mathbf{G}_j)|\mathbf{G}_j] + E[f(\mathbf{G}_i, \mathbf{G}_j)], \quad (2.2)$$

where $f(\mathbf{G}_i, \mathbf{G}_j)$ is a kernel function (e.g., IBS, Gaussian kernel). The test statistic is constructed such that when \mathbf{G} is significantly associated with the event of interest, we expect a large phenotype similarity, $M_i M_j$, weighted by a large genetic similarity, $\tilde{f}(\mathbf{G}_i, \mathbf{G}_j)$, leading to a large V^* value, therefore the test will have small p -value.

Now we want to derive the asymptotic null distribution of V . By Mercer's Theorem (Cris-

tianini and Shawe-Taylor, 2000), under regularity conditions, we can decompose $\tilde{f}(\mathbf{G}_i, \mathbf{G}_j)$ as $\tilde{f}(\mathbf{G}_i, \mathbf{G}_j) = \sum_{l=1}^{\infty} \lambda_l \psi(\mathbf{G}_i) \psi(\mathbf{G}_j)$, where $\{\lambda_l\}_{l=1}^{\infty}$ and $\{\psi_l(\cdot)\}_{l=1}^{\infty}$ are eigenvalues and eigenfunctions of $\tilde{f}(\cdot, \cdot)$. We then have

$$nV = \sum_{t=1}^{\infty} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \sqrt{\lambda_t} \psi_t(\mathbf{G}_i) M_i \right)^2. \quad (2.3)$$

Theorem 1. *Suppose that $\eta \equiv E(M_i^2)$, where $M = N(\infty) - \int_0^{\infty} Y(s) d\Lambda(s)$, is a finite-positive number and $E[\tilde{f}(\mathbf{G}_i, \mathbf{G}_j)] < \infty$. Under the null hypothesis that T is independent of \mathbf{G} , $nV \rightsquigarrow \eta \sum_{t=1}^{\infty} \lambda_t \chi_{1t}^2$, where χ_{1t}^2 's are independent χ^2 random variables with degree of freedom 1.*

Following Theorem 1, we can show that the asymptotic null distribution of nV is a weighted sum of independent degree one χ^2 random variables.

In real application, we estimate $\Lambda(t)$ by Nelson-Aalon estimator, $\hat{\Lambda}(t) = \int_0^t dN(s)/Y(s)$, where $N(t) = \sum_{i=1}^n N_i(t)$, $Y(t) = \sum_{i=1}^n Y_i(t)$. $\tilde{f}(\mathbf{G}_i, \mathbf{G}_j)$ is estimated from observed data by $(I - J)F(I - J)$, where $F = \{f(\mathbf{G}_j, \mathbf{G}_j)\}_{n \times n}$, I is a $n \times n$ diagonal matrix, and J is a $n \times n$ matrix with all elements equal 1. The test statistic can be written as

$$V = \hat{\mathbf{M}}(I - J)^T F(I - J)\hat{\mathbf{M}}, \quad (2.4)$$

where $\hat{\mathbf{M}} = (\hat{M}_1, \dots, \hat{M}_n)^T$, with $\hat{M}_i = N_i(\infty) - \int_0^{\infty} Y_i(s) d\hat{\Lambda}(s)$. We approximate $\lambda = E(M_i^2)$ by using $\hat{\lambda} = \hat{\mathbf{M}}^T \hat{\mathbf{M}}/n$ and η_t by using $\hat{\eta}_t = \tilde{\eta}_t/n$, where $\{\tilde{\eta}_t\}_{t=1}^n$ are eigenvalues of

$(I - J)^T F (I - J)$. The asymptotic null distribution of V can be approximated by

$$nV \sim \hat{\lambda} \sum_{t=1}^n \hat{\eta}_t \chi_{1t}^2, \quad (2.5)$$

where $\{\chi_{1t}^2\}_{t=1}^n$ are χ^2 random variables with 1 degree of freedom. Based on this linear combination of independent χ^2 random variables, the p -value can be calculated using Davies' method (Davies, 1980), $P(nV \geq nV_{obs})$, where V_{obs} is observed value of V . This is implemented using R function "*davies*" in R package "*CompQuadForm*".

In the absence of genetic heterogeneity, with adjustment covariates

The association test considering adjustment covariates, \mathbf{Z} , is similar to the test without covariate adjustment. We now assume that the survival time for subject i conditional on \mathbf{Z} follow a Cox PH model with below hazard function

$$\lambda(t|\mathbf{Z}_i) = \lambda_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{Z}_i\}$$

and covariate-centered genetic similarity as follow

$$\begin{aligned} & \tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j) \\ = & f(\mathbf{G}_i, \mathbf{G}_j) - E[f(\mathbf{G}_i, \mathbf{G}_k)u(\mathbf{Z}_k, \mathbf{Z}_j)|\mathbf{G}_i, \mathbf{Z}_j] - E[f(\mathbf{G}_j, \mathbf{G}_k)u(\mathbf{Z}_k, \mathbf{Z}_i)|\mathbf{G}_j, \mathbf{Z}_i] \\ & + E[u(\mathbf{Z}_i, \mathbf{Z}_m)f(\mathbf{G}_m, \mathbf{G}_k)u(\mathbf{Z}_k, \mathbf{Z}_j)|\mathbf{Z}_i, \mathbf{Z}_j], \end{aligned}$$

where $u(\mathbf{Z}_i, \mathbf{Z}_j) = (1, \mathbf{Z}_i^T)^T [E\{(1, \mathbf{Z}^T)^T (1, \mathbf{Z}^T)\}]^{-1} (1, \mathbf{Z}_j^T)^T$. The covariate-adjusted weighted V statistic can be written as

$$V_{\mathbf{Z}} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j) M_{\mathbf{Z},i} M_{\mathbf{Z},j}, \quad (2.6)$$

where $M_{\mathbf{Z},i} M_{\mathbf{Z},j}$ is considered ij -th V statistic kernel that measures phenotype similarity, $\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j)$ is covariate-centered kernel matrix that measures genetic similarity. The larger (smaller) phenotype similarity weighted by larger (smaller) genetic similarity, leads to larger (smaller) V statistic value and smaller (larger) p -value.

Theorem 2. *Suppose that $\xi \equiv E(M_{\mathbf{Z},i}^2)$ is a finite positive number and $E[\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j)] < \infty$. When T is independent of \mathbf{G} given \mathbf{Z} and \mathbf{G} is independent of \mathbf{Z} , $nV_{\mathbf{Z}} \rightsquigarrow \xi \sum_{t=1}^{\infty} \nu_t \chi_{1t}^2$, where $\{\chi_{1t}^2\}_{t=1}^{\infty}$ are independent χ^2 random variables with 1 degree of freedom and $\{\nu_t\}_{t=1}^{\infty}$ are eigenvalues of $\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j) = \sum_{t=1}^{\infty} \nu_t \phi(\mathbf{G}_i, \mathbf{Z}_i) \phi(\mathbf{G}_j, \mathbf{Z}_j)$ with $E(\phi_s(\mathbf{G}, \mathbf{Z}) \phi_t(\mathbf{G}, \mathbf{Z})) = I(s = t)$. The same asymptotic null distribution holds when $f(\mathbf{G}_i, \mathbf{G}_k)$ is cross-product kernel and \mathbf{G} is linearly related to \mathbf{Z} through $\mathbf{G} = \mathbf{a} + \mathbf{B}^T \mathbf{Z} + \mathbf{e}$, where \mathbf{a} and \mathbf{B} are constant intercept and regression coefficient matrix and \mathbf{e} is zero mean random error vector that is independent of \mathbf{Z} .*

When \mathbf{G} is independent of \mathbf{Z} , both test statistics, V and $V_{\mathbf{Z}}$, are asymptotically valid, with $V_{\mathbf{Z}}$ more powerful. When \mathbf{Z} is correlated with both \mathbf{G} and survival outcome, then \mathbf{Z} is an observable confounder. Under the linear confounding effect, Theorem 2 implies that the covariate-adjusted weighted V test with cross-product kernel genetic similarity is asymptotically correct.

In real application, we can obtain maximum partial-likelihood estimate for β and Breslow

estimator for cumulative baseline hazard function $\Lambda_0(t)$, denoted as $\hat{\beta}$ and $\hat{\Lambda}_0(t)$, by fitting a CoxPH model (Andersen and Gill, 1982) to the observed data D . With $\hat{\beta}$ and $\hat{\Lambda}_0(t)$, we have estimated martingale residual for subject i as $\hat{M}_{\mathbf{Z},i} = N_i(\infty) - \int_0^\infty Y_i(s) \exp(\hat{\beta}^T \mathbf{Z}) d\hat{\Lambda}_0(s)$. The covariate-centered genetic similarity kernel $\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j)$ is approximated using observed sample by $(I - H)^T F (I - H)$. The covariate adjusted weighted V statistic can be written as

$$V_{\mathbf{Z}} = \hat{\mathbf{M}}_{\mathbf{Z}}^T (\mathbf{I} - \mathbf{H})^T \mathbf{F} (\mathbf{I} - \mathbf{H}) \hat{\mathbf{M}}_{\mathbf{Z}}, \quad (2.7)$$

where $\hat{\mathbf{M}}_{\mathbf{Z}} = (\hat{M}_{\mathbf{Z},i}, \dots, \hat{M}_{\mathbf{Z},n})^T$, \mathbf{I} is $n \times n$ diagonal matrix, $\mathbf{H} = \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}^T \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{Z}}^T$ with $\tilde{\mathbf{Z}} = (\mathbf{1}, \mathbf{Z})$ is a $n \times (k + 1)$ matrix with elements in the first column equal 1. The asymptotic null distribution of $nV_{\mathbf{Z}}$ can be approximated by

$$nV_{\mathbf{Z}} \sim \hat{\xi} \sum_{t=1}^n \hat{\nu}_t \chi_{1t}^2, \quad (2.8)$$

where $\hat{\xi} = \hat{\mathbf{M}}_{\mathbf{Z}}^T \hat{\mathbf{M}}_{\mathbf{Z}} / n$ is an approximation of $\xi = E(M_{\mathbf{Z},i}^2)$, $\hat{\nu}_t = \tilde{\nu}_t / n$ with $\{\tilde{\nu}_t\}'s$ are eigenvalues of $(\mathbf{I} - \mathbf{H})^T \mathbf{F} (\mathbf{I} - \mathbf{H})$. When dimension of \mathbf{Z} , denoted as k ($k < n$), is large relative to n , formula 2.8 is replace with $\frac{n}{n-k-1} \hat{\xi} \sum_{t=1}^n \hat{\nu}_t \chi_{1t}^2$ to take into account of the projection operator $(\mathbf{I} - \mathbf{H})$ as mentioned in Wei and Lu (2017). Based on this distribution, linear combination of independent χ^2 random variable with 1 degree of freedom, p -value can be achieved using Davies' method (Davies, 1980), $P(nV_{\mathbf{Z}} \geq nV_{\mathbf{Z},obs})$, where $V_{\mathbf{Z},obs}$ is the observed value of $V_{\mathbf{Z}}$.

In the presence of genetic heterogeneity

In the association tests introduced above, we assume the effect of \mathbf{G} is homogeneous. However, genetic heterogeneity is more appropriate in genetic research of complex human disease.

In this section, we extend the proposed association tests, V and $V_{\mathbf{Z}}$, to take into consideration genetic heterogeneous effect, by replacing F matrix in hazard function 2.4 and 2.7 with $\mathbf{W} = (\mathbf{J} + \mathbf{K}) \odot \mathbf{F}$, where \odot is element-wise matrix product and $\mathbf{K} = \{k(\mathbf{X}_i, \mathbf{X}_j)\}_{n \times n}$ measures similarity of observable/latent subpopulation structure imposed by variable \mathbf{X} . The resulting weighted V statistic with or without covariates adjustment are

$$\begin{cases} V^H = \hat{\mathbf{M}}^T(\mathbf{I} - \mathbf{J})\mathbf{W}(\mathbf{I} - \mathbf{J})\hat{\mathbf{M}}, \\ V_{\mathbf{Z}}^H = \hat{\mathbf{M}}_{\mathbf{Z}}^T(\mathbf{I} - \mathbf{J})\mathbf{W}(\mathbf{I} - \mathbf{J})\hat{\mathbf{M}}_{\mathbf{Z}}. \end{cases}$$

The asymptotic null distribution of V^H and $V_{\mathbf{Z}}^H$ can be derived similarly as other association tests introduced in previous sections, by replacing \mathbf{F} with \mathbf{W} . Using Davies' method (Davies, 1980), the p -values can be computed, respectively, by $P(nV^H \geq nV_{obs}^H)$ and $P(nV_{\mathbf{Z}}^H \geq nV_{\mathbf{Z},obs}^H)$, where V_{obs}^H and $V_{\mathbf{Z},obs}^H$ are observed values of V^H and $V_{\mathbf{Z}}^H$.

2.1.2 G-G/G-E interaction test

Let $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ be the gene of interest. In testing G - G interaction effect, we assume the effect of \mathbf{G} depends on another gene $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_q)$. To test the G - G interaction effect, we assume that the survival time follow a Cox PH model with hazard function $\lambda(t|\mathbf{G}, \mathbf{H}) = \lambda_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{G} + \boldsymbol{\alpha}^T \mathbf{H}\}$. The maximum partial-likelihood estimates of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is denoted as $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$, $\Lambda_0(t)$ is estimated by using Breslow estimator, denoted as $\hat{\Lambda}_0(t)$. We then have $\hat{M}_i = N_i(\infty) - \int_0^\infty Y_i(s) \exp(\sum_{k=1}^p G_{ik}\hat{\beta}_k + \sum_{l=1}^q H_{il}\hat{\alpha}_l) d\hat{\Lambda}_0(s)$. The G - G interaction test is based on the following weighted V test statistic

$$V_I = \hat{\mathbf{M}}^T(\mathbf{I} - \mathbf{O})(\mathbf{F} \odot \mathbf{K})(\mathbf{I} - \mathbf{O})\hat{\mathbf{M}},$$

where $\mathbf{O} = \mathbf{Q}(\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T$ with $\mathbf{Q} = (\mathbf{1}, \mathbf{G}, \mathbf{H})$ is a $n \times (1 + p + q)$ matrix with all element in first column equal 1. $\mathbf{F} \odot \mathbf{K}$ is element-wise matrix product of \mathbf{F} and \mathbf{K} with $\mathbf{F} = \{f(\mathbf{G}_i, \mathbf{G}_j)\}_{n \times n}$ and $\mathbf{K} = \{k(\mathbf{H}_i, \mathbf{H}_j)\}_{n \times n}$. The asymptotic null distribution of nV_I can be approximated, similarly as $nV_{\mathbf{Z}}$, by linear combination of independent χ^2 distribution. When $p + q$ is large relative to n , we multiply the approximated asymptotic null distribution by $n/(n - p - q + 1)$ to take into account the project operator $(\mathbf{I} - \mathbf{O})$. The corresponding p -value is obtained using Davies' method (Davies, 1980), $P(nV_I \geq nV_{I,obs})$. In testing G - E interaction effect, the test can be conducted similarly as G - G interaction, except that \mathbf{H} in replaced with an environmental variable.

2.2 Simulations

We performed Monte Carlo simulation to assess the finite-sample performance of the weighted V test considering adjustment covariates, in absence of left-truncation time. In all simulation settings, unless otherwise specified, three different sample sizes, $n = \{500, 1000, 1500\}$, and three different SNP-set sizes, $p = \{5, 10, 15\}$ were considered. The genetic covariates we considered are SNP-set with each SNP generated from a $Binomial(2, 0.2)$. When considering confounding effects, SNP-set will be replaced by gene expression values. Two adjustment covariates were considered for each subject, one binary $Z_1 \sim Bernoulli(0.5)$ and one continuous $Z_2 \sim Uniform(0, 2)$. Survival times were assumed to follow the Cox PH model, which was generated from an exponential distribution with rate λ . The λ value varies in different scenarios to control the right censoring rate at moderate level (between 20% and 40%). The assessment of the empirical size and statistical power were based on, respectively, 2000 and 1000 Monte Carlo samples, and the significance level of a test was set at 0.05, unless otherwise specified in the scenario when testing genetic association under various stringent

p -value thresholds.

2.2.1 Testing genetic association in the absence of genetic heterogeneity

In this series of simulations, empirical size and power of test $V_{\mathbf{Z}}$ was assessed in detecting the association between a SNP-set and onset of target event. To demonstrate the advantages of weighted V test, the performance of $V_{\mathbf{Z}}$ was compared with three existing tests: coxKM (R package 'coxKM'), asymptotic version gt (R package 'globaltest'), and coxSKAT_LRT. Survival times for subject i ($i = 1, \dots, n$) was generated from exponential distribution with following rate

$$\lambda(t|\mathbf{G}_i, \mathbf{Z}_i) = \lambda_0(t) \exp\left\{\sum_{j=1}^p \beta_j G_{ij} + \sum_{k=1}^2 0.5 Z_{ik}\right\}, \quad (2.9)$$

where $\lambda_0(t)$ and $\{\beta_j\}_{j=1}^p$ vary between scenarios.

Computation time comparison between Weighted V statistic (WV), coxKM, gt (asymptotic version), and coxSKATs

We compared computation times of WV, coxKM, gt, and coxSKATs under the null hypothesis of no genetic association. Survival time for subject i was assumed to follow Cox PH model with hazard function 2.9 where $\lambda_0(t) = 1$ and $\{\beta_j\}'s = 0$. The cross-product kernel was used to measure genetic similarity. Three sample sizes $n = 150, 500, 1000$ and one SNP-set size $p = 5$ were considered. Table 2.1 shows that WV and asymptotic version gt are comparably fast, while coxKM is the slowest due to the resampling procedure to achieve the test statistic large sample null distribution and p -value.

Table 2.1 Computation time (seconds) of WV, coxKM, gt and coxSKATs under the null hypothesis of no genetic association and $p = 5$.

Sample Size	WV	coxKM	gt	coxSKATs
150	0.07	0.30	0.03	0.09
500	0.24	1.78	0.23	0.36
1000	1.41	8.22	1.76	2.13

Comparison of empirical size of WV, coxKM, gt (asymptotic version), and coxSKATs under moderate sample size

We investigated the empirical sizes of WV, coxKM, gt, coxSKATs under moderate sample size $n = 150$ and SNP-set size $p = 5$. Other simulation settings were the same as the simulation above. Figure 2.1 indicates that both WV and coxKM tests are asymptotically correct while gt and coxSKATs are not because the QQ plot of their p -values deviates from $Uniform(0, 1)$ distribution. Therefore, only WV and coxKM were investigated for the following simulations.

Empirical size and power under various n 's and p 's

In this simulation, the performance of WV and coxKM tests were investigated under various n and p . Survival time for subject i was assumed to follow Cox PH model with hazard function 2.9 where $\lambda_0(t) = 0.5$, $\{\beta_j\}$'s were set to 0 or 0.1 for empirical size and power assessment. IBS kernel is used to measure genetic similarity. Table 2.2 indicates that Type I error of both WV and coxKM are close to the nominal level and have comparable statistical power under various n 's and p 's.

Empirical size and power under linear confounding effect

In this simulation, the genetic covariates of our interest are gene expression values $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$, where $\mathbf{G}_i = Z_{i1} + Z_{i2} + \mathbf{e}_i$. $\mathbf{e}_i \sim MVN(\mathbf{0}, \Sigma)$ is p -dim multivariate normal

Table 2.2 Empirical size and power comparison between WV and coxKM in testing association of a SNP set with right-censored survival time in the absence of genetic heterogeneity, adjusting for covariates.

		Empirical Size (Power)		
		p=5, n=500	p=5, n=1000	p=5, n=1500
WV		0.054 (0.208)	0.049 (0.445)	0.055 (0.601)
coxKM		0.052 (0.244)	0.054 (0.430)	0.045 (0.627)
		p=10, n=500	p=10, n=1000	p=10, n=1500
WV		0.038 (0.335)	0.052 (0.631)	0.055 (0.819)
coxKM		0.041 (0.337)	0.051 (0.630)	0.040 (0.846)
		p=15, n=500	p=15, n=1000	p=15, n=1500
WV		0.046 (0.507)	0.049 (0.853)	0.046 (0.974)
coxKM		0.038 (0.473)	0.034 (0.886)	0.045 (0.976)

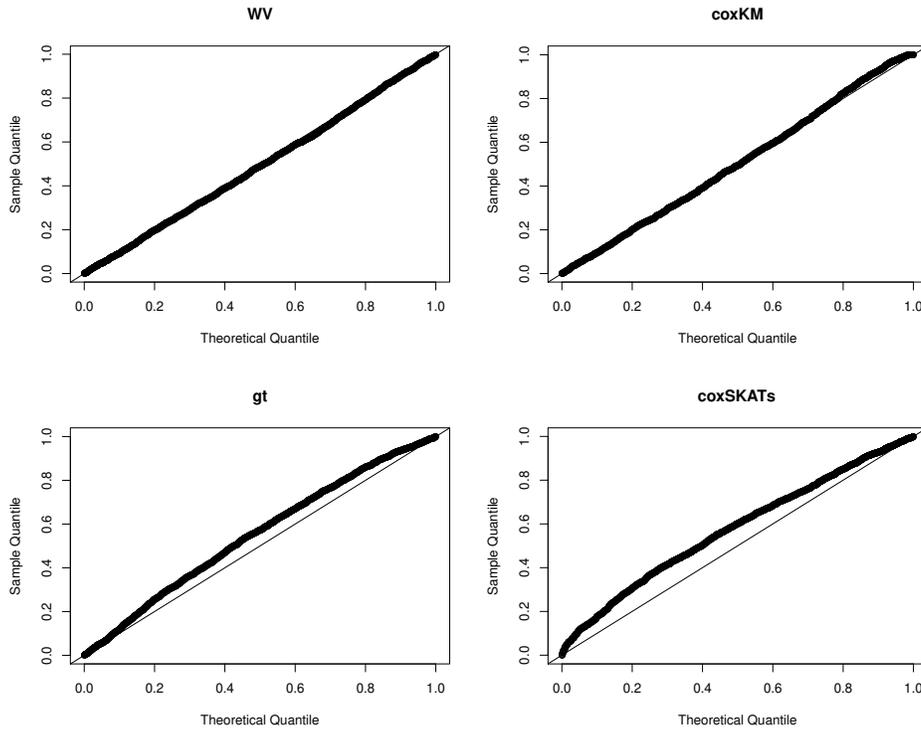


Figure 2.1 $Uniform(0, 1)$ Q-Q plots of p-values for WV, coxKM, gt and coxSKATs under $n = 150$, $p = 5$ and the null hypothesis of no genetic association.

random error with covariance matrix $\Sigma = \{0.5^{|i-j|}\}_{p \times p}$. Cross-product kernel was used to measure gene expression similarity for both WV and coxKM tests. Survival times were generated similarly as previous simulation, except \mathbf{G} is replaced by the gene expression

values. Sample size $n = 500$ and SNP-set size $p = 5$ is considered. Type 1 error of HWV is 0.0475, which is close to nominal level, while Type I error of coxKM (0.0185) is far from nominal level. This can also be observed from the *Uniform*(0, 1) Q-Q plot in Figure 2.2.

2.2.2 Testing genetic association in the presence of genetic heterogeneity

In this series of simulations, we investigated the performance of HWV in testing genetic association with age-to-onset of target event, considering genetic heterogeneity across four different sources: 1) heterogeneity across four observable subpopulations, 2) heterogeneity across two latent subpopulations, 3) heterogeneity across twenty latent subpopulations, and 4) heterogeneity across genome profile.

Genetic heterogeneity across four observable subpopulations

In this simulation, we investigated the empirical size and power of HWV and coxKM in the presence of genetic heterogeneity across four observable subpopulations inferred by $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$, a three-dimensional dummy variable that codes four equally distributed observable subpopulations. Survival time for subject i was generated from Cox PH model with the following hazard function

$$\lambda(t|\mathbf{G}_i, \mathbf{X}_i) = \exp \left\{ \sum_{k=1}^p (\beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4) G_{ik} + 0.6 X_{i1} + 0.6 X_{i2} + 0.2 X_{i3} \right\}.$$

We set $(\beta_1, \beta_2, \beta_3, \beta_4)$ to be $(0, 0, 0, 0)$ and $(0.5, 0.005, 0.001, 0.0005)$ for, empirical size assessment and power assessment, respectively. IBS kernel was used to measure genetic similarity and identity kernel, $\{I(\mathbf{X}_i = \mathbf{X}_j)\}_{n \times n}$, was used to measure subpopulation similarity in HWV. Two sample sizes $n = \{500, 750\}$ and three SNP-set sizes $p = \{5, 10, 15\}$ were

considered. Table 2.3 shows that both HWV and coxKM control Type I error well around nominal level under various sample sizes and SNP-set sizes. However, HWV is more powerful than coxKM in presence of genetic heterogeneity across four observable subpopulations, by taking advantage of the background similarity when conducting the test.

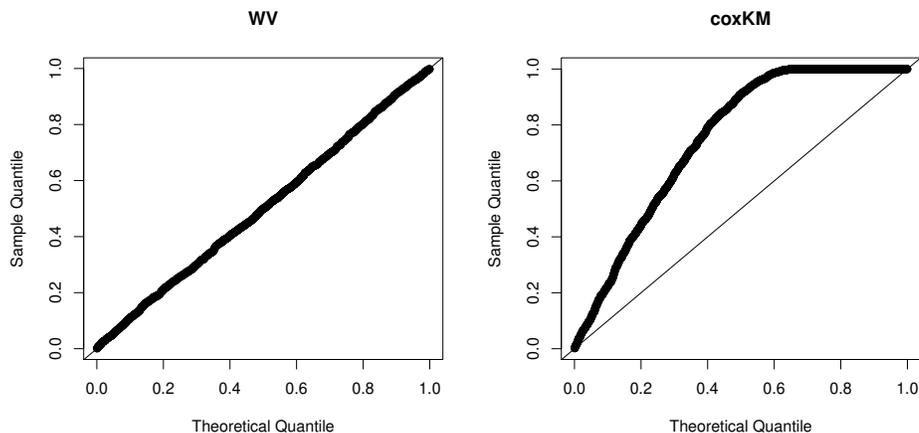


Figure 2.2 $Uniform(0, 1)$ Q-Q plots of p-values for WV and coxKM under linear confounding and the null hypothesis of no genetic association.

Table 2.3 Empirical size and power comparison between HWV and coxKM in the presence of genetic heterogeneity across four observable sub-populations, adjusting for covariates.

	Empirical Size (Power)	
	p=5, n=500	p=5, n=750
HWV	0.054 (0.801)	0.049 (0.963)
coxKM	0.047 (0.618)	0.047 (0.811)
	p=10, n=500	p=10, n=750
HWV	0.052 (0.942)	0.046 (0.997)
coxKM	0.045 (0.799)	0.055 (0.975)
	p=15, n=500	p=15, n=750
HWV	0.043 (0.931)	0.050 (0.998)
coxKM	0.044 (0.857)	0.049 (0.984)

Genetic heterogeneity across two latent subpopulations

In this simulation, we investigated the empirical size and power of HWV and coxKM in the presence of genetic heterogeneity across two latent subpopulations. The survival time for

subject i in subpopulation j was generated from Cox PH model with the following hazard function

$$\lambda(t|\mathbf{G}_{ij}, \mathbf{Z}_{ij}) = 0.5 \exp\left\{\sum_{k=1}^p G_{ijk}\beta_{jk} + 0.5Z_{ij1} + 0.5Z_{ij2}\right\}, \quad (2.10)$$

where β_{jk} represents the effect of \mathbf{G}_k ($k = 1, \dots, p$) in subpopulation j ($j = 1, 2$). One-dimensional covariate $\mathbf{X} = \{X_i\}_{i=1}^n$ was generated to infer two latent subpopulations. Each element is $X_i = a_i + 1 + e_i$, where $a_i \sim \text{Bernoulli}(0.5)$ and $e_i \sim \text{Normal}(0, 0.25)$. IBS kernel was used to measure genetic similarity and unified kernel to measure latent subpopulation similarity. Table 2.4 shows that Type I error of HWV is close to nominal level under various n 's and p 's, while coxKM could be conservative when p is large relative to n (e.g., under $p=15, n=500$). Table 2.5 shows power comparison of HWV and coxKM under four different scenarios: T1, T2, T3, and T4. It can be observed that in the absence of genetic heterogeneity (scenario T1 $\beta_{1k} = \beta_{2k} \neq 0$), HWV is less powerful than coxKM since the genetic effect in the underlying population is homogeneous. However, in the presence of genetic heterogeneity (scenario T2, T3, T4), HWV is more powerful than coxKM and as heterogeneity size ($|\beta_{1k} - \beta_{2k}|$) increases, HWV gains more power.

Genetic heterogeneity across twenty latent subpopulations

In this simulation, we investigated empirical sizes and power of HWV and coxKM in the presence of genetic heterogeneity across twenty latent subpopulations. The simulation setting is same as above except the number of subpopulations increased to twenty. The regression coefficients, $\{\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp}), \text{ s.t. } \beta_{j1} = \dots, \beta_{jp}\}_{j=1}^{20}$, in hazard function 2.10 were set to 0 for empirical size assessment. In power assessment, $\{\boldsymbol{\beta}_j\}_{j=1}^{20}$ were generated from uniform distribution with mean μ_β and standard deviation σ_β . To infer the twenty subpopulation

Table 2.4 Empirical size comparison between HWV and coxKM in testing genetic association in the presence of genetic heterogeneity across two latent sub-populations, adjusting for covariates.

	Empirical Size		
	p=5, n=500	p=5, n=1000	p=5, n=1500
HWV	0.050	0.051	0.056
coxKM	0.048	0.061	0.051
	p=10, n=500	p=10, n=1000	p=10, n=1500
	HWV	0.050	0.054
coxKM	0.034	0.049	0.046
	p=15, n=500	p=15, n=1000	p=15, n=1500
	HWV	0.051	0.047
coxKM	0.038	0.039	0.042

structure, we used $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{25})$ to infer subpopulation structure. $\mathbf{X}_k = \mathbf{a}_k + \mathbf{e}_k$ ($k = 1, \dots, 25$), with \mathbf{a}_k is a length n bootstrap sample from $\{1, \dots, 20\}$ and $\mathbf{e}_k = (e_{k1}, \dots, e_{kn}) \sim Normal(0, 0.5)$ is random error vector. IBS kernel was used to measure genetic similarity in HWV and coxKM. Unified kernel was used to measure subpopulation similarity in HWV. Table 2.6 shows both HWV and coxKM control Type I error well around nominal level, with an minor exception that under $p = 15, n = 500$, coxKM is slightly conservative due to relatively small sample size issue. Figure 2.3 shows power comparison under $p = 15, n = 1000$. It can be observed that as average genetic effect size increases ($\mu_\beta = 0, 0.02, 0.03$), both HWV and coxKM gains power. However, for a fixed average genetic effect (μ_β), as genetic heterogeneity effect increases ($\sigma_\beta = 0.02, 0.04, 0.06$), power of coxKM does not change much, while HWV gains substantially more power by considering subpopulation structure in performing the association test.

Table 2.5 Power comparison between HWV and coxKM in the presence of genetic heterogeneity across two latent sub-populations, adjusting for baseline covariates.

		Heterogeneity Scenario†							
		T1		T2		T3		T4	
	β_{1k}	+0.09	+0.13	+0.02	+0.04	0	0	+0.01	-0.01
	β_{2k}	+0.09	+0.13	-0.02	-0.04	+0.02	+0.04	+0.05	+0.03
p=5, n=500	HWV	0.11	0.25	0.10	0.27	0.06	0.11	0.09	0.10
	coxKM	0.28	0.60	0.04	0.06	0.06	0.07	0.08	0.04
p=5, n=1000	HWV	0.24	0.59	0.18	0.48	0.08	0.17	0.18	0.15
	coxKM	0.57	0.91	0.04	0.04	0.05	0.07	0.09	0.04
p=5, n=1500	HWV	0.42	0.86	0.20	0.64	0.09	0.21	0.22	0.21
	coxKM	0.84	0.99	0.05	0.04	0.06	0.08	0.11	0.05
p=10, n=500	HWV	0.10	0.25	0.27	0.77	0.10	0.30	0.29	0.27
	coxKM	0.46	0.84	0.04	0.04	0.04	0.06	0.07	0.05
p=10, n=1000	HWV	0.25	0.69	0.48	0.98	0.16	0.51	0.55	0.51
	coxKM	0.83	1.00	0.05	0.04	0.05	0.07	0.10	0.04
p=10, n=1500	HWV	0.44	0.94	0.69	1.00	0.21	0.67	0.71	0.70
	coxKM	0.96	1.00	0.05	0.05	0.07	0.08	0.10	0.04
p=15, n=500	HWV	0.09	0.22	0.54	0.98	0.19	0.58	0.58	0.56
	coxKM	0.56	0.93	0.05	0.04	0.04	0.05	0.06	0.05
p=15, n=1000	HWV	0.22	0.72	0.86	1.00	0.34	0.87	0.87	0.86
	coxKM	0.93	1.00	0.03	0.05	0.05	0.08	0.12	0.05
p=15, n=1500	HWV	0.43	0.98	0.94	1.00	0.46	0.95	0.96	0.95
	coxKM	0.99	1.00	0.05	0.06	0.05	0.09	0.16	0.06

† Various heterogeneity scenarios were investigated in this simulation, determined by values of β_{1k} and β_{2k} , including same effect size with same direction (T1), same effect size with opposite directions (T2), no effect in one sub-population while positive effect in the other (T3), and different effect sizes with same or opposite directions (T4).

Genetic heterogeneity across individual genome profile

In this simulation, instead of assuming a 'categorical' subpopulation structure like previous two simulations did, we now assume that the subpopulation structure is 'continuous'. In other words, each individual is itself a subpopulation and genetic variants have varied effect across individuals. We used a genome profile of 1,000 SNPs, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{1000})$, to infer subpopulation structure. \mathbf{X} was generated in three steps, 1) generate genetic effect sizes $\{\beta_i = (\beta_{i1}, \dots, \beta_{ip}), s.t. \beta_{i1} = \dots = \beta_{ip}\}_{i=1}^n$ from uniform distribution with mean μ_β and variance σ_β^2 , 2) generate a n -dim vector \mathbf{X}_d ($d = 1, \dots, 1000$) from $MVN(\mathbf{0}, \Sigma)$, where

Table 2.6 Empirical size comparison between HWV and coxKM in the presence of genetic heterogeneity across 20 latent sub-populations, adjusting for covariates.

	Empirical Size		
	p=5, n=500	p=5, n=1000	p=5, n=1500
HWV	0.052	0.055	0.051
coxKM	0.042	0.053	0.045
	p=10, n=500	p=10, n=1000	p=10, n=1500
	HWV	0.048	0.056
coxKM	0.046	0.043	0.050
	p=15, n=500	p=15, n=1000	p=15, n=1500
	HWV	0.047	0.040
coxKM	0.038	0.044	0.051

Σ is a $n \times n$ identity matrix under null hypothesis and $\{\exp(-|\beta_{i1} - \beta_{j1}|/\sigma_\beta)\}_{n \times n}$ under alternative hypothesis, 3) categorize \mathbf{X}_d ($d = 1, \dots, 1000$) into SNPs coded by 0, 1, or 2 by using predetermined quantile cutoffs: a^2 , $2a(1 - a)$, and $(1 - a)^2$ of a standard normal distribution where a is MAF generated from $Beta(2, 5)$, to ensure the resulting population is in HWE. Survival times were generated from Cox PH model with hazard function 2.10, with $\{\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp}), \text{ s.t. } \beta_{j1} = \dots = \beta_{jp}\}_{j=1}^n$ set to 0 for empirical size assessment and sampled from uniform distribution with mean μ_β and variance σ_β^2 for power assessment. IBS kernel was used to measure genetic and genome profile similarity.

Results in Table 2.7 indicate that Type I error of both HWV and coxKM are close to nominal level under various sample sizes (n) and SNP-set sizes (p). Figure 2.4 shows power comparison of HWV and coxKM under $p = 10, n = 1000$. Figure 2.4 indicates that as average genetic effect (μ_β) increases, both tests gain power. However, for a fixed μ_β , as genetic heterogeneity size (σ_β) increases, HWV gains substantially more power than coxKM does.

Table 2.7 Empirical size comparison of HWV and coxKM considering genetic heterogeneity due to genomic profile, adjusting for covariates.

	Empirical Size		
	p=5, n=500	p=5, n=1000	p=5, n=1500
HWV	0.041	0.057	0.049
coxKM	0.042	0.046	0.057
	p=10, n=500	p=10, n=1000	p=10, n=1500
HWV	0.044	0.054	0.057
coxKM	0.045	0.044	0.057
	p=15, n=500	p=15, n=1000	p=15, n=1500
HWV	0.042	0.047	0.050
coxKM	0.039	0.040	0.045

2.2.3 Testing G-G/G-E interaction effect

In this simulation, we investigated the performance of the weighted V test for G - G and G - E interaction, denoted as WVI, and compared it with the likelihood ratio test (LRT).

For testing G - G interaction effect, two genes, $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ and $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_q)$, were generated using a two-step procedure: 1) generate n samples from $MVN(\mathbf{0}, \Sigma_{\mathbf{G}})$ and from $MVN(\mathbf{0}, \Sigma_{\mathbf{H}})$, where $\Sigma_{\mathbf{G}} = \{0.3^{|k-l|}\}_{p \times p}$ and $\Sigma_{\mathbf{H}} = \{0.3^{|k-l|}\}_{q \times q}$, to mimic LD structure between k - and l -th SNP within gene \mathbf{G} and \mathbf{H} ; 2) categorize each column into 0, 1 or 2 using pre-defined cut-off values to achieve Hardy-Weinberg Equilibrium (HWE) and set MAF to 0.2. We assume survival time of subject i ($i = 1, \dots, n$) follow Cox PH model with the following hazard function

$$\lambda(t|\mathbf{G}_i, \mathbf{H}_i) = 0.5 \exp\left\{\sum_{k=1}^p 0.05G_{ik} + \sum_{l=1}^q 0.05H_{il} + \sum_{m=1}^{pq} \beta_m(GH)_{im}\right\},$$

where $(GH)_{im}$ represents m -th interaction term of \mathbf{G}_i and \mathbf{H}_i . $\{\beta_m\}_{m=1}^{pq}$ were set to 0 and 0.1 for empirical size and power assessment, respectively. Three sample sizes $n = 500, 1000, 1500$

and two SNP-set sizes $p = q = 2, 3$ were considered. Cross-product kernel was used to measure genetic similarity for both \mathbf{G} and \mathbf{H} .

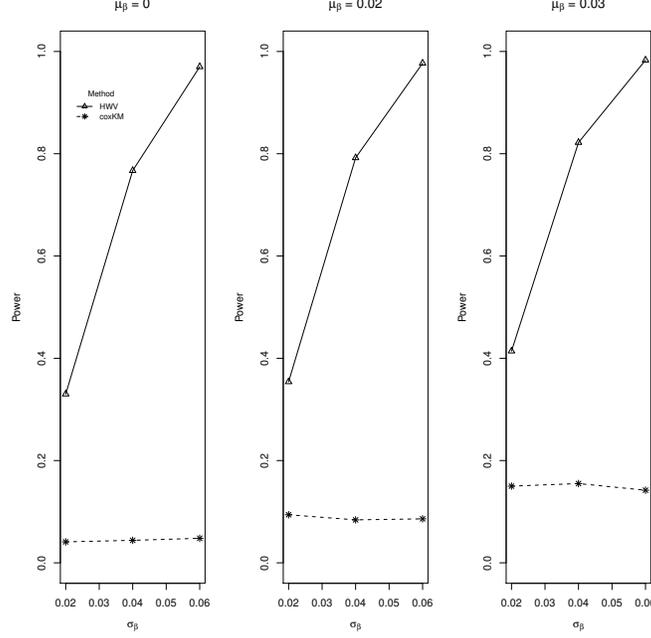


Figure 2.3 Power comparison between HWV and coxKM in the presence of genetic heterogeneity across twenty latent subpopulations when $p = 15$ and $n = 1000$.

For testing G - E interaction effect, the simulation setting is similar to that of testing G - G interaction, except that $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$ is an two-dimensional environmental variable, with $\mathbf{H}_1 \sim \text{Bernoulli}(0.5)$ and $\mathbf{H}_2 \sim \text{Uniform}(0, 2)$. The survival time for subject i ($i = 1, \dots, n$) was generated from CoxPH model with the following hazard function

$$\lambda(t|\mathbf{G}_i, \mathbf{H}_i) = 0.5 \exp \left\{ \sum_{k=1}^p 0.1 G_{ik} + \sum_{l=1}^2 0.1 H_{il} + \sum_{m=1}^{2p} \beta_m (GH)_m \right\}, \quad (2.11)$$

where $(GE)_m$ is m -th interaction term of \mathbf{G} and \mathbf{H} . One sample size $n = 1000$ and one SNP-set size $p = 10$ was considered (q was fixed at 2 since \mathbf{H} has two-dimension). IBS kernel to measure genetic similarity and unified kernel for environmental similarity.

Table 2.8 and 2.9 indicates that, in testing $G-G$ and $G-E$ interaction effect, Type I error of WVI is close to nominal level while LRT could be slightly inflated under small sample size and relative large SNP-set. Under all n 's and p 's WVI is more powerful than LRT as expected because the asymptotic null distribution of WVI has a more effective degree of freedom than LRT, therefore more powerful than LRT. The power advantages of WVI over LRT increase with the strength of LD, measured by ρ in $\Sigma = \{\rho^{|k-l|}\}_{p \times p}$.

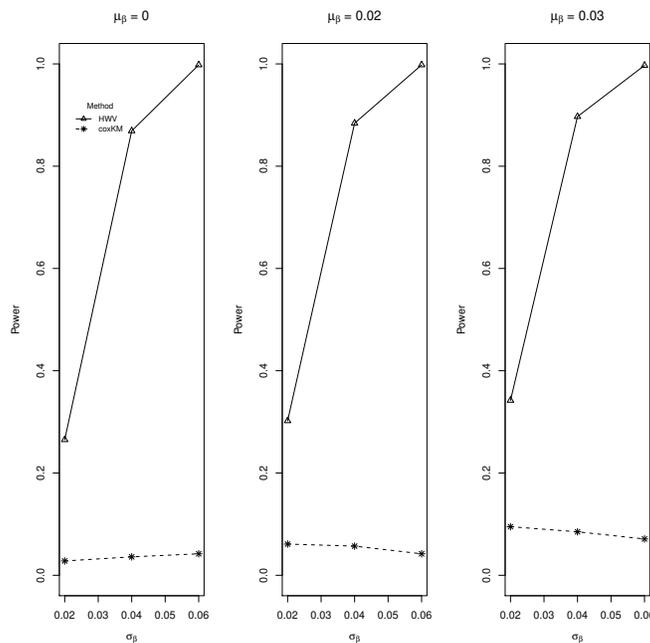


Figure 2.4 Power comparison between HWV and coxKM in the presence of genetic heterogeneity across individual genome profiles when $p = 10$ and $n = 1000$.

2.2.4 Empirical size under stringent p-value thresholds

So far, we have investigated the empirical size of WV and HWV under a significance level of 0.05. However, to account for multiple testing issues in GWAS, a more strict p -value (5×10^{-3} or lower) is often considered. In this simulation, we investigated the empirical size of 1) WV in absence of genetic heterogeneity, with $p = 5, n = 1000$ and 2) HWV in the presence of heterogeneity across observable subpopulations with $p = 10, n = 1000$ under

Table 2.8 Empirical sizes and powers of WVI and LRT in testing G-G interaction without considering left truncation.

	Empirical Size (Power)		
	p=2, q=2, n=500	p=2, q=2, n=1000	p=2, q=2, n=1500
WVI	0.050 (0.255)	0.056 (0.435)	0.051 (0.629)
LRT	0.055 (0.205)	0.054 (0.328)	0.052 (0.494)
	Empirical Size (Power)		
	p=3, q=3, n=500	p=3, q=3, n=1000	p=3, q=3, n=1500
WVI	0.057 (0.488)	0.059 (0.837)	0.057 (0.952)
LRT	0.072 (0.341)	0.064 (0.664)	0.052 (0.864)

Table 2.9 Empirical sizes and powers of HWV and LRT in testing G-E interaction without considering left truncation ($p = 10, q = 2, n = 1000$).

	Empirical Size (Power)
	HWV
LRT	0.071 (0.855)

p -value thresholds that are more strict than 0.05. 350K Monte Carlo samples were generated in the same way as corresponding association tests in previous sections. Table 2.10 shows that the empirical size of WV and HWV is close to the nominal level.

Table 2.10 Empirical size for WV and HWV that considers heterogeneity across individual genome profiles and adjusts for covariates, under several stringent p -value thresholds.

Threshold	Empirical Size†	
	WV (p=5, n=1000)	HWV (p=10, n=1000)
0.05	0.0494	0.0487
0.005	0.00474	0.00453
0.0005	0.000400	0.000489
0.00005	0.0000429	0.0000371

† Proportion of p -values in 350K replicates that are smaller or equal to the corresponding threshold.

2.3 A Real Application

We applied the developed WV and HWV test on the ROSMAP dataset by assuming the age-to-onset of AD follows the Cox PH model. ROSMAP is a GWAS that includes demen-

tia exam data from two large longitudinal studies of aging and dementia: 1) the Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP) (Bennett et al., 2018). The ROSMAP genotype dataset includes 1,639 subjects, each has 750,173 SNPs. We first performed quality control on genotype dataset 1) at SNP level by removing SNPs with $MAF > 0.01$, HWE test's p -value $> 10^{-6}$, or missing rate $> 2\%$, then 2) at the subject level by removing subjects with missing genotype rate $> 2\%$. This results in a quality-controlled genotype dataset with 1,618 subjects each having 619,061 SNPs. The missing genotypes were then imputed with random samples generated from $Binomial(2, MAF)$, where MAF was estimated from the genotype dataset. PCA was performed using Plink software (version 1.9) to obtain the first 20 principal components that represent the underlying ancestry information. After that, we grouped all available SNPs into 21,329 different genes by using human genome annotation (GRCh38/hg38) obtained from the UCSC Genome Browser. We also formed a new gene APOE4 which was coded as the count of the APOE4- ϵ 4 allele.

The ROSMAP phenotype dataset includes information from 2,376 subjects. Each subject has 1) baseline covariates: race, gender, years of education, and binary study cohort indicator (ROS or MAP) and 2) follow-up information: subject's age and clinical cognitive diagnosis result at each examination center visit including at study registry. For our analysis, we include only subjects who are disease-free in the study registry. Age at baseline served as left-truncation time, the age at first AD diagnosis served as survival time, and the age at study termination served as right censoring time if the subject is disease-free throughout the study follow-up period.

We merged preprocessed genotype and phenotype datasets by keeping only the overlapping subjects to generate the analysis-ready dataset, which includes 1,433 subjects, with a

censoring rate of 67.8%. We are interested in 1) testing genetic covariates that are associated with the time-to-onset of first AD diagnosis and 2) detecting potential genetic heterogeneity across subpopulations inferred by either baseline covariates or individual genome profiles (represented by a random sample of 200K SNPs), in the presence of left-truncation time and adjust for gender, year of education, study cohort indicator (ROS served as baseline), and first 20 principal components. Note that race is not adjusted because 1,432 out of 1,433 subjects are white. IBS kernel was used to measure genetic similarity, while the choice of similarity kernel for subpopulation similarity depends on the source of heterogeneity, which is elaborated in Table 2.11.

Various types of heterogeneity have been considered including no genetic heterogeneity (S1), heterogeneity across individual genome profiles (S2), heterogeneity due to years of education (S3), heterogeneity across three education attainment categories (0: yrs of education \leq 12; 1: yrs of education \in (12, 16]; 2: yrs of education $>$ 16) with education similarity between subjects i and j measured by $\kappa_{ij} = I(\mathbf{X}_i = \mathbf{X}_j)$ (S4), heterogeneity across the three education attainment categories with education similarity measured by IBS kernel (S5), and heterogeneity due to sex (S6). FDR-based p-value thresholds were calculated according to Benjamini and Hochberg (1995) under arbitrary dependence assumption, i.e., $\text{Threshold}_i = \frac{\alpha i}{m \sum_{k=1}^m (1/k)}$ ($i = 1, \dots, m$), where m is the number of tests and α is the target FDR.

Table 2.11 shows the analysis results of the ROSMAP dataset. To account for multiple testing issues, the p -value thresholds were obtained by controlling the FDR under 10% by Benjamini-Hochberg procedure Benjamini and Hochberg (1995). When in the absence of genetic heterogeneity (S1), APOE4 appeared to be the most significant gene ($p=1.21\text{E-}10$)

followed by APOC1 ($p=1.04E-07$). When in the presence of genetic heterogeneity (S2-S6), even though APOE4 and APOC1 were still the most significant genes, their p -values vary across scenarios, with notably smaller p -values in S4 that considered genetic heterogeneity across re-categorized year-of-education, with similarity measured by IBS kernel. This indicates the potential genetic heterogeneity due to education levels. In other words, the effect of APOE4 and APOC1 on age-to-onset of AD varies across subjects' education levels.

Table 2.11 Top five significant genes discovered by WV/HWV from a genome-wide association analysis of ROSMAP dataset by assuming age-to-onset of AD follows Cox PH model.

Heterogeneity	Gene names and p-values				
S1	APOE4	APOC1	CAMSAP1	XPR1	GRIP1
	1.21e-10	1.04e-07	2.34e-05	1.48e-04	3.04e-04
S2	APOE4	APOC1	CAMSAP1	XPR1	KU.MEL.3
	4.03e-10	2.31e-07	7.56e-06	6.92e-05	1.97e-04
S3	APOE4	APOC1	CAMSAP1	XPR1	PPM1A
	6.72e-11	4.54e-08	7.94e-05	1.33e-04	3.93e-04
S4	APOE	APOC1	CAMSAP1	XPR1	LINC00911
	3.68e-11	1.68e-08	3.07e-05	4.30e-05	7.88e-05
S5	APOE	APOC1	CAMSAP1	LETMD1	CSRNP2
	4.30e-11	3.64e-08	1.15e-05	3.57e-05	6.17e-05
S6	APOE	APOC1	CAMSAP1	XPR1	PPM1A
	2.87e-10	2.29e-07	2.05e-05	9.72e-05	1.19e-04
Threshold	4.45e-07	8.89e-07	1.33e-06	1.78e-06	2.22e-06

2.4 Discussion

We developed a suite of multi-variant association and interaction tests for survival phenotypes based on weighted V statistics. They have three main advantages over the state-of-the-art methods. First, the new tests are faster. Second, they can account for possible heterogeneity in genetic effects to enhance power. Third, they can handle linear confounding.

Although the covariate adjustment in the (heterogeneity) weighted V tests is based on a Cox model, the tests are robust against model mis-specification. This is because, under a

null model that is not the Cox model,

$$\begin{aligned}
\hat{M}_{\mathbf{Z},i} &= N_i(\infty) - \int_0^\infty Y_i(s) \exp(\mathbf{Z}_i^T \hat{\boldsymbol{\gamma}}) d\hat{\Lambda}_0(s) \\
&= N_i(\infty) - \int_0^\infty \{Y_i(s) \exp(\mathbf{Z}_i^T \boldsymbol{\gamma}^*) + o_p(1)\} \left[\frac{E\{dN_i(s)\}}{E\{Y_i(s) \exp(\mathbf{Z}_i^T \boldsymbol{\gamma}^*)\}} + o_p(1) \right] \\
&\equiv M_{\mathbf{Z},i}^* + o_p(1),
\end{aligned}$$

where $\boldsymbol{\gamma}^*$ is defined as in Lin and Wei (1989), $E(M_{\mathbf{Z},i}^*) = 0$, and thus Theorem 2 with $M_{\mathbf{Z},i}$ replaced by $M_{\mathbf{Z},i}^*$ still holds.

The covariate adjustment proposed for the (heterogeneity) weighted V tests cannot control Type I error rates around the nominal level in the presence of nonlinear confounding, i.e., the genetic covariates follow a nonlinear model about the adjustment covariates. However, our simulation results suggested that the empirical sizes of the proposed tests are smaller, not larger, than the nominal level in this situation (results not shown here due to space limitations). A conservative test does not violate the fundamental principle of small-probability events in hypothesis testing.

The proposed G-G/G-E interaction test assumes linearity on the main effects of the two genes (or the gene and the environmental determinant). When the true main effects are nonlinear, the test may have an incorrect size. A more robust test could estimate the null model using the kernel machine Cox regression (Cai, Tonini, and Lin, 2011), which, however, significantly increases the computational cost.

One may improve the power of the heterogeneity weighted V tests by changing w_{ij} to

$(\rho + \kappa_{ij})f(\mathbf{G}_i, \mathbf{G}_j)$, where ρ is a positive unknown parameter, and then using the test statistic

$$S = \sup_{\rho \in \mathcal{I}} nV^H(\rho)$$

or

$$S_{\mathbf{Z}} = \sup_{\rho \in \mathcal{I}} nV_{\mathbf{Z}}^H(\rho),$$

where \mathcal{I} is a range of ρ under consideration. The asymptotic null distribution of S or $S_{\mathbf{Z}}$ is hard to derive. A workaround is to calculate S or $S_{\mathbf{Z}}$ for many permutations of the n subjects' martingale residuals to obtain a permutation-based p-value. But this permutation test requires an assumption that there is no relationship between the genetic variables and the adjustment covariates when the latter is present.

We can also develop a set of weighted U tests in parallel to the weighted V tests developed herein. For example, a covariate-adjusted weighted U test for genetic association is based on the following weighted U statistic,

$$U_{\mathbf{Z}} = \frac{1}{n(n-1)} \sum_{i \neq j} \hat{f}_{\mathbf{Z}}(\mathbf{g}_i, \mathbf{g}_j) \hat{M}_{\mathbf{Z},i} \hat{M}_{\mathbf{Z},j},$$

where $\hat{f}_{\mathbf{Z}}(\mathbf{g}_i, \mathbf{g}_j)$ is the ij -th element of $(I - H)^T F (I - H)$. Following the derivation of the large-sample distribution of $nV_{\mathbf{Z}}$, it is easy to show that when n is large, the null distribution of $nU_{\mathbf{Z}}$ can be approximated by

$$nU_{\mathbf{Z}} \sim \hat{\xi} \sum_{t=1}^n \hat{v}_t (\chi_{1t}^2 - 1).$$

The weighted U and V tests can be used interchangeably, as they have the same asymptotic efficiency.

CHAPTER 3 Multi-marker genetic association and interaction tests with interval-censored survival outcomes

3.1 Methods

3.1.1 Association test

Consider a cohort study of n independent subjects, who are disease-free at the study registry. Denote genetic covariates $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$, where \mathbf{G}_j is the j -th genetic variant that can be either a SNP (coded as 0, 1, or 2) or gene expression values. $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_d)$ is a d -dim adjustment covariates. Survival times, T , are subject to the interval censoring and possible left truncation. Interval left and right endpoints are denoted as L and R , and left truncation time is denoted as A . Two sets of tests were developed to detect genetic association with interval-censored survival outcomes by assuming the effect of \mathbf{G} is homogeneous or heterogeneous across subpopulations indicated explicitly by \mathbf{X} or by latent variables inferred by \mathbf{X} . Therefore, we denote the observed data as $D = \{(\mathbf{G}_i, \mathbf{Z}_i, L_i, R_i, A_i, \mathbf{X}_i)\}_{i=1}^n$ (\mathbf{X}_i only used in the presence of genetic heterogeneity). We assume that examination times $\{(L_i, R_i)\}_{i=1}^n$ are independent of event risk (In that case one can in the analysis ignore the distribution of the examination times, and treat the examination times as fixed.). We also assume non-informative censoring and truncation time given \mathbf{G} and \mathbf{Z} (and \mathbf{X} in the presence of genetic heterogeneity).

Consider the scenario when in the absence of genetic heterogeneity. Under the null hypothesis that \mathbf{G} is independent of T given \mathbf{Z} , we assume a semi-parametric transformation model for the cumulative hazard function of T given \mathbf{Z} as follow

$$\Lambda(t|\mathbf{Z}) = G[\Lambda_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{Z}\}],$$

where $\Lambda_0(t)$ is unspecified cumulative baseline hazard function with $\Lambda_0(0) = 0$. $G(x) = -\log \int_0^\infty e^{-xt} f(t) dt$, where $f(t)$ is a known density function whose support is \mathcal{R}^+ . Choose $f(t)$ to be a gamma density with mean 1 and variance r yields the logarithmic transformation: $G(x) = \log(1 + rx)/x$, which becomes Cox proportional hazard and proportional odds model when setting $r=0$ and $r=1$.

Motivated by association test proposed in Chapter 2, we introduce a subject-level frailty (h) and use a working semiparametric transformation frailty model with cumulative hazard function: $\Lambda(t|\mathbf{Z}, h) = G[\Lambda_0(t) \exp\{\beta^T \mathbf{Z} + h\}]$. The covariate-adjusted weighted V test statistic can be constructed similarly to equation 2.6 by replacing martingale residuals $\{M_{\mathbf{Z},i}\}_{i=1}^n$ by score function $\{S_{\mathbf{Z},i}\}_{i=1}^n$, where $S_{\mathbf{Z},i} = \partial \log \mathcal{L}_i(\Lambda_0, \boldsymbol{\beta}, h_i) / \partial h_i |_{h_i=0}$. $\mathcal{L}_i(\Lambda_0, \boldsymbol{\beta}, h_i)$ is conditional likelihood function of subject i given frailty h_i and left-truncation time A_i , as follows

$$\mathcal{L}_i(\Lambda_0, \boldsymbol{\beta}, h_i) = \frac{\exp(-G[\Lambda_0(L_i)e^{\beta^T \mathbf{Z}_i + h_i}]) - \exp(-G[\Lambda_0(R_i)e^{\beta^T \mathbf{Z}_i + h_i}])}{\exp(-G[\Lambda_0(A_i)e^{\beta^T \mathbf{Z}_i + h_i}])}.$$

The weighted V test statistics is as follow,

$$V_{\mathbf{Z},IC} = n^{-2} \sum_{i=1}^n \sum_{j=1}^n \tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j) S_{\mathbf{Z},i} S_{\mathbf{Z},j}, \quad (3.1)$$

where $S_{\mathbf{Z},i} S_{\mathbf{Z},j}$ is considered as V statistic kernel that measures phenotype similarity, weighted

by covariate-centered genetic similarity $\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j)$,

$$\begin{aligned}\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j) &= f(\mathbf{G}_i, \mathbf{G}_j) - E[f(\mathbf{G}_i, \mathbf{G}_k)u(\mathbf{Z}_k, \mathbf{Z}_j)|\mathbf{G}_i, \mathbf{Z}_j] \\ &\quad - E[f(\mathbf{G}_j, \mathbf{G}_k)u(\mathbf{Z}_k, \mathbf{Z}_i)|\mathbf{G}_j, \mathbf{Z}_i] \\ &\quad + E[u(\mathbf{Z}_i, \mathbf{Z}_m)f(\mathbf{G}_m, \mathbf{G}_k)u(\mathbf{Z}_k, \mathbf{Z}_j)|\mathbf{Z}_i, \mathbf{Z}_j],\end{aligned}$$

where $f(\mathbf{G}_i, \mathbf{G}_j)$ is genetic similarity and $u(\mathbf{Z}_i, \mathbf{Z}_j) = (\mathbf{1}, \mathbf{Z}_i^T)[E\{(\mathbf{1}, \mathbf{Z}^T)^T(\mathbf{1}, \mathbf{Z}^T)\}]^{-1}(\mathbf{1}, \mathbf{Z}_j^T)^T$.

When \mathbf{G} has effect on T , the larger (smaller) phenotype similarity, weighted by larger (smaller) genetic similarity, results in larger (smaller) weighted V statistic and more (less) significant p -value.

When \mathbf{Z} is independent of \mathbf{G} , the asymptotic null distribution of $V_{\mathbf{Z},IC}$ can be obtained similarly following Theorem 2 in Chapter 2. Denote $\xi = E(S_{\mathbf{Z}}^2)$, $\{\nu_t\}_{t=1}^{\infty}$ are eigenvalues of $\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j)$, the asymptotic null distribution of weighted V statistic is

$$nV_{\mathbf{Z},IC} \sim \xi \sum_{t=1}^{\infty} \nu_t \chi_{1t}^2, \quad (3.2)$$

where $\{\chi_{1t}^2\}_{t=1}^{\infty}$ are independent χ^2 random variables of 1 degree of freedom.

When \mathbf{Z} and \mathbf{G} are linearly correlated, $\mathbf{G} = \mathbf{a} + \mathbf{b}^T \mathbf{Z} + \mathbf{e}$, where \mathbf{a} and \mathbf{b} are intercept and regression coefficients and \mathbf{e} is a normally distributed random error that is independent of \mathbf{Z} . Theorem 2 in Chapter 2 suggests that the distribution in 3.2 is still asymptotically correct.

In real application the regression coefficients $\hat{\beta}$ and cumulative baseline hazard function $\hat{\Lambda}_0(t)$ in 3.1 are obtained by fitting model $\Lambda(t|\mathbf{Z}, h = 0) = G[\Lambda_0(t)e^{\beta^T \mathbf{Z}}]$ with $r = 0$ using

unpenalized version of penalized nonparametric maximum likelihood estimation (PNPMLE) for Cox regression and possibly left truncated data (Li, Pak, and Todem, 2020). When $r > 0$ is considered, a more general method (Zeng, Mao, and Lin, 2016) can be used to obtain NPMLE of $\boldsymbol{\beta}$ and $\Lambda_0(t)$. Due to the lack of available estimation methods for the Porportional Odds model with left-truncation and interval-censored data, left truncation was considered only under the Cox proportional hazards model. $\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j)$ is estimated by its sample mean denoted as $(\mathbf{I} - \mathbf{H})\mathbf{F}(\mathbf{I} - \mathbf{H})$, where $\mathbf{F} = \{f(\mathbf{G}_i, \mathbf{G}_j)\}_{n \times n}$ with $f(\mathbf{G}_i, \mathbf{G}_j)$ the genetic similarity between any pair of subjects i and j , $\mathbf{H} = \tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}^T \tilde{\mathbf{Z}})^{-1} \tilde{\mathbf{Z}}^T$ with $\tilde{\mathbf{Z}} = (\mathbf{1}, \mathbf{Z})$. The weighted V statistic can be rewritten as

$$V_{\mathbf{Z}, IC} = \hat{\mathbf{S}}_{\mathbf{Z}}^T (\mathbf{I} - \mathbf{H}) \mathbf{F} (\mathbf{I} - \mathbf{H}) \hat{\mathbf{S}}_{\mathbf{Z}}, \quad (3.3)$$

where $\hat{\mathbf{S}}_{\mathbf{Z}} = (\hat{\mathbf{S}}_{\mathbf{Z},1}, \dots, \hat{\mathbf{S}}_{\mathbf{Z},n})$.

Estimate of $\xi = E(S_{\mathbf{Z}}^2)$, denoted as $\hat{\xi}$, can be obtained by $\hat{S}_{\mathbf{Z}}^T \hat{S}_{\mathbf{Z}}/n$, where $\hat{S}_{\mathbf{Z},i} = \partial \mathcal{L}_i(\Lambda_0(t), \boldsymbol{\beta}, h_i) / \partial h_i |_{\Lambda_0(t)=\hat{\Lambda}_0(t), \boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, h_i=0}$. The estimate of ν_t is estimated as $\hat{\nu}_t = \tilde{\nu}_t/n$, where $\tilde{\nu}_t$'s are eigenvalues of $\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j)$. The asymptotic null distribution of $nV_{\mathbf{Z}, IC}$ can be approximated by

$$nV_{\mathbf{Z}, IC} \sim \hat{\xi} \sum_{t=1}^n \hat{\nu}_t \chi_{t1}^2,$$

which is linear combination of independent χ^2 random variables of 1 degree of freedom. p -value can be computed using Davies' method (Davies, 1980) $P(nV_{\mathbf{Z}, IC} \geq nV_{\mathbf{Z}, IC}^{obs})$, where $V_{\mathbf{Z}, IC}^{obs}$ is the observed value of $V_{\mathbf{Z}, IC}$.

When in the absence of adjustment covariates \mathbf{Z} , the centered genetic similarity is denoted

as $\tilde{f}(\mathbf{G}_i, \mathbf{G}_j)$ and is estimated by sample mean $(\mathbf{I} - \mathbf{J})\mathbf{F}(\mathbf{I} - \mathbf{J})$, where $\mathbf{J} = \{1\}_{n \times n}$. The corresponding weighted V statistic is $V_{IC} = \hat{\mathbf{S}}^T(\mathbf{I} - \mathbf{J})\mathbf{F}(\mathbf{I} - \mathbf{J})\hat{\mathbf{S}}$, and the asymptotic null distribution can be approximated similarly as above.

When in the presence of genetic heterogeneity, the effect of \mathbf{G} varies across different subpopulations, the association tests V_{IC} and $V_{\mathbf{Z},IC}$ is extended to consider the subpopulation structure, thereby improving test power. The setting is the same as the association test in the absence of heterogeneity, except that \mathbf{X} is now introduced to infer explicit or latent subpopulation structure. The heterogeneity weighted V statistics, with or without covariate adjustment are

$$\begin{cases} V_{IC}^H = \hat{\mathbf{S}}^T(\mathbf{I} - \mathbf{J})\mathbf{W}(\mathbf{I} - \mathbf{J})\hat{\mathbf{S}} \\ V_{\mathbf{Z},IC}^H = \hat{\mathbf{S}}^T(\mathbf{I} - \mathbf{H})\mathbf{W}(\mathbf{I} - \mathbf{H})\hat{\mathbf{S}}, \end{cases}$$

where $\mathbf{W} = (\mathbf{J} + \mathbf{K}) \odot \mathbf{F}$. $\mathbf{K} = \{k(\mathbf{X}_i, \mathbf{X}_j)\}_{n \times n}$ with the ij -th element represents subpopulation similarity of subject i and j . \odot is the element-wise matrix multiplication. The asymptotic null distribution of nV_{IC}^H and $nV_{\mathbf{Z},IC}^H$ can be approximated similarly as that of nV_{IC} and $nV_{\mathbf{Z},IC}$ using linear combination of independent χ^2 random variable with 1 degree of freedom. Based on this distribution, p -value is calculated using Davies' method (Davies, 1980) by $P(nV_{IC}^H \geq nV_{IC,obs}^H)$ and $P(nV_{\mathbf{Z},IC}^H \geq nV_{\mathbf{Z},IC,obs}^H)$. $V_{IC,obs}^H$ and $V_{\mathbf{Z},IC,obs}^H$ are observed values of V_{IC}^H and $V_{\mathbf{Z},IC}^H$.

3.1.2 G-G/G-E interaction test

Two interaction tests were developed to detect the interaction effect between two genes (G - G) or between a gene and environmental variable (G - E). Let $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ be the interested gene, let $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_q)$ be a different gene or an environmental variable when there exists G - G or G - E interaction effect.

To test G - G interaction effect, we assume that given \mathbf{G} and \mathbf{H} , the survival time T for subject i follow a semi-parametric transformation model with cumulative hazard function as follow

$$\Lambda(t|\mathbf{G}_i, \mathbf{H}_i) = G[\Lambda_0(t) \exp\{\boldsymbol{\beta}^T \mathbf{G}_i + \boldsymbol{\alpha}^T \mathbf{H}_i\}]. \quad (3.4)$$

The V statistic kernel is $S_i S_j$ with $S_i = \partial \mathcal{L}_i(\Lambda_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, h_i) / \partial h_i |_{\Lambda_0 = \hat{\Lambda}_0, \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}, \boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}, h_i = 0}$. The conditional likelihood function for subject i , $\mathcal{L}_i(\Lambda_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, h_i)$, can be written as follow

$$\mathcal{L}_i(\Lambda_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, h_i) = \frac{\exp(-G[\Lambda_0(L_i) e^{\boldsymbol{\beta}^T \mathbf{G}_i + \boldsymbol{\alpha}^T \mathbf{H}_i h_i}]) - \exp(-G[\Lambda_0(R_i) e^{\boldsymbol{\beta}^T \mathbf{G}_i + \boldsymbol{\alpha}^T \mathbf{H}_i h_i}])}{\exp(-G[\Lambda_0(A_i) e^{\boldsymbol{\beta}^T \mathbf{G}_i + \boldsymbol{\alpha}^T \mathbf{H}_i h_i}])}. \quad (3.5)$$

The weighted V statistic for G - G interaction is $V_{IC}^I = \hat{\mathbf{S}}^T (\mathbf{I} - \mathbf{O})(\mathbf{F} \odot \mathbf{K})(\mathbf{I} - \mathbf{O})\hat{\mathbf{S}}$, where $\mathbf{O} = \mathbf{Q}(\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T$, $\mathbf{Q} = (\mathbf{1}, \mathbf{G}, \mathbf{H})$. $\mathbf{F} = \{f(\mathbf{G}_i, \mathbf{G}_j)\}_{n \times n}$ and $\mathbf{K} = \{k(\mathbf{H}_i, \mathbf{H}_j)\}_{n \times n}$ are genetic similarity matrices. The asymptotic null distribution of nV_{IC}^I can be approximated similarly as that of $nV_{\mathbf{Z}, IC}$. When $p + q$ is large relative to sample size n , the asymptotic null distribution of V_{IC}^I is . p -value can be computed using Davies' method (Davies, 1980) $P(nV_{IC}^I \geq nV_{IC, obs}^I)$, where $V_{IC, obs}^I$ is observed value of V_{IC}^I . The test for G - E interaction effect is performed in the same way as that of G - G interaction test except that \mathbf{H} is replaced by environmental variable.

3.2 Simulations

Monte Carlo simulation was performed to assess the finite-sample performance of (heterogeneity) weighted V statistic in detecting genetic association with time-to-onset of interval-censored and possibly left-truncated survival outcomes. In all simulation settings, unless otherwise specified, two sample sizes $n = 500, 1000$ and two SNP-set sizes $p = 4, 8$ were

considered. For each subject, we assume three examine center visits at ages $\{V_1, V_2, V_3\}$, where $V_1 \sim Uniform(L, R)$, $V_2 = V_1 + Uniform(L, R)$, and $V_3 = V_2 + Uniform(L, R)$ with predetermined positive values L and R to control the left censoring rate $\sim 25\%$ and right censoring rate $\sim 35\%$. When adjust for covariates, one binary variable $Z_1 \sim Binomial(0.5)$ and one continuous variable $Z_2 \sim Uniform(0, 2)$ were considered. Genetic covariates $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ is a $n \times p$ SNP set generated using a two-step procedure 1) generate n random samples from multivariate normal distribution $MVN(\mathbf{0}, \Sigma)$, where $\Sigma = \{0.5^{|k-l|}\}_{p \times p}$ is covariance matrix that used to mimic LD between any pair of SNPs, 2) categorize each quantitative column into SNP with three levels labeled as 0, 1, and 2 by using predetermined quantile cutoffs: a^2 , $2a(1-a)$, and $(1-a)^2$ of a standard normal distribution with $a = 0.2$, so that the resulting population is in HWE and each SNP has MAF=0.2. When in the presence of genetic heterogeneity, \mathbf{X} was generated accordingly to infer explicit/latent subpopulation structure. Survival time T was assumed to follow either Cox proportional hazard or proportional odds model, two special cases of the class of semiparametric transformation model. However, left-truncated survival time was not considered under the proportional odds model, due to the lack of an available estimation method, and the simulation results were shown in Appendix. In all simulations, 1000 Monte Carlo samples were generated and the significance level was set to 0.05.

3.2.1 Testing genetic association in the absence of genetic heterogeneity

In this series of simulation, we investigated the performance of weighted V statistic in detecting genetic association considering adjustment covariates and left-truncation times, in the absence of genetic heterogeneity. Survival time for subject i ($i = 1, \dots, n$) follows Cox

PH model (constant baseline hazard) with the following form of hazard function,

$$\lambda(t|\mathbf{Z}_i, \mathbf{G}_i) = \exp\{\boldsymbol{\beta}^T \mathbf{G}_i + \boldsymbol{\alpha}^T \mathbf{Z}_i\}, \quad (3.6)$$

where the value of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ vary in different simulation settings to achieve moderate left- and right-censoring rate ($\sim 30\%$).

Empirical size and power under various n 's and p 's

In this simulation, we assumed that \mathbf{Z} is independent of \mathbf{G} , and the empirical size and power of the weighted V test were evaluated under various n 's and p 's. Regression coefficients in hazard function 3.6 are set to $\boldsymbol{\beta} = \mathbf{0}$, $\boldsymbol{\alpha} = \mathbf{0.5}$ for empirical size assessment and $\boldsymbol{\beta} = \boldsymbol{\alpha} = \mathbf{0.5}$ for power assessment. We set $L = 0.1$ and $R = 0.6$ as endpoints of uniform distribution used in follow-up time generation. IBS kernel was used to measure genetic similarity. Table 3.1 shows that the Type I error of WV-IC is close to the nominal level and the power increases with sample size.

Table 3.1 Empirical size and power of WV-IC with $r = 0$ in testing genetic effects in the absence of left truncation.

	Empirical Size (Power)	
	p=4, n=500	p=4, n=1000
WV-IC	0.047 (0.231)	0.057 (0.410)
	p=8, n=500	p=8, n=1000
	WV-IC	0.048 (0.415)

Empirical size and power under linear confounding

In this simulation, we assumed that there exists linear confounding effect, when \mathbf{Z} has linear relationship with \mathbf{G} . The genetic covariate of our interest, $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$, are gene expression values with $\mathbf{G}_i = Z_{i1} + Z_{i2} + \mathbf{e}_i$ ($i = 1, \dots, n$). The random error is $\mathbf{e}_i \sim$

$MVN(\mathbf{0}, \Sigma)$ with $\Sigma = \{0.5^{|k-l|}\}_{p \times p}$. Regression coefficients in hazard function 3.6 was set to $\beta = \mathbf{0}$, $\alpha = \mathbf{0.05}$ for empirical size assessment and $\beta = \alpha = \mathbf{0.02}$ for power assessment. We set $L = 0.1$ and $R = 0.6$ in follow-up time generation. Cross-product kernel was used to measure gene expression similarity. Table 3.2 shows that WV-IC controls Type I error well around nominal level and gains power as sample size increases.

Table 3.2 Empirical size and power of WV-IC with $r = 0$ in testing genetic effects under linear confounding and no left truncation.

	Empirical Size (Power)	
	p=4, n=500	p=4, n=1000
WV-IC	0.050 (0.158)	0.055 (0.313)
	p=8, n=500	p=8, n=1000
	WV-IC	0.041 (0.320)

3.2.2 Testing genetic association in the presence of genetic heterogeneity

So far, we have assumed that the effect of \mathbf{G} is homogeneous in the population. In this series of simulations, we introduced genetic heterogeneity, and the performance of HWV-IC and WV-IC will be evaluated under four sources of heterogeneity 1) observable subpopulations, 2) two latent subpopulations, 3) twenty latent subpopulations, and 4) individual genome profile.

Genetic heterogeneity across observable subpopulations

In this simulation, we investigated the performance of HWV-IC and WV-IC in the presence of genetic heterogeneity across four observable subpopulations, which was inferred by $\mathbf{Z} = (\mathbf{Z}_1 = I(1 \leq \mathbf{U} < 2), \mathbf{Z}_2 = I(2 \leq \mathbf{U} < 3), \mathbf{Z}_3 = I(3 \leq \mathbf{U} \leq 4))$, where $\mathbf{U} = (U_1, \dots, U_n) \sim Uniform(0, 4)$ and $I(\cdot)$ is an indicator function. The genetic covariate \mathbf{G} was generated following the two-step procedure, but four different sets of predetermined

MAFs ($\sim Uniform(0.005, 0.05)$) and LD structures (measured by ρ through $\Sigma_{p \times p} = \{\rho^{|k-l|}\}$) were used as follows,

MAFs = $\{0.0118, 0.0304, 0.0476, 0.0450, 0.0393, 0.0076, 0.0459, 0.0054\}$ and $\rho = 0.2$ in subpopulation 1;

MAFs = $\{0.0378, 0.0421, 0.0312, 0.0467, 0.0156, 0.0085, 0.0362, 0.0336\}$ and $\rho = 0.5$ in subpopulation 2;

MAFs = $\{0.0157, 0.0299, 0.0220, 0.0119, 0.0272, 0.0093, 0.0063, 0.0354\}$ and $\rho = 0.7$ in subpopulation 3;

MAFs = $\{0.0257, 0.0323, 0.0289, 0.0252, 0.0078, 0.0319, 0.0439, 0.0193\}$ and $\rho = 0.6$ in subpopulation 4.

Survival time for subject i ($i = 1, \dots, n$) was generated from Cox PH model with the following hazard function,

$$\lambda(t|\mathbf{G}_i, \mathbf{Z}_i) = \exp\left\{\sum_{k=1}^p (\beta_1 Z_{i1} + \beta_2 Z_{i2} + \beta_3 Z_{i3} + \beta_4) G_{ik} + 0.6Z_{i1} + 0.6Z_{i2} + 0.2Z_{i3}\right\},$$

where (Z_{i1}, Z_{i2}, Z_{i3}) is a three-dimensional dummy variable that indicates which subpopulation subject i belongs to. We set $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ for empirical size assessment and $\beta_1 = 0.4, \beta_2 = 0.002, \beta_3 = 0.0005, \beta_4 = 0.0005$ for power assessment. $L = 0.1$ and $R = 0.5$ were the endpoints of uniform distribution used in follow-up time generation. IBS kernel was used to measure genetic similarity and identity kernel for subpopulation inferred by \mathbf{Z} . Table 3.3 indicates that Type I error of both HWV-IC and WV-IC are close to nominal level and HWV-IC is more powerful than WV-IC as expected since HWV-IC takes advantages of

subpopulation structure in performing the test to increase statistical power.

Table 3.3 Comparison of performance of WV-IC and HWV-IC ($r = 0$) in testing genetic association under genetic heterogeneity across four observable subpopulations and no left truncation.

	Empirical Size (Power)	
	p=4, n=500	p=4, n=1000
HWV-IC	0.040 (0.535)	0.046 (0.878)
WV-IC	0.042 (0.384)	0.049 (0.647)
	p=8, n=500	p=8, n=1000
HWV-IC	0.046 (0.801)	0.053 (0.998)
WV-IC	0.047 (0.552)	0.044 (0.889)

Genetic heterogeneity across two latent subpopulations

In this simulation, we investigated performance of HWV-IC in the presence of two latent subpopulations, inferred by one-dim vector $\mathbf{X} = \{X_i\}_{i=1}^n$, with $X_i = a_i + e_i$, where $a_i = \text{Bernoulli}(0.5)$ and $e_i \sim \text{Normal}(0, 0.5)$. Survival time for subject i in subpopulation j was generated from Cox PH model with the following hazard function,

$$\lambda(t|\mathbf{G}_{ij}, \mathbf{Z}_{ij}) = \exp\left\{\sum_{k=1}^p G_{ijk}\beta_{jk} + 0.05(Z_{ij1} + Z_{ij2})\right\}, \quad (3.7)$$

where $\{\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp}) \text{ s.t. } \beta_{j1} = \dots = \beta_{jp}\}_{j=1}^2$, denoted as $\{\boldsymbol{\beta}_j\}_{j=1}^2$, are effect sizes of \mathbf{G} in two subpopulations. $\{\boldsymbol{\beta}_j\}_{j=1}^2$ were set to 0 for empirical assessment and set to different values power assessment under five different scenarios representing four different forms of genetic heterogeneity. We set $L = 0.1$ and $R = 0.5$ for follow-up time generation. IBS kernel was used to measure genetic similarity and Gaussian kernel for subpopulation similarity. Table 3.4 indicates that Type I error of HWV-IC is close to nominal level. Table 3.5 shows power assessment under no heterogeneity (T1) and four different forms of heterogeneity (T2

- T5). It can be observed that power of HWV-IC increase with sample size and size of heterogeneity ($|\beta_{1k} - \beta_{2k}|$).

Table 3.4 Empirical size of HWV-IC with $r = 0$ in testing genetic effects under genetic heterogeneity across two latent subpopulations and no left truncation.

	Empirical Size	
	p=4, n=500	p=4, n=1000
HWV-IC	0.042	0.049
HWV-IC	p=8, n=500	p=8, n=1000
	0.047	0.045

Table 3.5 Power of HWV-IC with $r = 0$ in testing genetic effects under genetic heterogeneity across two latent subpopulations and no left truncation. Various heterogeneity scenarios were considered, determined by the values of β_{1k} and β_{2k} , including the same effect size and the same effect direction (T1), identical sizes but opposite directions (T2), no effect in one subpopulation while positive effect in the other (T3), different sizes and opposite directions (T4), and different sizes but the same direction (T5).

	Heterogeneity Scenario									
	T1		T2		T3		T4		T5	
β_{1k}	0.05	0.1	-0.05	-0.1	0	0	-0.025	-0.025	0.02	0.02
β_{2k}	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05	0.1
p=4, n=500	0.153	0.508	0.079	0.213	0.063	0.185	0.083	0.201	0.079	0.202
p=4, n=1000	0.257	0.862	0.136	0.402	0.110	0.369	0.108	0.305	0.162	0.440
p=8, n=500	0.210	0.790	0.232	0.735	0.139	0.441	0.167	0.437	0.142	0.450
p=8, n=1000	0.423	0.989	0.477	0.976	0.230	0.801	0.280	0.788	0.258	0.812

Genetic heterogeneity across twenty latent subpopulations

In this simulation, we investigated the performance of HWV-IC in the presence of twenty latent subpopulations. The simulation setting is similar to simulation above except that the number of subpopulation increased from two to twenty. $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{25})$ was generated to infer subpopulation structure with $\mathbf{X}_k = \mathbf{a}_k + \mathbf{e}_k$, \mathbf{a}_k is a length n bootstrap sample from $\{1, 2, \dots, 20\}$ and $\mathbf{e}_k \sim Normal(0, 0.5)$. Regression coefficients, $\{\beta_j\}_{j=1}^{20}$, in hazard function 3.7 were set to 0 for empirical size assessment. For power assessment, $\{\beta_j\}_{j=1}^{20}$ were generated

from uniform distribution with mean μ_β and variance σ_β^2 . We set $L = 0.1$ and $R = 0.6$ for follow-up time generation. IBS kernel was used to measure genetic similarity and Gaussian kernel for subpopulation similarity. Table 3.6 indicates that Type I error of HWV-IC is close to nominal level under various n 's and p 's. Table 3.7 indicates the power of HWV-IC increases as mean effect size (μ_β) increases. For a mean effect size (μ_β), the power increases with size of genetic heterogeneity (σ_β).

Table 3.6 Empirical size of HWV-IC with $r = 0$ in testing genetic effects under genetic heterogeneity across twenty latent subpopulations and no left truncation.

	Empirical Size	
	p=4, n=500	p=4, n=1000
HWV-IC	0.051	0.050
	p=8, n=500	p=8, n=1000
	0.045	0.052

Table 3.7 Power of HWV-IC with $r = 0$ under genetic heterogeneity across twenty latent subpopulations and no left truncation.

	Power								
	0	0	0	0.025	0.025	0.025	0.05	0.05	0.05
μ_β	0.05	0.1	0.15	0.05	0.1	0.15	0.05	0.1	0.15
σ_β	0.063	0.148	0.369	0.105	0.240	0.483	0.224	0.422	0.640
p=4, n=500	0.108	0.376	0.804	0.252	0.562	0.906	0.532	0.796	0.969
p=4, n=1000	0.179	0.766	0.993	0.283	0.862	0.996	0.549	0.933	0.999
p=8, n=500	0.434	0.989	1.000	0.687	0.999	1.000	0.908	1.000	1.000
p=8, n=1000									

Genetic heterogeneity across individual genome profile

In this simulation, instead of assuming the subpopulation structure to be 'categorical' (two or twenty subpopulations), we investigated the performance of HWV-IC in the presence of heterogeneity across 'continuous' subpopulation structure. In other words, each individual is a 'subpopulation'. Regression coefficients in hazard function 3.7, $\{\beta_j\}_{j=1}^n$, were set to 0

for empirical assessment and sampled from uniform distribution with mean μ_β and variance σ_β^2 for power assessment. We generated a 1000-SNP genome profile, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{1000})$, to infer subpopulation structure. The generation of \mathbf{X} follow a two-step procedure, 1) generate a length n sample from $MVN(\mathbf{0}, \Sigma)$, repeat for 1,000 times to get a $n \times 1000$ matrix, the covariance matrix Σ was set to $I_{n \times n}$ under null hypothesis and $\exp\{-|\beta_{il} - \beta_{jl}|/\sigma_\beta\}$ under alternative hypothesis. 2) categorize each quantitative column into SNP coded as 0, 1, and 2 using prespecified cutoffs in order to achieve HWE and MAFs that follow uniform distribution $Uniform(0.05, 0.2)$. We set (L, R) to be $(0.2, 0.6)$ and $(0.1, 0.5)$, respectively, for empirical size and power assessment. IBS kernel was used to measure both genetic and genome profile similarity. Table 3.8 indicates that Type I error of HWV-IC is close to nominal level. Table 3.9 indicates that HWV-IC gains power as mean effect size (μ_β) increases. For a fixed mean effect size (μ_β), HWV-IC gains power with heterogeneity size (σ_β).

Table 3.8 Empirical size of HWV-IC with $r = 0$ in testing genetic effects under genetic heterogeneity across individual genome profiles and no left truncation.

	Empirical Size	
	p=4, n=500	p=4, n=1000
HWV-IC	0.043	0.052
	p=8, n=500	p=8, n=1000
	HWV-IC	0.047

3.2.3 Testing G-G/G-E interaction effect

In this simulation, we investigated the performance of WVI-IC and LRT in testing G - G and G - E interaction effect. For testing G - G interaction between two genes $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$, $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_q)$. The survival time for subject i was generated from Cox PH model with

Table 3.9 Power of HWV-IC with $r = 0$ under genetic heterogeneity across individual genome profiles and no left truncation.

	Power								
	0	0	0	0.02	0.02	0.02	0.04	0.04	0.04
μ_β	0.04	0.08	0.12	0.04	0.08	0.12	0.04	0.08	0.12
σ_β	0.062	0.076	0.103	0.089	0.116	0.149	0.170	0.190	0.252
p=4, n=500	0.069	0.087	0.223	0.125	0.204	0.369	0.330	0.442	0.590
p=4, n=1000	0.094	0.280	0.693	0.170	0.388	0.765	0.347	0.589	0.850
p=8, n=500	0.117	0.750	0.996	0.326	0.856	0.999	0.699	0.957	1.000
p=8, n=1000									

the following hazard function

$$\lambda(t|\mathbf{G}_i, \mathbf{H}_i) = \exp\left\{\sum_{k=1}^p 0.02G_{ik} + \sum_{l=1}^q 0.02H_{il} + \sum_{m=1}^{pq} \beta_m(GH)_{im}\right\}, \quad (3.8)$$

where $(GH)_{im}$ is the m -th column of interaction between \mathbf{G}_i and \mathbf{H}_i . $\{\beta_m\}_{m=1}^{pq}$ were set to 0 for empirical size assessment and 0.08 for power assessment. We set $L = 0.1$ and $R = 0.5$ for follow-up time generation. Cross-product kernel was used to measure genetic similarity for both genes. Two sets of SNP-set size were used ($p = 4, q = 2$) and ($p = 6, q = 3$).

For testing G - E interaction, the simulation setting is similar to G - G interaction test, except that $\mathbf{H} = (H_1, \dots, H_n) \sim \text{Binomial}(0.5)$ is now a one-dim binary vector representing subjects' gender. Survival time for subject i was generated from Cox proportional hazard model with hazard function 3.8. $\{\beta_m\}_{m=1}^{pq}$ were set to 0 and 0.1 for empirical size and power assessment. Cross-product kernel was used to measure genetic similarity and identity kernel for gender similarity. Two sets of SNP-set size ($p = 4, q = 1$) and ($p = 8, q = 1$) were investigated.

Table 3.10 and 3.11 indicate that, in testing both G - G and G - E interaction, Type I error of WVI-IC is close to nominal level while LRT is inflated. Also, in both scenarios, WVI-IC

is more powerful than LRT as expected because the asymptotic null distribution of WVI-IC has a more efficient degree of freedom than LRT, therefore more powerful than LRT.

Table 3.10 Empirical sizes and powers of WVI-IC and LRT in testing G-G interaction under no left truncation.

	Empirical Size (power)	
	p=4, q=2, n=500	p=4, q=2, n=1000
WVI-IC	0.046 (0.458)	0.052 (0.793)
LRT	0.069 (0.305)	0.073 (0.519)
	Empirical Size (power)	
	p=6, q=3, n=500	p=6, q=3, n=1000
WVI-IC	0.060 (0.864)	0.057 (0.989)
LRT	0.115 (0.590)	0.076 (0.863)

Table 3.11 Empirical sizes and powers of WVI-IC and LRT in detecting G-E interaction under no left truncation.

	Empirical Size (power)	
	p=4, q=1, n=500	p=4, q=1, n=1000
WVI-IC	0.051 (0.236)	0.056 (0.435)
LRT	0.055 (0.204)	0.054 (0.350)
	Empirical Size (power)	
	p=8, q=1, n=500	p=8, q=1, n=1000
WVI-IC	0.047 (0.450)	0.043 (0.768)
LRT	0.069 (0.341)	0.068 (0.575)

3.2.4 Empirical size under stringent p-value thresholds

In case-control studies, there is usually a huge amount of genetic variants are tested. To account for the multiple testing issue, Bonferroni correction is used and results in p -value threshold 5×10^{-3} or smaller. In this simulation, we investigated the performance of 1) WV-IC in the absence of genetic heterogeneity and 2) HWV-IC in the presence of genetic heterogeneity across observable subpopulations under stringent p -value thresholds. The corresponding simulation settings for significance level 0.5 were adopted, except that only scenario $p = 4, n = 1000$ was investigated and 150K Monte Carlo samples were generated.

The empirical size is calculated as the proportion of p -values smaller than the corresponding p -value threshold. Table 3.12 shows that the Type I error of both WV-IC and HWV-IC is close to the nominal level, indicating the capability of WV-IC and HWV-IC for large-scale GWAS.

Table 3.12 Empirical sizes of WV-IC and HWV-IC under stringent p -value thresholds and no left truncation.

Threshold	Empirical Size	
	WV-IC	HWV-IC
0.005	0.0046	0.0047
0.0005	0.00039	0.00053
0.00005	0.000044	0.000047

3.3 A Real Application

We applied the proposed WV-IC and HWV-IC tests to a GWAS Dental Caries: A Whole-Genome Association and Gene X Environment studies. The study includes 5,418 consented subjects ascertained through four study sites: PITT, IOWA, DRDR, and GEIRS. Each subject has 601,273 typed SNPs and 84 subject-level phenotypes such as age at the dental exam, gender, and number of DMFS (Decay, Missing, Filling, Sound) teeth/surfaces. Since the goal of this analysis is to detect genetic association with age-to-onset of ECC, in the absence/presence of genetic heterogeneity, we keep only subjects who were younger than 6 at the dental exam. Also, since each subject has one dental exam, the age-to-onset of ECC is current status data, one special case of interval censoring survival outcome.

To avoid systematic error, we performed SNP- and subject-level quality control on the genotype data using PLINK 1.9. SNPs were removed if $MAF > 0.01$, HWE test p -value $> 10^{-6}$, or missing rate $> 2\%$. Subjects were removed if the missing genotype rate was $> 2\%$. In

addition to quality control, we also removed subjects whose genotype or phenotype is missing. The resulting genotype data has 1,125 subjects and 553,194 SNPs. The genotype data still has $\sim 0.01\%$ missing values, which were then imputed by random samples generated from $Binomial(2, MAF)$, where MAF was estimated from the phenotype dataset. PCA was performed using PLINK 1.9 to obtain the first 10 principal components. The 553,194 SNPs were then grouped into 23,008 different genes based on human UCSC human genome annotation (NCBI36/hg18).

Each of the 23,008 genes was tested for genetic association and potential heterogeneity effect using WV-IC and HWV-IC adjust for the top 10 principal components, study site indicator (PITT as baseline), gender, and race. To account for the multiple testing issue, we adopted the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) to control the FDR at 10%. The p -value threshold for i -th most significant gene is obtained by $l_i = (i\alpha)\{m \sum_{k=1}^m (1/k)\}^{-1}$ ($i = 1, \dots, 23008$), where $\alpha = FDR$ and $m=23,008$. When in the presence of genetic heterogeneity, four sources were considered 1) across different races (White, Asian, Black American Indian, Bi- or multi-racial, and Other), 2) across two water fluoride levels, sufficient (>0.7 mg/L) and insufficient (≤ 0.7 mg/L), 3) across two sexes, and 4) across genetic background inferred by a random sample of 200,000 SNPs from the whole genome. Note that, because subjects ascertained from the DRDR study site have missing water fluoride levels, the sample size drops from 1,125 to 652 and the count of unique study sites drops from 4 to 3 when water fluoride level is included in the analysis. In all analysis scenarios, two kernels were used to measure genetic similarity: IBS and Cross-product, two special cases of the semiparametric transformation model were considered: the Cox proportional hazard model ($r=0$) and proportional odds model ($r=1$).

Table 3.13 shows the top five genes detected by WV-IC (scenario S1) and HWV-IC (scenarios S2-S5) and the corresponding p -values. Even though none of the 23,008 genes was shown to be statistically significantly associated with age-to-onset of ECC, the analysis did suggest that MPPED2 and TSPAN2 were frequently found to have the smallest p -value. MPPED2 exhibited suggestive evidence for association with ECC development in the first genome-wide association scan for dental caries in the study of 1,305 US children (Shaffer et al., 2011). The biological role of MPPED2 was later verified via a meta-analysis of five childhood samples (Staley et al., 2017). For TSPAN2, no strong evidence was found to show its biological role in ECC.

3.4 Discussion

We developed the first set of multi-marker tests for genetic associations and G-G/G-E interactions with interval-censored and possibly left-truncated survival outcomes. The new tests can adjust for covariates based on a semiparametric transformation model. The proposed HWV-IC can also account for genetic heterogeneity to increase the power of association testing.

Our methods are directly applicable to interval-censored competing risks data if the observation process is independent of the process of competing risks given the covariates. Specifically, by applying our methods to the reduced interval-censored competing risks data about the failure cause of interest (Hudgens, Li, and Fine, 2014), one can test the effect of a marker set or a G-G/G-E interaction on the cumulative incidence function of that cause.

It is worthwhile to extend the proposed tests to multivariate survival phenotypes. This extension has applications to genetic studies of chronic diseases that can occur at multiple sites in the human body, such as dental caries, diabetic retinopathy, and age-related macular

Table 3.13 Top five genes discovered by WV-IC and HWV-IC from a gene-based genome-wide association analysis of the DC-WGAGE dataset. PH and PO stand for the Cox proportional hazard model and the proportional odds model respectively. CP and IBS stand for cross-product kernel and IBS kernel respectively. Various types of heterogeneity were considered, including no genetic heterogeneity (S1), heterogeneity across races (S2), heterogeneity between the two home water fluoride levels (S3), heterogeneity between sexes (S4), and heterogeneity across genetic backgrounds (S5).

Scenario		Genes and p-values				
PO+CP	S1	MPPED2	OR2B11	NECAB3	E2F1	CYB5R2
		5.01E-05	7.61E-05	8.62E-05	8.62E-05	1.69E-04
	S2	MPPED2	HOXC13-AS	CYB5R2	NECAB3	E2F1
		3.21E-05	1.29E-04	1.37E-04	1.79E-04	1.79E-04
	S3	TSPAN2	CRCT1	HOXC13-AS	LCE5A	TSHB
		1.17E-05	2.23E-05	2.55E-05	2.73E-05	4.78E-05
	S4	MPPED2	OR2B11	NECAB3	E2F1	CDK5RAP1
		5.88E-05	9.91E-05	1.10E-04	1.11E-04	1.65E-04
	S5	MPPED2	OR2B11	NECAB3	E2F1	CYB5R2
		4.94E-05	7.84E-05	8.70E-05	8.70E-05	1.69E-04
PH+CP	S1	MPPED2	OR2B11	NECAB3	E2F1	CYB5R2
		8.23E-05	8.48E-05	8.90E-05	8.90E-05	9.26E-05
	S2	MPPED2	CYB5R2	LINC00511	HOXC13-AS	NCR3LG1
		5.75E-05	9.12E-05	1.69E-04	1.90E-04	2.08E-04
	S3	TSPAN2	HOXC13-AS	CRCT1	ECRG4	LCE5A
		1.27E-05	3.05E-05	4.53E-05	5.07E-05	5.50E-05
	S4	CYB5R2	MPPED2	TSPAN2	NCR3LG1	OR2B11
		2.16E-04	2.21E-04	2.49E-04	2.64E-04	2.73E-04
	S5	TSPAN2	MPPED2	CYB5R2	TPM3P9	NCR3LG1
		1.77E-04	1.93E-04	2.02E-04	2.07E-04	2.09E-04
PO+IBS	S1	LOC101927989	MPPED2	PAX9	NECAB3	E2F1
		4.68E-05	4.89E-05	7.04E-05	7.61E-05	7.61E-05
	S2	MPPED2	STARD5	LOC101927989	CCDC185	CYB5R2
		2.78E-05	3.56E-05	1.19E-04	1.31E-04	1.50E-04
	S3	TSPAN2	CRCT1	TSHB	LCE5A	HOXC13-AS
		5.84E-06	5.63E-05	6.04E-05	6.23E-05	6.48E-04
	S4	LOC101927989	CCDC185	MPPED2	PAX9	NECAB3
		5.03E-05	6.12E-05	6.46E-05	8.93E-05	9.72E-05
	S5	LOC101927989	MPPED2	PAX9	NECAB3	E2F1
		4.79E-05	4.85E-05	7.13E-05	7.79E-05	7.79E-05
PH+IBS	S1	CCDC185	MPPED2	NECAB3	E2F1	PAX9
		5.76E-05	6.89E-05	7.58E-05	7.58E-05	8.95E-05
	S2	STARD5	MPPED2	CCDC185	CYB5R2	PPM1E
		4.28E-05	4.36E-05	7.93E-05	1.07E-04	1.65E-04
	S3	TSPAN2	HOXC13-AS	TSHB	CRCT1	ECRG4
		6.21E-06	7.55E-05	7.83E-05	1.14E-04	1.25E-04
	S4	CCDC185	STARD5	TSPAN2	MPPED2	CYB5R2
		2.04E-05	1.02E-04	1.63E-04	2.01E-04	2.51E-04
	S5	CCDC185	STARD5	TSPAN2	MPPED2	CYB5R2
		2.79E-05	7.71E-05	1.12E-04	1.60E-04	2.14E-04
p-value threshold		4.09E-07	8.18E-07	1.23E-06	1.64E-06	2.05E-06

degeneration. Joint analysis of the failure times at the different sites could increase the power to detect the association of such a disease with genes. The extension hinges on finding an appropriate phenotype similarity for multivariate interval-censored survival endpoints.

CHAPTER 4 Multi-marker genetic association and interaction tests based on the accelerated failure time model

4.1 Methods

4.1.1 Association tests

Consider a cohort of n independent individuals who have not experienced any competing events at baseline. We assume that each individual is subject to multiple potential sources of failure event and the occurrence of one event impedes the occurrence of all other events. Let $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ be genetic covariates that is either a set of SNPs, gene expression values or a genetic pathway. We are interested in testing the genetic association with age-to-onset of Cause 1 (primary event) with or without considering genetic heterogeneity. Denote A as left truncation time (e.g., age at study registry), $\tilde{T} = \min(T, C)$ is observed survival time subject to right censoring, where T is survival time and C is right censoring time. $\Delta = I(T \leq C)$ and ϵ is cause of failure, $\epsilon = 1$ for primary cause, $\epsilon = 0$ for all other causes. Baseline covariates, $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_k)$ is adjusted to reduce confounding and/or increase power. When in the presence of genetic heterogeneity, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_d)$ is used to infer subpopulation structure. Observed data for n subjects is denoted as $D = (D_1, \dots, D_n)$, with $\{D_i = (A_i, \tilde{T}_i, \Delta_i, \Delta_i \epsilon_i, \mathbf{G}_i^T, \mathbf{Z}_i^T)\}_{i=1}^n$ and also includes \mathbf{X}_i in the presence of genetic heterogeneity. We assume the survival time T , the residual censoring time $C - A$ and the left truncation time A are conditionally independent given \mathbf{G} and \mathbf{Z} (and \mathbf{X} if considering genetic heterogeneity).

When in the absence of genetic heterogeneity, under the null hypothesis that \mathbf{G} is not associated with survival time T , we assume that, given adjustment covariate \mathbf{Z} , the survival

time of subject i ($i = 1, \dots, n$) follows AFT model with the following CSH for Cause 1,

$$\lambda(t|\mathbf{Z}_i) = \lambda_0(te^{-\boldsymbol{\beta}^T \mathbf{Z}_i})e^{-\boldsymbol{\beta}^T \mathbf{Z}_i},$$

where $\lambda_0(\cdot)$ is an unknown baseline CSH function. In real application, the regression parameter $\boldsymbol{\beta}$ and cumulative baseline hazard function $\Lambda_0(t)$ are estimated by applying rank-based estimation method (Chiou and Xu, 2017) to observed data D with the (working) AFT model $\log(T^*) = \boldsymbol{\beta}^T \mathbf{Z} + \epsilon$. The resulting estimators are $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0(\boldsymbol{\beta}, t)$. Then we have $\Lambda_\epsilon(t) = \int_{-\infty}^t \lambda_\epsilon(s)ds = \int_{-\infty}^t e^s \lambda_0(s)ds$, which can be estimated using Nelson-Aalon type estimator,

$$\hat{\Lambda}_\epsilon(\boldsymbol{\beta}, t) = \int_{-\infty}^t \frac{\sum_{i=1}^n dN_i(\boldsymbol{\beta}, s)}{\sum_{i=1}^n \nu_i(\boldsymbol{\beta}, s)Y_i(\boldsymbol{\beta}, s)},$$

where $N_i(\boldsymbol{\beta}, t) = I(e_i(\boldsymbol{\beta}) \leq t, \Delta_i \epsilon_i = 1)$, $Y_i(\boldsymbol{\beta}, t) = I(e_i(\boldsymbol{\beta}) \geq t)$, and $\nu_i(\boldsymbol{\beta}, t) = I(e_i^a(\boldsymbol{\beta}) < t)$, with $e_i(\boldsymbol{\beta}) = \log \tilde{T}_i - \boldsymbol{\beta}^T \mathbf{Z}_i$ and $e_i^a(\boldsymbol{\beta}) = \log A_i - \boldsymbol{\beta}^T \mathbf{Z}_i$. Define martingale residual to be $M_i = M_i(\Lambda_\epsilon, \boldsymbol{\beta}) = N_i(\boldsymbol{\beta}, \infty) - \int_{-\infty}^{\infty} \nu_i(\boldsymbol{\beta}, t)Y_i(\boldsymbol{\beta}, t)d\Lambda_\epsilon(t)$, $\hat{M}_i = M_i(\hat{\Lambda}_\epsilon(\boldsymbol{\beta}, t), \boldsymbol{\beta})$ and $\tilde{M}_i = M_i(\hat{\Lambda}_\epsilon(\hat{\boldsymbol{\beta}}, t), \hat{\boldsymbol{\beta}})$. The weighted V test statistics can be written as

$$\begin{aligned} V_{\mathbf{Z}} &= \tilde{\mathbf{M}}^T \mathbf{F} \tilde{\mathbf{M}} \\ &= \sum_{i=1}^n \sum_{j=1}^n f(\mathbf{G}_i, \mathbf{G}_j) \tilde{M}_i \tilde{M}_j, \end{aligned}$$

where $\tilde{\mathbf{M}} = (\tilde{M}_1, \dots, \tilde{M}_n)$ and $\mathbf{F} = \{f(\mathbf{G}_i, \mathbf{G}_j)\}_{n \times n}$. $\tilde{M}_i \tilde{M}_j$ is V statistic kernel that measures phenotype similarity weighted by $f(\mathbf{G}_i, \mathbf{G}_j)$, which measures genetic similarity, for a pair of subjects (i, j) . The value of weighted V statistic is the weighted sum over all pairs of (i, j) . Therefore, the larger (smaller) the phenotype similarity weighted by larger (smaller) genetic

similarity results in larger (smaller) $V_{\mathbf{Z}}$ value, therefore smaller (larger) p -value.

To get the asymptotic null distribution of $V_{\mathbf{Z}}$, we decompose $\mathbf{F} = E_1^T E_1$ for some $m \times n$ matrix E_1 , where $m(1 \leq m \leq n)$ is the rank of \mathbf{F} . We then have weighted V statistic in quadratic form as: $V_{\mathbf{Z}} = (\tilde{\mathbf{M}}^T E_1^T)(E_1 \tilde{\mathbf{M}})$, and the asymptotic null distribution can be approximated by $\sum_{j=1}^m \lambda_j \chi_{1j}^2$, which is linear combination of independent χ^2 variables of 1 degree of freedom weighted by $\{\lambda_i\}_{i=1}^m$, the eigenvalues of estimator of $Cov(E_1 \tilde{\mathbf{M}})$ (derivation shown in Appendix). Based on this distribution, the p -values can be computed using Davies' method (Davies, 1980), $P(V_{\mathbf{Z}} \geq V_{\mathbf{Z},obs})$, where $V_{\mathbf{Z},obs}$ is the observed value of $V_{\mathbf{Z}}$.

When in the presence of genetic heterogeneity inferred by d -dimensional covariate $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_d)$, the heterogeneity weighted V test statistic is an extension of 5.3 as follow,

$$V_{\mathbf{Z}}^H = \tilde{\mathbf{M}}^T \mathbf{W} \tilde{\mathbf{M}},$$

where $\mathbf{W} = (\mathbf{J} + \mathbf{H}) \odot \mathbf{F}$, $\mathbf{H} = \{h(\mathbf{X}_i, \mathbf{X}_j)\}_{n \times n}$, $h(\mathbf{X}_i, \mathbf{X}_j)$ is a kernel function that measures subpopulation similarity between subject i and j . \odot is element-wise matrix multiplication. The asymptotic null distribution of $V_{\mathbf{Z}}^H$ can be approximated similarly as that of $V_{\mathbf{Z}}$. Based on the distribution, p -value can be computed using Davies' method (Davies, 1980), $P(V_{\mathbf{Z}}^H \geq V_{\mathbf{Z},obs}^H)$, where $V_{\mathbf{Z},obs}^H$ is observed value of $V_{\mathbf{Z}}^H$.

4.1.2 G-G/G-E interaction test

We developed two interaction tests to detect the potential association between G - G and G - E interaction effect on age-to-onset of the primary event (Cause 1) considering left-truncation time.

In testing G - G interaction effect two genes $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ and $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_q)$.

Under the null hypothesis of no G - G interaction effect, the adjustment covariate \mathbf{Z} is replaced with (\mathbf{G}, \mathbf{H}) and the genetic main effect is homogeneous. We assume the cause specific hazard for subject i given \mathbf{G} and \mathbf{H} is,

$$\lambda(t|\mathbf{G}_i, \mathbf{H}_i) = \lambda_0(te^{-\beta^T \mathbf{G}_i - \alpha^T \mathbf{H}_i})e^{-\beta^T \mathbf{G}_i - \alpha^T \mathbf{H}_i}, \quad (4.1)$$

where $\lambda_0(\cdot)$ is unspecified baseline hazard function. In real application, the regression parameter β and cumulative baseline hazard function $\Lambda_0(t)$ are estimated by applying rank-based estimation method (Chiou and Xu, 2017) to observed data D with the (working) AFT model $\log(T^*) = \beta^T \mathbf{G} + \alpha^T \mathbf{H} + \epsilon$. The resulting estimators are $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\Lambda}_0(\beta, \alpha, t)$. Then we have $\Lambda_\epsilon(t) = \int_{-\infty}^t \lambda_\epsilon(s) ds = \int_{-\infty}^t e^s \lambda_0(s) ds$, which can be estimated using Nelson-Aalon type estimator,

$$\hat{\Lambda}_\epsilon(\beta, \alpha, t) = \int_{-\infty}^t \frac{\sum_{i=1}^n dN_i(\beta, \alpha, s)}{\sum_{i=1}^n \nu_i(\beta, \alpha, s) Y_i(\beta, \alpha, s)}, \quad (4.2)$$

where $N_i(\beta, \alpha, t) = I(e_i(\beta, \alpha) \leq t, \Delta_i \epsilon_i = 1)$, $Y_i(\beta, \alpha, t) = I(e_i(\beta, \alpha) \geq t)$, and $\nu_i(\beta, \alpha, t) = I(e_i^a(\beta, \alpha) < t)$, with $e_i(\beta, \alpha) = \log \tilde{T}_i - \beta^T \mathbf{G}_i - \alpha^T \mathbf{H}_i$ and $e_i^a(\beta, \alpha) = \log A_i - \beta^T \mathbf{G}_i - \alpha^T \mathbf{H}_i$. Let $M_i = M_i(\Lambda_\epsilon, \beta, \alpha) = N_i(\beta, \alpha, \infty) - \int_{-\infty}^{\infty} \nu_i(\beta, \alpha, t) Y_i(\beta, \alpha, t) d\Lambda_\epsilon(t)$ be martingale residual, $\hat{M}_i = M_i(\hat{\Lambda}_\epsilon(\beta, \alpha, t), \beta, \alpha)$ and $\tilde{M}_i = M_i(\hat{\Lambda}_\epsilon(\hat{\beta}, \hat{\alpha}, t), \hat{\beta}, \hat{\alpha})$. The weighted V test statistics can be written as

$$\begin{aligned} V_I &= \tilde{\mathbf{M}}^T (\mathbf{F} \odot \mathbf{K}) \tilde{\mathbf{M}} \\ &= \sum_{i=1}^n \sum_{j=1}^n f(\mathbf{G}_i, \mathbf{G}_j) k(\mathbf{H}_i, \mathbf{H}_j) \tilde{M}_i \tilde{M}_j, \end{aligned} \quad (4.3)$$

where $\tilde{\mathbf{M}} = (\tilde{M}_1, \dots, \tilde{M}_n)$, $\mathbf{F} = \{f(\mathbf{G}_i, \mathbf{G}_j)\}_{n \times n}$, and $\mathbf{K} = \{k(\mathbf{H}_i, \mathbf{H}_j)\}_{n \times n}$. $\tilde{M}_i \tilde{M}_j$ in 4.3 is

the V statistic kernel that measures phenotype similarity weighted by $f(\mathbf{G}_i, \mathbf{G}_j)k(\mathbf{H}_i, \mathbf{H}_j)$, which measures genetic similarity, for a pair of subjects (i, j) . The value of weighted V statistic is the weighted sum over all pairs of (i, j) . Therefore, the larger (smaller) the phenotype similarity weighted by larger (smaller) genetic similarity results in larger (smaller) $V_{\mathbf{Z}}$ value, therefore smaller (larger) p -value.

The asymptotic null distribution of V_I can be approximated similarly as $V_{\mathbf{Z}}$. Based on the distribution, we use Davies' method (Davies, 1980) to compute p -value, $P(V_I \geq V_{I,obs})$, where $V_{I,obs}$ is the observed value of V_I .

For testing G - E interaction effect, the weighted V test statistic and its asymptotic null distribution can be obtained in the same way, except that \mathbf{H} is replaced by an environmental variable (e.g., gender). The proposed interaction tests are expected to be more powerful than regular tests (e.g., LRT, Wald test) when there exists strong LD because the weighted V statistic has an effective degree of freedom lower than that of regular tests.

4.1.3 Small sample adjustment

The association and interaction tests introduced above work efficiently when the sample size is large. However, as shown in the QQ plot in the simulation below, the distribution might not be accurate when n is relatively small (as compared to p), which is frequently encountered in whole-exome sequencing studies (Lee et al., 2012; Emond et al., 2012). Therefore, motivated by (Chen et al., 2016), we propose a small sample correction on $V_{\mathbf{Z}}$, $V_{\mathbf{Z}}^H$ and V_I , denoted, respectively, as $V_{\mathbf{Z}}^c$, $V_{\mathbf{Z}}^{H,c}$ and V_I^c as follow,

$$V_{\mathbf{Z}}^c = \frac{\tilde{\mathbf{M}}^T \mathbf{F} \tilde{\mathbf{M}}}{\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}}, \quad V_{\mathbf{Z}}^{H,c} = \frac{\tilde{\mathbf{M}}^T \mathbf{W} \tilde{\mathbf{M}}}{\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}}, \quad \text{and} \quad V_I^c = \frac{\tilde{\mathbf{M}}^T (\mathbf{F} \odot \mathbf{K}) \tilde{\mathbf{M}}}{\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}},$$

where $\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}$ accounts for the variability of martingale residual. After accounting for that, we expect that the approximated asymptotic null distribution of corrected weighted V statistics, $V_{\mathbf{Z}}^c$, $V_{\mathbf{Z}}^{H,c}$, and V_I^c have exact mixture of χ^2 distribution. For $V_{\mathbf{Z}}^c$, based on the distribution, we can apply Davies' method (Davies, 1980) to compute p -values by $P(V_{\mathbf{Z}}^c \geq V_{\mathbf{Z},obs}^c) = P(\frac{\tilde{\mathbf{M}}^T \mathbf{F} \tilde{\mathbf{M}}}{\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}} \geq V_{\mathbf{Z},obs}^c) = P(\tilde{\mathbf{M}}^T (\mathbf{F} - V_{\mathbf{Z},obs}^c \mathbf{I}) \tilde{\mathbf{M}} \geq 0)$. The p -value of $V_{\mathbf{Z}}^{H,c}$ and V_I^c can be computed similarly.

4.2 Simulations

We performed a Monte Carlo simulation to assess the finite-sample performance of (heterogeneity) weighted V tests on survival time following the AFT model, considering adjustment covariates, left-truncation time, and each individual potentially experiencing two competing events (Cause 1 and Cause 2). In all simulations, unless otherwise specified, two sample sizes $n = 400, 500$ and two SNP set sizes $p = 3, 5$ were considered. Genetic covariates, denoted as $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$, is a set of p SNPs (coded by 0, 1, or 2), unless otherwise specified. The generation of \mathbf{G} follows a two-step procedure that considered potential LD structure, 1) sample n vectors independently from $MVN(\mathbf{0}, \Sigma)$, where $\Sigma = \{0.5^{|k-l|}\}_{p \times p}$, the kl -th element is the covariance between \mathbf{G}_k and \mathbf{G}_l ; 2) categorize each quantitative column into SNP with three levels 0, 1, and 2 by using predetermined quantile cutoffs: a^2 , $(1-a)^2$, and $2a(1-a)$ of standard normal distribution with $a \sim Beta(2, 5)$, to ensure the resulting population is in Hardy-Weinberg equilibrium (HWE) and MAFs follow $Beta(2, 5)$ distribution. Two baseline covariates were adjusted, one binary $Z_1 \sim Binomial(0.5)$ and one continuous $Z_2 \sim Uniform(0, 2)$. Unless otherwise specified, we assumed that there is no confounding effect, in other words, \mathbf{Z} is independent of \mathbf{G} . The failure of all competing events was assumed to follow the AFT model, and the observed survival times were generated following

steps in section 3.2 of Beyersmann et al. (Beyersmann et al., 2009). Left truncation times were generated from $Uniform(0, 1)$ and the residual censoring times from $Exponential(0.3)$ to control the right censoring rate around 35%. In all simulations, 1000 Monte Carlo samples were generated and the significance level of a test was set at 0.05 unless otherwise specified in testing genetic association under the scenarios where stringent p -value thresholds were considered.

4.2.1 Association test in the absence of genetic heterogeneity

In this series of simulation, we investigated the empirical size and power of $V_{\mathbf{Z}}$ in the absence of genetic heterogeneity and in the presence of left truncation time. Left truncation time was generated from $Uniform(0, 1)$. For subject i , both competing events were assumed to follow AFT model with following cause specific hazard (CSH) function,

$$\lambda(t|\mathbf{G}_i, \mathbf{Z}_i) = \lambda_0(t \cdot \exp\{-\sum_{j=1}^p \beta_{1j}G_{ij} - \sum_{l=1}^2 \beta_{2l}Z_{il}\}) \exp\{-\sum_{j=1}^p \beta_{1j}G_{ij} - \sum_{l=1}^2 \beta_{2l}Z_{il}\},$$

where $\lambda_0(x) = x^2 + x$ is baseline hazard function.

Empirical size and power under various n 's and p 's

In this simulation, we assessed the empirical size and power of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$. Cause specific hazard (CSH) of Cause 1 and Cause 2 competing events for subject i are as follows,

$$\left\{ \begin{array}{l} \lambda_1(t|\mathbf{Z}_i, \mathbf{G}_i) = \lambda_0(t \cdot \exp\{-\sum_{j=1}^p \beta_j G_{ij} - \sum_{l=1}^2 0.1 Z_{il}\}) \exp\{-\sum_{j=1}^p \beta_j G_{ij} - \sum_{l=1}^2 0.1 Z_{il}\}, \\ \lambda_2(t|\mathbf{Z}_i, \mathbf{G}_i) = \lambda_0(t \cdot \exp\{-\sum_{j=1}^p 0.2 G_{ij} - \sum_{k=1}^2 0.2 Z_{ik}\}) \exp\{-\sum_{j=1}^p 0.2 G_{ij} - \sum_{k=1}^2 0.2 Z_{ik}\}, \end{array} \right.$$

where λ_1 is CSH of Cause 1 (primary event). We set $\beta_1 = \dots = \beta_p$ to 0 and 0.1 for empirical size and power assessment, while CSH for Cause 2 stays the same. IBS kernel was used to measure genetic similarity in both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$. Gaussian kernel was used to measure similarity of adjustment covariates \mathbf{Z} for test $V_{\mathbf{Z}}^H$. Table 4.1 shows that the Type I error of both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are close to nominal level and their power are comparable under various n 's and p 's. This indicates that when the genetic effect is homogeneous, $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ have similar performance.

Table 4.1 Empirical size and power comparison of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under covariate adjustment and left truncation.

	Empirical Size (Power)	
	p=3, n=400	p=3, n=500
$V_{\mathbf{Z}}$	0.051 (0.688)	0.054 (0.805)
$V_{\mathbf{Z}}^H$	0.050 (0.690)	0.054 (0.798)
	p=5, n=400	p=5, n=500
$V_{\mathbf{Z}}$	0.052 (0.865)	0.048 (0.945)
$V_{\mathbf{Z}}^H$	0.045 (0.849)	0.047 (0.940)

Empirical size and power under quadratic confounding

In this simulation, we assessed the empirical size and power of $V_{\mathbf{Z}}$ in the absence of genetic heterogeneity. We assumed that \mathbf{G} is a quadratic function of \mathbf{Z} , in other words, there is quadratic confounding effect. The genetic covariate for subject i was replaced by gene expression values generated by $\mathbf{G}_i = 0.5Z_{i1} + 0.5Z_{i2} + 0.25Z_{i1}^2 + 0.25Z_{i2}^2 + 0.5Z_{i1}Z_{i2} + \mathbf{e}_i$ ($i = 1, \dots, n$). The random error $\mathbf{e}_i \sim MVN(\mathbf{0}, \Sigma)$, where $\Sigma = \{0.5^{|k-l|}\}_{p \times p}$. CSH for Cause

1 and Cause 2 competing event of subject i are as follow

$$\left\{ \begin{array}{l} \lambda_1(t|\mathbf{Z}_i, \mathbf{G}_i) = \lambda_0(t \cdot \exp\{-\sum_{j=1}^p \beta_j G_{ij} - \sum_{l=1}^2 0.1 Z_{il}\}) \\ \exp\{-\sum_{j=1}^p \beta_j G_{ij} - \sum_{l=1}^2 0.1 Z_{il}\}, \\ \lambda_2(t|\mathbf{Z}_i, \mathbf{G}_i) = \lambda_0(t \cdot \exp\{-\sum_{j=1}^p 0.1 G_{ij} - \sum_{k=1}^2 0.2 Z_{ik}\}) \\ \exp\{-\sum_{j=1}^p 0.1 G_{ij} - \sum_{k=1}^2 0.2 Z_{ik}\}, \end{array} \right.$$

where λ_1 is CSH for primary event (Cause 1), $\beta_1 = \dots = \beta_p$ are set to 0 and 0.03 for empirical size and power assessment, while CSH for Cause 2 stays the same. Gaussian kernel was used to measure gene expression similarity. Figure 4.1 shows that p -value of $V_{\mathbf{Z}}$ is uniformly distributed and table 4.2 shows Type I error of $V_{\mathbf{Z}}$ is close to nominal level and power increases with sample size. The results show that $V_{\mathbf{Z}}$ is able to handle quadratic confounding effect.

Table 4.2 Empirical size and power of $V_{\mathbf{Z}}$ in testing genetic effects under quadratic confounding and left truncation.

	Empirical Size (Power)	
	p=3, n=400	p=3, n=500
$V_{\mathbf{Z}}$	0.050 (0.178)	0.049 (0.220)
	p=5, n=400	p=5, n=500
$V_{\mathbf{Z}}$	0.047 (0.279)	0.057 (0.325)

4.2.2 Association test in the presence of genetic heterogeneity

In this series of simulations, we investigated the performance of association test statistic, $V_{\mathbf{Z}}^H$, in the presence of genetic heterogeneity across 1) observable subpopulations, 2) two latent subpopulations, 3) twenty subpopulations, and 4) individual genome profile. The

advantage of $V_{\mathbf{Z}}^H$ over $V_{\mathbf{Z}}$ is obvious through the performance comparison shown below. For all simulations in this section, we assume that the effect of \mathbf{G} exists on survival time of both competing events, but the form of effect is heterogeneous for Cause 1 (primary event) and homogeneous for Cause 2.

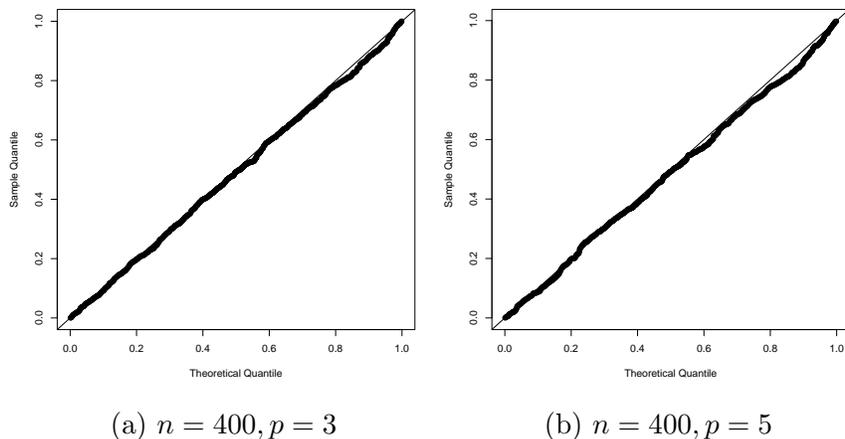


Figure 4.1 Q-Q plot of the null p-value of $V_{\mathbf{Z}}$ under quadratic confounding with $n = 400$ and $p = 3$ or 5 .

Genetic heterogeneity across observable subpopulation

In this simulation, we assessed the empirical size and power of heterogeneity weighted V statistic, $V_{\mathbf{Z}}^H$, in the presence of heterogeneity across two observable subpopulations. We used a one-dim vector to infer two equally weighted subpopulations, $\mathbf{X} = \{X_i\}_{i=1}^n$, where $X_i = \{0, 1\}$. Therefore, two genetic covariates (each has sample size $n/2$) were generated using two different sets of MAFs ($\sim \text{Beta}(2, 5)$) and LD structure ($\Sigma_{p \times p} = \{\rho^{|k-l|}\}$ depends on ρ) as follow,

MAFs = $\{0.144, 0.046, 0.042, 0.013, 0.166\}$ and $\rho = 0.5$ in Subpopulation 1,

MAFs = $\{0.095, 0.038, 0.450, 0.125, 0.265\}$ and $\rho = 0.3$ in Subpopulation 2.

The first p MAFs will be used. We used a one-dimensional vector $\mathbf{X} = (X_1, \dots, X_n) \sim \text{Binomial}(0.5)$ to infer two observable subpopulations (e.g., male and female). The CSH of Cause 1 and Cause 2 for subject i are as follow,

$$\begin{cases} \lambda_1(t|\mathbf{Z}_i, \mathbf{G}_i) = \lambda_0(t \cdot e^{-\sum_{j=1}^p (\beta_0 + \beta_1 X_i) G_{ij} - 0.1 X_i}) e^{-\sum_{j=1}^p (\beta_0 + \beta_1 X_i) G_{ij} - 0.1 X_i} \\ \lambda_2(t|\mathbf{Z}_i, \mathbf{G}_i) = \lambda_0(t \cdot e^{-\sum_{j=1}^p 0.02 G_{ij} - \sum_{k=1}^2 0.2 Z_{ik}}) e^{-\sum_{j=1}^p 0.02 G_{ij} - \sum_{k=1}^2 0.2 Z_{ik}}, \end{cases}$$

where $\beta_0 = \beta_1 = 0$ for empirical size assessment and $(\beta_0, \beta_1) = \{(0.002, 0.1), (0.002, 0.2)\}$ for power assessment under two different genetic heterogeneity sizes, measured by β_1 . IBS kernel was used to measure genetic similarity and identity kernel used to measure subpopulation similarity. Performance of weighted V statistic, $V_{\mathbf{Z}}$, was also included. Table 4.3 shows that Type I error of both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are close to nominal level. Table 4.4 shows that $V_{\mathbf{Z}}^H$ is more powerful than $V_{\mathbf{Z}}$. For a fixed mean effect size (β_0), as heterogeneity size (β_1) increases, $V_{\mathbf{Z}}^H$ gains more power by taking advantage of the subpopulation structure in performing the test.

Table 4.3 Empirical sizes of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ in testing genetic association under genetic heterogeneity across two observable subpopulations and left truncation.

	Empirical Size	
	p=3, n=400	p=3, n=500
$V_{\mathbf{Z}}^H$	0.054	0.058
$V_{\mathbf{Z}}$	0.054	0.057
	p=5, n=400	p=5, n=500
$V_{\mathbf{Z}}^H$	0.046	0.049
$V_{\mathbf{Z}}$	0.045	0.051

Genetic heterogeneity across two latent subpopulations

In this simulation, we assessed the empirical size and power of heterogeneity weighted V statistic in the presence of genetic heterogeneity across two latent subpopulations. A one-

Table 4.4 Powers of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ in testing genetic association under genetic heterogeneity across two observable subpopulations and left truncation.

	Power			
	p=3, n=400		p=3, n=500	
β_1	0.1	0.2	0.1	0.2
$V_{\mathbf{Z}}^H$	0.226	0.646	0.261	0.749
$V_{\mathbf{Z}}$	0.185	0.543	0.234	0.645
	p=5, n=400		p=5, n=500	
	0.1	0.2	0.1	0.2
$V_{\mathbf{Z}}^H$	0.289	0.733	0.368	0.860
$V_{\mathbf{Z}}$	0.259	0.604	0.303	0.711

dim vector $\mathbf{X} = (X_1, \dots, X_n)$ was generated to infer two latent subpopulation structure. $X_i = a_i + 1 + e_i$ ($i = 1, \dots, n$), where $a_i \sim \text{Binomial}(0.5)$ and $e_i \sim \text{Normal}(0, 0.5)$. The CSH of Cause 1 and Cause 2 competing event for subject i in subpopulation j are as follows

$$\left\{ \begin{array}{l} \lambda_1(t|\mathbf{Z}_i, \mathbf{G}_{ij}) = \lambda_0(t \cdot \exp\{-\sum_{k=1}^p G_{ijk}\beta_{jk} - 0.1Z_{i1} - 0.1Z_{i2}\}) \cdot \\ \exp\{-\sum_{k=1}^p G_{ijk}\beta_{jk} - 0.1Z_{i1} - 0.1Z_{i2}\} \\ \lambda_2(t|\mathbf{Z}_i, \mathbf{G}_{ij}) = \lambda_0(t \cdot \exp\{-\sum_{k=1}^p 0.02G_{ijk} - 0.2Z_{i1} - 0.2Z_{i2}\}) \cdot \\ \exp\{-\sum_{k=1}^p 0.02G_{ijk} - 0.2Z_{i1} - 0.2Z_{i2}\}, \end{array} \right. \quad (4.4)$$

where $\{\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp}), s.t. \beta_{j1} = \dots = \beta_{jp}\}_{j=1}^2$ are respectively genetic effects of p SNPs in two latent subpopulations. $\{\boldsymbol{\beta}_j\}_{j=1}^2$ were set to 0 for empirical size assessment and set to different values to indicate different genetic heterogeneity scenarios for power assessment. IBS kernel was used to measure genetic similarity and Gaussian kernel for subpopulation similarity. Table 4.5 shows that Type I error of both $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ are close to nominal level. Table 4.6 shows that when in the absence of genetic heterogeneity (T1), $V_{\mathbf{Z}}$ is more powerful,

however, when in the presence of genetic heterogeneity, $V_{\mathbf{Z}}^H$ is more powerful and gains more power as heterogeneity size increases.

Table 4.5 Empirical sizes of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across two latent subpopulations, covariate adjustment and left truncation.

	Empirical Size	
	p=3, n=400	p=3, n=500
$V_{\mathbf{Z}}^H$	0.047	0.050
$V_{\mathbf{Z}}$	0.052	0.058
	p=5, n=400	p=5, n=500
$V_{\mathbf{Z}}^H$	0.053	0.044
$V_{\mathbf{Z}}$	0.045	0.049

Table 4.6 Powers of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across two latent subpopulations, covariate adjustment and left truncation. Various heterogeneity scenarios were considered, determined by the values of β_{1k} and β_{2k} , including the same effect size and the same effect direction (T1), identical sizes but opposite directions (T2), no effect in one subpopulation while positive effect in the other (T3), and different sizes but the same direction (T4).

		Heterogeneity Scenario							
		T1		T2		T3		T4	
	β_{1k}	0.04	0.08	-0.05	-0.1	0	0	0.03	0.03
	β_{2k}	0.04	0.08	0.05	0.1	0.08	0.12	0.08	0.1
p=3,n=400	$V_{\mathbf{Z}}^H$	0.252	0.768	0.775	0.972	0.719	0.906	0.601	0.784
	$V_{\mathbf{Z}}$	0.376	0.891	0.056	0.043	0.370	0.636	0.596	0.746
p=3,n=500	$V_{\mathbf{Z}}^H$	0.327	0.825	0.826	0.979	0.768	0.945	0.703	0.868
	$V_{\mathbf{Z}}$	0.376	0.927	0.057	0.047	0.467	0.716	0.702	0.817
p=5,n=400	$V_{\mathbf{Z}}^H$	0.403	0.900	0.984	1.000	0.950	0.994	0.855	0.958
	$V_{\mathbf{Z}}$	0.594	0.976	0.062	0.069	0.557	0.775	0.829	0.894
p=5,n=500	$V_{\mathbf{Z}}^H$	0.467	0.968	0.988	1.000	0.968	0.998	0.935	0.979
	$V_{\mathbf{Z}}$	0.729	0.995	0.047	0.068	0.659	0.866	0.905	0.958

Genetic heterogeneity across twenty latent subpopulation

In this simulation, we assessed empirical size and power of heterogeneity weighted V statistic, $V_{\mathbf{Z}}^H$, in the presence of twenty latent subpopulations. The simulation setting is same to

simulation above, except that we increased the number of latent subpopulation to twenty. We used $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{25})$ to infer subpopulation structure. $\mathbf{X}_k = \mathbf{a}_k + \mathbf{e}_k$ ($k = 1, \dots, 25$) where \mathbf{a}_k is a length n bootstrap sample from $\{1, \dots, 20\}$ that represents subgroup assignments. Random error $\mathbf{e}_k = (e_{k1}, \dots, e_{kn}) \sim Normal(0, 0.5)$. Regression coefficients for j subpopulations, $\{\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp}), s.t. \beta_{j1} = \dots = \beta_{jp}\}_{j=1}^{20}$, in CSH function 4.4 were set to 0 for empirical size assessment. In power assessment, $\{\boldsymbol{\beta}_j\}_{j=1}^{20}$ were sampled from uniform distribution with mean μ_β and variance σ_β^2 . IBS kernel was used to measure genetic similarity and Gaussian kernel for subpopulation similarity. Table 4.7 shows that Type I error of both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are close to nominal level. Table 4.8 shows that $V_{\mathbf{Z}}^H$ is more powerful than $V_{\mathbf{Z}}$ and gains power as heterogeneity size (σ_β) increases.

Table 4.7 Empirical sizes of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across twenty latent subpopulations, covariate adjustment and left truncation.

	Empirical Size	
	p=3, n=400	p=3, n=500
$V_{\mathbf{Z}}^H$	0.045	0.043
$V_{\mathbf{Z}}$	0.046	0.042
	p=5, n=400	p=5, n=500
$V_{\mathbf{Z}}^H$	0.043	0.044
$V_{\mathbf{Z}}$	0.042	0.048

Genetic heterogeneity across individual genome profile

In this simulation, we investigated the empirical size and power of heterogeneity weighted V statistic, $V_{\mathbf{Z}}^H$, in the presence of genetic heterogeneity across individual genome profile. In stead of considering two or twenty 'categorical' subpopulation structure, we now assume a 'continuous' subpopulation structure. In other words, each individual itself is a single 'subpopulation'. We used a genome profile of 1,000 SNPs, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{1000})$, to infer

Table 4.8 Power comparison of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ under genetic heterogeneity across twenty latent subpopulations, covariate adjustment and left truncation.

	Power			
	p=3, n=400		p=3, n=500	
	(0.02, 0.04)	(0.02, 0.08)	(0.02, 0.04)	(0.02, 0.08)
$(\mu_{\beta}, \sigma_{\beta})$				
$V_{\mathbf{Z}}^H$	0.396	0.742	0.448	0.843
$V_{\mathbf{Z}}$	0.336	0.506	0.390	0.616
	p=5, n=400		p=5, n=500	
	(0.02, 0.04)	(0.02, 0.08)	(0.02, 0.04)	(0.02, 0.08)
$(\mu_{\beta}, \sigma_{\beta})$				
$V_{\mathbf{Z}}^H$	0.679	0.955	0.785	0.988
$V_{\mathbf{Z}}$	0.473	0.634	0.581	0.754

subpopulation structure. \mathbf{X} was generated in three steps, 1) generate genetic effect sizes in each subgroup $\{\beta_i = (\beta_{i1}, \dots, \beta_{ip}), s.t. \beta_{i1} = \dots = \beta_{ip}\}_{i=1}^n$, 2) generate a n -dim vector \mathbf{X}_d ($d = 1, \dots, 1000$) from $MVN(\mathbf{0}, \Sigma)$, where Σ is a $n \times n$ identity matrix under null hypothesis and $\{\exp(-|\beta_{i1} - \beta_{j1}|/\sigma_{\beta})\}_{n \times n}$ under alternative hypothesis, 3) categorize \mathbf{X}_d ($d = 1, \dots, 1000$) into SNPs coded by 0, 1, or 2 by using predetermined quantile cutoffs: a^2 , $2a(1 - a)$, and $(1 - a)^2$ of a standard normal distribution where a is MAF generated from $Beta(2, 5)$, to ensure the resulting population is in HWE. IBS kernel was used to measure both genetic and subpopulation similarity. Table 4.9 shows that Type I error of both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are close to nominal level 0.05. Table 4.10 shows that $V_{\mathbf{Z}}^H$ is more powerful than $V_{\mathbf{Z}}$. Furthermore, for fixed mean effect size ($\mu_{\beta} = 0.03$), as heterogeneity size (σ_{β}) increases from 0.02 to 0.04, $V_{\mathbf{Z}}^H$ gains substantially more power while the power of $V_{\mathbf{Z}}$ does not change much as heterogeneity size increases.

4.2.3 Testing G-G/G-E interaction effect

In this simulation, we investigated the performance of weighted V statistic, V_I , in testing interaction effect. The performance was also compared with Wald test. For testing $G-G$

Table 4.9 Empirical sizes of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ in testing genetic effects under genetic heterogeneity across individual genome profiles, covariate adjustment and left truncation.

	Empirical Size	
	p=3, n=400	p=3, n=500
$V_{\mathbf{Z}}^H$	0.052	0.058
$V_{\mathbf{Z}}$	0.053	0.058
	p=5, n=400	p=5, n=500
$V_{\mathbf{Z}}^H$	0.046	0.049
$V_{\mathbf{Z}}$	0.047	0.051

Table 4.10 Powers of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ under genetic heterogeneity across individual genome profiles, covariate adjustment and left truncation.

	Power			
	p=3, n=400		p=3, n=500	
$(\mu_\beta, \sigma_\beta)$	(0.03, 0.02)	(0.03, 0.04)	(0.03, 0.02)	(0.03, 0.04)
$V_{\mathbf{Z}}^H$	0.253	0.328	0.308	0.400
$V_{\mathbf{Z}}$	0.241	0.242	0.290	0.292
	p=5, n=400		p=5, n=500	
$(\mu_\beta, \sigma_\beta)$	(0.03, 0.02)	(0.03, 0.04)	(0.03, 0.02)	(0.03, 0.04)
$V_{\mathbf{Z}}^H$	0.449	0.649	0.550	0.756
$V_{\mathbf{Z}}$	0.390	0.381	0.480	0.477

interaction effect between two genes $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ and $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_q)$. We assumed that two genes come from the same population, therefore have the same MAF and LD structures. CSH of Cause 1 and Cause 2 competing events for subject i are as follows

$$\left\{ \begin{array}{l} \lambda_1(t|\mathbf{Z}_{ij}, \mathbf{G}_{ij}) = \lambda_0(t \cdot \exp\{-\sum_{j=1}^p 0.2G_{ij} - \sum_{j=1}^q 0.2H_{ij} - \sum_{m=1}^{pq} \beta_m(GH)_m\}) \cdot \\ \exp\{-\sum_{j=1}^p 0.2G_{ij} - \sum_{j=1}^q 0.2H_{ij} - \sum_{m=1}^{pq} \beta_m(GH)_m\} \\ \lambda_2(t|\mathbf{Z}_{ij}, \mathbf{G}_{ij}) = \lambda_0(t \cdot \exp\{-\sum_{j=1}^p 0.2G_{ij} - \sum_{j=1}^q 0.2H_{ij}\}) \cdot \\ \exp\{-\sum_{j=1}^p 0.2G_{ij} - \sum_{j=1}^q 0.2H_{ij}\}, \end{array} \right.$$

where λ_1 is CSH for Cause 1 competing event (primary event), $(GH)_m$ is m -th column of interaction between \mathbf{G} and \mathbf{H} . Two SNP-set sizes ($p = 1, q = 2$) and ($p = 2, q = 2$) and two sample sizes $n = 600, 1000$ were used. We set $\beta_m = 0$ under null hypothesis of no G - G interaction effect to Cause 1 competing event and $\beta=1$ under alternative hypothesis, when G - G interaction effect exists only in Cause 1 (primary event). Cross-product kernel was used to measure genetic similarity for both genes.

For testing G - E interaction effect, the simulation setting is the same as above except that \mathbf{H} is replaced by a one-dim vector binary environmental variable generated from $Binomial(0.5)$. Two SNP-set sizes $p = 2, 3$ and two sample sizes $n = 500, 900$ were used. Identity kernel was used to measure environmental variable similarity.

Table 4.11 and 4.12 show that Type I error of V_I is close to nominal level while Wald test is slightly inflated under small sample size (e.g., under $p = 2, q = 2, n = 600$). In all scenarios, V_I is more powerful than the Wald test as expected, since the asymptotic null distribution of V_I has a more effective degree of freedom.

Table 4.11 Empirical sizes and powers of V_I and the Wald test in testing G-G interaction under left truncation.

	Empirical Size (Power)	
	p=1, q=2, n=600	p=1, q=2, n=1000
V_I	0.054 (0.217)	0.059 (0.289)
Wald	0.048 (0.139)	0.059 (0.206)
	p=2, q=2, n=600	p=2, q=2, n=1000
V_I	0.065 (0.408)	0.060 (0.550)
Wald	0.075 (0.202)	0.059 (0.375)

Table 4.12 Empirical sizes and powers of V_I and the Wald test in testing G-E interaction under left truncation.

	Empirical Size (Power)	
	p=2, q=1, n=500	p=2, q=1, n=900
V_I	0.052 (0.329)	0.055 (0.561)
Wald	0.047 (0.276)	0.054 (0.507)
	p=3, q=1, n=500	p=3, q=1, n=900
V_I	0.061 (0.381)	0.053 (0.617)
Wald	0.085 (0.342)	0.045 (0.529)

4.2.4 Empirical size under stringent p-value thresholds

In case-control studies, there is usually a huge amount of genetic variants are tested. To account for the multiple testing issue, Bonferroni correction is used and results in p -value threshold 5×10^{-3} or smaller. In this simulation, we investigated the performance of 1) $V_{\mathbf{Z}}$ in the absence of genetic heterogeneity and 2) $V_{\mathbf{Z}}^H$ in the presence of genetic heterogeneity across observable subpopulations under stringent p -value thresholds. The corresponding simulation settings for significance level 0.05 were adopted, except that only $p = 3, n = 500$ was considered and 500K Monte Carlo samples were generated. The Type I errors were the proportion of p -values that are smaller than the corresponding p -value thresholds. Table 4.13 shows that Type I error of both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are close to nominal level, indicating the capability of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ for large-scale GWAS.

4.2.5 Small-sample correction

In this series of simulations, we investigated the performance of small sample adjusted (heterogeneity) weighted V statistic in the following three scenarios: 1) $V_{\mathbf{Z}}$ in the absence of genetic heterogeneity, 2) $V_{\mathbf{Z}}^H$ in the presence of genetic heterogeneity across observable subpopulations, and 3) V_I in the presence of G - G and G - E interaction effect.

Table 4.13 Empirical sizes of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ under stringent p -value thresholds and left truncation.

Threshold	Empirical Size	
	$V_{\mathbf{Z}}$	$V_{\mathbf{Z}}^H$
0.05	0.051	0.049
0.005	0.0052	0.0047
0.0005	0.00054	0.00048
0.00005	0.000060	0.000054

In the absence of genetic heterogeneity under small sample size

In this simulation, we assessed the empirical size and power of $V_{\mathbf{Z}}^c$ and $V_{\mathbf{Z}}$ in testing genetic association in the absence of genetic heterogeneity. The simulation settings were the same as that of the simulations shown above, except that we used sample size $n = 100$ and SNP-set size $p = 15$. Table 4.14 and Figure 4.2 show that small sample corrected weighted V statistic, $V_{\mathbf{Z}}^c$, is asymptotically correct without loss of power, while unadjusted weighted V statistic, $V_{\mathbf{Z}}$, is not.

Table 4.14 Empirical size and power of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^c$ considering left truncation, with $n = 100$, $p = 15$.

	Empirical Size (power)
$V_{\mathbf{Z}}^c$	0.045 (0.417)
$V_{\mathbf{Z}}$	0.034 (0.346)

In the presence of genetic heterogeneity across observable subpopulations under a small sample size

In this simulation, we assessed the empirical size and power of $V_{\mathbf{Z}}^{H,c}$ and $V_{\mathbf{Z}}^H$ in testing genetic association in the presence of genetic heterogeneity across observable subpopulation. The simulation settings were the same as that of simulations shown above, except that we used sample size $n = 200$ and SNP-set size $p = 10$. In power assessment, two genetic

heterogeneity sizes were considered: $\beta_1 = 0.02, 0.04$. Table 4.15 and figure 4.3 show that small sample corrected heterogeneity weighted V statistic, $V_{\mathbf{Z}}^{H,c}$, is asymptotically correct without loss of power, while $V_{\mathbf{Z}}^H$ is not.

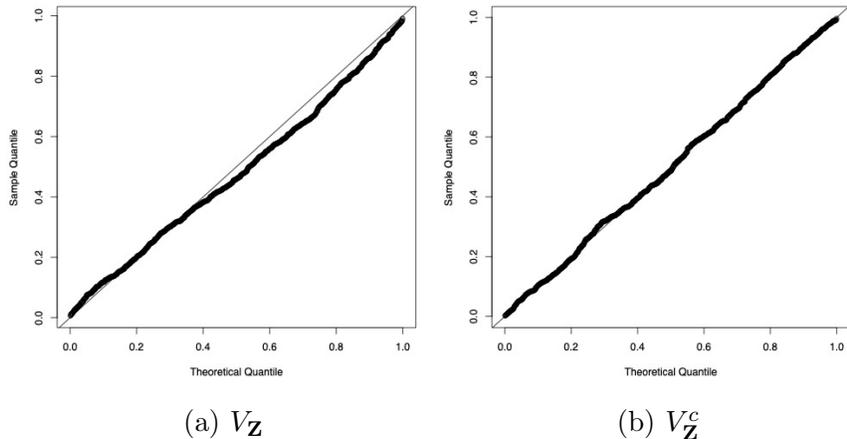


Figure 4.2 Comparison of uniform QQ plot of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^c$ in detecting genetic association without considering genetic heterogeneity effect, with $n=100$ and $p=15$.

Table 4.15 Empirical size and power of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}^{H,c}$ considering genetic heterogeneity across two observable subpopulations and left truncation, with $n = 200$, $p = 10$.

	Empirical Size	Power	
β_1	0	0.02	0.04
$V_{\mathbf{Z}}^{H,c}$	0.051	0.343	0.395
$V_{\mathbf{Z}}^H$	0.045	0.342	0.394

Interaction test under small sample size

In this simulation, we assessed the empirical size and power of weighted V statistics, V_I^c and V_I , in testing $G-G$ and $G-E$ interaction effect. The simulation settings were the same as that of simulations with large sample size in previous sections, except that we used $n = 600$ and $p = q = 2$ for testing $G-G$ interaction effect and $n = 500$ and $p = 2, q = 1$ for testing $G-E$ interaction effect. Table 4.16, 4.17 and figure 4.4, 4.5 show that the p -values of small sample corrected weighted V statistic, V_I^c , is close to $Uniform(0,1)$ distribution, while V_I

is not.

Table 4.16 Empirical size and power of test V_I and V_I^c in testing G-G interaction effect, considering left truncation, with $n = 600$, $p = q = 2$.

	Empirical Size (power)
V_I^c	0.064 (0.413)
V_I	0.065 (0.408)

Table 4.17 Empirical size and power of test V_I and V_I^c in testing G-E interaction effect, considering left truncation, with $n = 500$, $p = 2$.

	Empirical Size (power)
V_I^c	0.054 (0.334)
V_I	0.052 (0.329)

4.3 A Real Application

We applied the developed small sample adjusted (heterogeneity) weighted V test statistics, $V_{\mathbf{Z}}^c$ and $V_{\mathbf{Z}}^{H,c}$, on the ROSMAP dataset by assuming the age-to-onset of Alzheimer’s disease follows the AFT model. ROSMAP is a GWAS that includes dementia exam data from two large longitudinal studies of aging and dementia: 1) the Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP) (Bennett et al., 2018). The ROSMAP genotype dataset includes 1,679 subjects, each has 750,173 SNPs. We first performed quality control on genotype dataset 1) at SNP level by removing SNPs with $MAF > 0.01$, HWE test’s p -value $> 10^{-6}$, or missing rate $> 2\%$, then 2) at the subject level by removing subjects with missing genotype rate $> 2\%$. This results in a quality-controlled genotype dataset with 1,618 subjects each having 619,061 SNPs. There were still $\sim 0.3\%$ missing genotypes, which were then imputed with random samples generated from $Binomial(2, MAF)$, where MAF was estimated from the genotype dataset. PCA was performed using Plink software (version 1.9) to obtain the first 6 principal components that represent the underlying ancestry information.

After that, we grouped all available SNPs into 21,285 different genes by using human genome annotation (GRCh38/hg38) obtained from UCSC Genome Browser. We also formed a new gene APOE4 which was coded as the count of the APOE4- ϵ 4 allele.

The ROSMAP phenotype dataset includes information from 2,543 subjects. Each subject has 1) baseline covariates: race, gender, years of education, and binary study cohort indicator (ROS or MAP) and 2) follow-up information: subject’s age and clinical cognitive diagnosis result at each examination center visit including at the study registry. For our analysis, we include only subjects who are disease-free in the study registry. Age at baseline served as left-truncation time, the age at first AD diagnosis served as survival time, and the age at study termination served as right censoring time if the subject is disease-free throughout the study follow-up period.

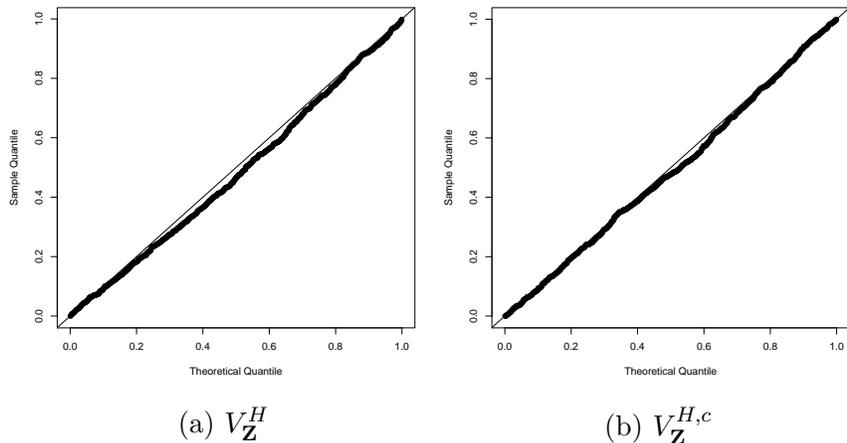


Figure 4.3 Uniform QQ plot of $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}^{H,c}$, considering genetic heterogeneity across two observable subpopulations, with $n=200$ and $p=10$.

We merged preprocessed genotype and phenotype datasets by keeping only the overlapping subjects to generate the analysis-ready dataset, which includes 1,440 subjects, with a censoring rate of 62.5%. We are interested in 1) testing genetic covariates that are associated

with the time-to-onset of first AD diagnosis and 2) detecting potential genetic heterogeneity across subpopulations inferred by either baseline covariates or individual genome profiles (represented by a random sample of 200K SNPs), in the presence of left-truncation time and adjust for gender, year of education, study cohort indicator (ROS served as baseline), and first 6 principal components. Note that race is not adjusted because 1,439 out of 1,440 subjects are white. IBS kernel was used to measure genetic similarity, while the choice of similarity kernel for subpopulation similarity depends on the source of heterogeneity, which is elaborated in Table 4.18.

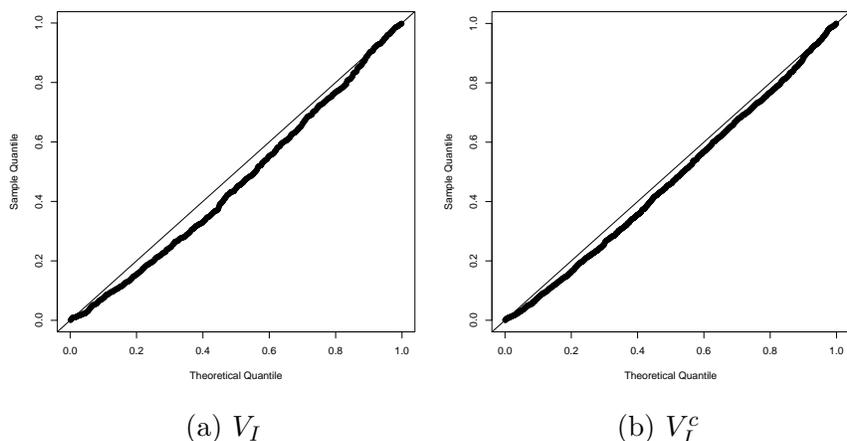


Figure 4.4 Comparison of uniform QQ plot of test V_I and V_I^c in testing G-G interaction effect, with $n=600$, $p=q=2$.

Table 4.18 shows the analysis results of the ROSMAP dataset. To account for multiple testing issues, the p -value thresholds were obtained by controlling the FDR under 10% using the Benjamini-Hochberg procedure Benjamini and Hochberg (1995). Four different kernels were used to measure genetic similarity. When in the absence of genetic heterogeneity (scenario S1), APOE4 appeared to be the most significant gene followed by APOC1. When in the presence of genetic heterogeneity (scenarios S2-S4), even though APOE4 and APOC1

were still the most significant genes, their p -values varies across scenarios and sources of heterogeneity, with notably smaller p -values when Laplacian kernel was used under scenario S2 and S4 that considered genetic heterogeneity across sexes and genetic background. In other words, the effect of APOE4 and APOC1 on age-to-onset of AD varies across subjects' gender and genetic background. In addition to APOE4 and APOC1, IGSF23 is frequently detected as the next most significant gene. The function of IGSF23 is listed in table 4.19.

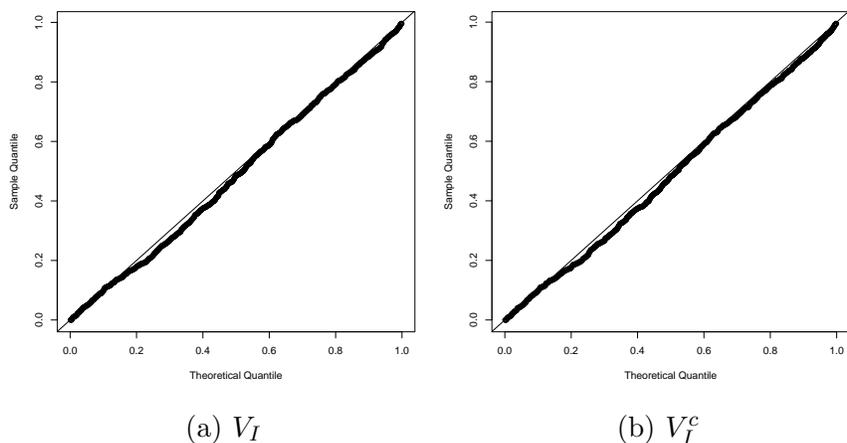


Figure 4.5 Comparison of uniform QQ plot of test V_I and V_I^c in testing G-E interaction effect, with $n=500$, $p=2$, $q=1$.

4.4 Discussion

We developed a suite of novel multi-marker genetic association and interaction tests based on weighted V statistics. The proposed tests have four major advantages: 1) they can handle competing risks, 2) they are accurate under a small sample size, 3) the heterogeneity weighted V statistic is powerful when the genetic effect is heterogeneous in the underlying population, and 4) they can handle quadratic confounding effect.

The asymptotic null distribution of weighted V statistic can be well approximated by a rescaled χ^2 distribution, $c_0\chi_{d_0}^2$, using the Satterthwaite method (Liu, Lin, and Ghosh, 2007).

Table 4.18 Top five genes discovered by $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^{H,c}$ from ROSMAP dataset by assuming the age-to-onset of Alzheimer’s disease follows AFT model. IBS, CP, Lap and Quad stand for IBS, cross-product, Laplacian and Quadratic kernel, respectively, that measures genetic similarity. Various types of heterogeneity were considered, including no genetic heterogeneity (S1), heterogeneity between sexes (S2), heterogeneity across education categories (S3), and heterogeneity across genetic backgrounds (S4).

Kernel	Scenario	Genes and p-values				
IBS	S1	APOE4	APOC1	IGSF23	TBCC	HSBP1
		3.46E-14	1.72E-09	6.99E-05	7.20E-05	1.21E-04
	S2	APOE4	APOC1	PLEKHG5,TNFRSF25	TBCC	GRIP1
		6.45E-14	2.61E-09	9.45E-05	1.41E-04	1.91E-04
	S3	APOE4	APOC1	EFEMP2	MUS81	TBCC
		6.08E-10	1.01E-07	1.58E-04	6.08E-04	9.74E-04
	S4	APOE4	APOC1	IGSF23	TBCC	HSBP1
		7.63E-14	1.50E-09	7.09E-05	7.26E-05	1.22E-04
CP	S1	APOE4	APOC1	IGSF23	MTMR2	HSBP1
		6.66E-16	2.35E-10	4.14E-05	9.32E-05	9.91E-05
	S2	APOE4	APOC1	IGSF23	GRIP1	MTMR2
		1.44E-15	1.11E-09	9.46E-05	1.06E-04	1.42E-04
	S3	APOE4	APOC1	TBCC	IGSF23	TTC1
		1.15E-14	1.08E-09	3.40E-05	9.31E-05	1.50E-04
	S4	APOE4	APOC1	IGSF23	MTMR2	HSBP1
		6.21E-15	3.19E-10	4.23E-05	9.45E-05	9.84E-05
Lap	S1	APOE4	APOC1	IGSF23	GRIP1	TBCC
		1.00E-12	3.83E-08	6.80E-05	8.70E-05	1.23E-04
	S2	APOE4	APOC1	GRIP1	MTMR2	IGSF23
		4.57E-13	2.72E-08	1.13E-04	2.07E-04	2.09E-04
	S3	APOE4	APOC1	EFEMP2	TBCC	DCAF12
		6.05E-12	1.27E-08	1.40E-04	3.15E-04	3.88E-04
	S4	APOE4	APOC1	IGSF23	GRIP1	TBCC
		3.26E-13	5.11E-09	6.92E-05	9.29E-05	1.24E-04
Quad	S1	APOE4	APOC1	GRIP1	EHHADH-AS1	MTMR2
		9.99E-16	2.52E-10	1.07E-04	1.31E-04	2.12E-04
	S2	APOE4	APOC1	EHHADH-AS1	GRIP1	C16orf54
		6.66E-16	1.07E-09	8.77E-05	9.69E-05	1.33E-04
	S3	APOE4	APOC1	TBCC	GRIP1	TTC1
		2.44E-15	2.07E-09	1.14E-04	1.60E-04	1.78E-04
	S4	APOE4	APOC1	GRIP1	EHHADH-AS1	GABBR1
		4.11E-15	1.23E-08	1.08E-04	1.29E-04	2.14E-04
p-value threshold		4.46E-07	8.91E-07	1.34E-06	1.78E-06	2.23E-06

Table 4.19 Function of IGSF23, the third most frequent top gene, according to Entrez.

Gene	Function
IGSF23	This gene encodes a protein that has one immunoglobulin (Ig) domain and is a member of the immunoglobulin superfamily. Proteins in this superfamily are usually found on or in cell membranes and act as receptors in immune response pathways.

c_0 is the scaled parameter and d_0 is the χ^2 distribution’s degree of freedom estimated by matching the mean and variance of weighted V statistics and $c_0\chi_{d_0}^2$. In testing $G-G$ and $G-E$ interaction effect, the scaled χ^2 distribution of weighted V statistic has d_0 degree of freedom,

which is smaller than pq , the number of G - G interaction terms, when there is a weak or strong correlation between SNPs. Therefore the weighted V statistic is more powerful than regular tests (e.g., LRT, Wald test).

CHAPTER 5 Multi-marker genetic association and interaction tests based on the additive hazards model

5.1 Methods

A series of association and interaction tests were developed for survival times based on an additive hazards model possibly under competing risks and/or left truncation.

5.1.1 Association test

Consider a cohort of n independent individuals who have not experienced any competing events at baseline. We assume that each individual is subject to multiple potential sources of failure event and the occurrence of one event impedes the occurrence of all other events. Let $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ be genetic covariates that is either a set of SNPs, gene expression values or a genetic pathway. We are interested in detecting the genetic association with age-to-onset of Cause 1 competing event (primary event) with or without considering genetic heterogeneity. Denote A as left truncation time (e.g., age at study registry), $\tilde{T} = \min(T, C)$ is observed survival time subject to right censoring, where T is survival time and C is right censoring time. $\Delta = I(T \leq C)$ and ϵ is cause of failure, $\epsilon = 1$ for Cause 1 competing event (primary event) and $\epsilon = 0$ for all other causes. Baseline covariates, $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_k)$ is adjusted and when in the presence of genetic heterogeneity, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_d)$ is used to infer subpopulation structure. Observed data for n subjects is denoted as $D = (D_1, \dots, D_n)$, with $D_i = (A_i, \tilde{T}_i, \Delta_i, \Delta_{i\epsilon_i}, \mathbf{G}_i^T, \mathbf{Z}_i^T, \mathbf{X}_i^T)$ (\mathbf{X}_i is included only when in the presence of genetic heterogeneity). We assume the survival time T , the residual censoring time $C - A$ and the left truncation time A are conditionally independent given \mathbf{G} and \mathbf{Z} (and \mathbf{X} if considering genetic heterogeneity).

When in the absence of genetic heterogeneity, under the null hypothesis that \mathbf{G} is in-

dependent of survival time of Cause 1 competing event (primary event), we assume that survival time for subject i ($i = 1, \dots, n$), given adjustment covariate \mathbf{Z} , follows additive hazards model with the following CSH function,

$$\lambda(t|\mathbf{Z}_i) = \lambda_0(t) + \boldsymbol{\beta}^T \mathbf{Z}_i, \quad (5.1)$$

where $\lambda_0(t)$ is an unknown baseline hazard function. In real application, the regression parameter $\boldsymbol{\beta}$ and cumulative baseline hazard function $\Lambda_0(t)$ are estimated by fitting a semi-parametric additive hazards model (Lin and Ying, 1994) to observed data D . The resulting estimators are $\hat{\boldsymbol{\beta}}$ and

$$\hat{\Lambda}_0(\boldsymbol{\beta}, t) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)\boldsymbol{\beta}^T \mathbf{Z}_i\}}{\sum_{j=1}^n Y_j(u)}, \quad (5.2)$$

where $N_i(u) = I(\tilde{T} \leq u, \Delta_i \epsilon_i = 1)$ is counting process and $Y_i(u) = I(A_i < u \leq \tilde{T})$ is at-risk process for subject i at time u . Define $M_i = M_i(\Lambda_0, \boldsymbol{\beta}) = N_i(\infty) - \int_0^\infty Y_i(t)\{d\Lambda_0(t) + \boldsymbol{\beta}^T \mathbf{Z}_i dt\}$ the martingale residual, $\hat{M}_i = M_i(\hat{\Lambda}_0(\boldsymbol{\beta}, t), \boldsymbol{\beta})$. By plugging in the $\hat{\boldsymbol{\beta}}$ or $\hat{\Lambda}_0(\boldsymbol{\beta}, t)$ in equation 5.2, we have $\tilde{M}_i = M_i(\hat{\Lambda}_0(\hat{\boldsymbol{\beta}}, t), \hat{\boldsymbol{\beta}})$. The weighted V test statistics can be written as

$$\begin{aligned} V_{\mathbf{Z}} &= \tilde{\mathbf{M}}^T \mathbf{F} \tilde{\mathbf{M}} \\ &= \sum_{i=1}^n \sum_{j=1}^n f(\mathbf{G}_i, \mathbf{G}_j) \tilde{M}_i \tilde{M}_j, \end{aligned} \quad (5.3)$$

where $\tilde{\mathbf{M}} = (\tilde{M}_1, \dots, \tilde{M}_n)$ and $\mathbf{F} = \{f(\mathbf{G}_i, \mathbf{G}_j)\}_{n \times n}$. $\tilde{M}_i \tilde{M}_j$ in 5.3 is considered as V statistic kernel that measures phenotype similarity weighted by $f(\mathbf{G}_i, \mathbf{G}_j)$, measures genetic similar-

ity, for a pair of subjects (i, j) . The weighted V statistic value is the weighted sum over all pairs of (i, j) . Therefore, the larger (smaller) the phenotype similarity weighted by larger (smaller) genetic similarity results in larger (smaller) $V_{\mathbf{Z}}$ value, therefore smaller (larger) p -value.

To get the asymptotic null distribution of $V_{\mathbf{Z}}$, we decompose $\mathbf{F} = E_1^T E_1$ for some $m \times n$ matrix E_1 , where $m(1 \leq m \leq n)$ is the rank of \mathbf{F} . We then have weighted V statistic in quadratic form, $V_{\mathbf{Z}} = (\tilde{\mathbf{M}}^T E_1^T)(E_1 \tilde{\mathbf{M}})$, and the asymptotic null distribution can be approximated by $\sum_{j=1}^m \lambda_j \chi_{1j}^2$, which is linear combination of independent χ^2 variables of 1 degree of freedom weighted by $\{\lambda_i\}_{i=1}^m$, the eigenvalues of estimator of $Cov(E_1 \tilde{\mathbf{M}})$ (derivation shown in Appendix). Based on this distribution, the p -values can be computed using Davies' method (Davies, 1980), $P(V_{\mathbf{Z}} \geq V_{\mathbf{Z},obs})$, where $V_{\mathbf{Z},obs}$ is the observed value of $V_{\mathbf{Z}}$.

When in the presence of genetic heterogeneity inferred explicitly or implicitly by $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_d)$, the heterogeneity weighted V test statistic is an extension of 5.3, denoted as $V_{\mathbf{Z}}^H = \tilde{\mathbf{M}}^T \mathbf{W} \tilde{\mathbf{M}}$, where $\mathbf{W} = (\mathbf{J} + \mathbf{H}) \odot \mathbf{F}$, $\mathbf{H} = \{h(\mathbf{X}_i, \mathbf{X}_j)\}_{n \times n}$, $h(\mathbf{X}_i, \mathbf{X}_j)$ is a kernel function that measures subpopulation similarity between subject i and j . \odot is element-wise matrix multiplication. The asymptotic null distribution of $V_{\mathbf{Z}}^H$ can be approximated similarly as that of $V_{\mathbf{Z}}$ introduced above by replacing \mathbf{F} with \mathbf{W} in equation 5.3. Based on the estimated distribution, p -value can be computed by $P(V_{\mathbf{Z}}^H \geq V_{\mathbf{Z},obs}^H)$ using Davies' method (Davies, 1980), where $V_{\mathbf{Z},obs}^H$ is observed value of $V_{\mathbf{Z}}^H$.

5.1.2 G-G/G-E interaction test

We developed two interaction tests to detect the potential association between G - G and G - E interaction effect on age-to-onset of Cause 1 competing event (primary event) considering left-truncation time.

In testing G - G interaction effect of two genes $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ and $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_q)$. Under the null hypothesis of no G - G interaction effect on Cause 1 competing event, the adjustment covariate \mathbf{Z} is replaced with (\mathbf{G}, \mathbf{H}) and the genetic main effect is assumed to be homogeneous. Under the null hypothesis, the CSH function of subject i given \mathbf{G} and \mathbf{H} is as follows

$$\lambda(t|\mathbf{G}_i, \mathbf{H}_i) = \lambda_0(t) + \boldsymbol{\beta}^T \mathbf{G}_i + \boldsymbol{\alpha}^T \mathbf{H}_i, \quad (5.4)$$

where $\lambda_0(t)$ is unspecified baseline hazard function. In real application, we fit a semiparametric additive hazards model (Lin and Ying, 1994) to observed data D and obtain the estimates of regression coefficients $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and cumulative baseline hazard function Λ_0 , denoted as $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\alpha}}$, and $\hat{\Lambda}_0$, where $\hat{\Lambda}_0$ is as below

$$\hat{\Lambda}_0(\boldsymbol{\beta}, \boldsymbol{\alpha}, t) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(u) - Y_i(u)\{\boldsymbol{\beta}^T \mathbf{G}_i + \boldsymbol{\alpha}^T \mathbf{H}_i\}\}}{\sum_{j=1}^n Y_j(u)}, \quad (5.5)$$

where $N_i(u)$ is counting process and $Y_i(u)$ is at-risk process for subject i . Define martingale residual $M_i = M_i(\Lambda_0, \boldsymbol{\beta}, \boldsymbol{\alpha}) = N_i(\infty) - \int_0^\infty Y_i(t)\{d\Lambda_0(t) + (\boldsymbol{\beta}^T \mathbf{G}_i + \boldsymbol{\alpha}^T \mathbf{H}_i)dt\}$, $\hat{M}_i = M_i(\hat{\Lambda}_0(\boldsymbol{\beta}, \boldsymbol{\alpha}, t), \boldsymbol{\beta}, \boldsymbol{\alpha})$, and $\tilde{M}_i = M_i(\hat{\Lambda}_0(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, t), \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$. The weighted V test statistic for G - G interaction effect is

$$\begin{aligned} V_I &= \tilde{\mathbf{M}}^T (\mathbf{F} \odot \mathbf{K}) \tilde{\mathbf{M}} \\ &= \sum_{i=1}^n \sum_{j=1}^n f(\mathbf{G}_i, \mathbf{G}_j) k(\mathbf{H}_i, \mathbf{H}_j) \tilde{M}_i \tilde{M}_j. \end{aligned}$$

where $\mathbf{F} = \{f(\mathbf{G}_i, \mathbf{G}_j)\}_{n \times n}$ and $\mathbf{K} = \{k(\mathbf{H}_i, \mathbf{H}_j)\}_{n \times n}$ are genetic similarity matrices, and \odot is element-wise matrix multiplication. The asymptotic null distribution of V_I can be

approximated similarly to $V_{\mathbf{Z}}$ by replacing \mathbf{F} with $\mathbf{F} \odot \mathbf{K}$ in equation 5.3. Based on this estimated distribution, we use Davies' method (Davies, 1980) to compute p -value, $P(V_I \geq V_{I,obs})$, where $V_{I,obs}$ is the observed value of V_I .

For testing G - E interaction effect, the weighted V test statistic is performed in the same way as testing G - G interaction effect, except that \mathbf{H} is replaced by an environmental variable (e.g., gender). The proposed interaction tests are expected to be more powerful than regular tests (e.g., LRT, Wald test), and the advantage over regular tests increases as LD gets stronger because the asymptotic null distribution of weighted V statistic has an effective degree of freedom lower than that of regular tests.

5.1.3 Small sample adjustment

The association and interaction tests introduced above work efficiently when sample size is large. However, as shown in the QQ plot in Figure 5.1 (a) and Figure 5.2 (a) in simulation below, the distribution might not be accurate when n is relative small (as compared to p), which is frequently encountered in whole-exome sequencing studies (Lee et al., 2012; Emond et al., 2012). Therefore, motivated by Chen et al. (2016), we propose a small sample correction on $V_{\mathbf{Z}}$, $V_{\mathbf{Z}}^H$ and V_I , denoted, respectively, as $V_{\mathbf{Z}}^c$, $V_{\mathbf{Z}}^{H,c}$ and V_I^c as follow,

$$V_{\mathbf{Z}}^c = \frac{\tilde{\mathbf{M}}^T \mathbf{F} \tilde{\mathbf{M}}}{\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}}, \quad V_{\mathbf{Z}}^{H,c} = \frac{\tilde{\mathbf{M}}^T \mathbf{W} \tilde{\mathbf{M}}}{\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}}, \quad \text{and} \quad V_I^c = \frac{\tilde{\mathbf{M}}^T (\mathbf{F} \odot \mathbf{K}) \tilde{\mathbf{M}}}{\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}},$$

where $\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}$ accounts for the variability of martingale residual. We expect that the approximated asymptotic null distribution of corrected (heterogeneity) weighted V statistics, $V_{\mathbf{Z}}^c$, $V_{\mathbf{Z}}^{H,c}$, and V_I^c have exact mixture of χ^2 distribution. For $V_{\mathbf{Z}}^c$, based on the distribution, we can apply Davies' method (Davies, 1980) to compute p -values by $P(V_{\mathbf{Z}}^c \geq V_{\mathbf{Z},obs}^c) = P(\frac{\tilde{\mathbf{M}}^T \mathbf{F} \tilde{\mathbf{M}}}{\tilde{\mathbf{M}}^T \tilde{\mathbf{M}}} \geq$

$V_{\mathbf{Z},obs}) = P(\tilde{\mathbf{M}}^T(\mathbf{F} - V_{\mathbf{Z},obs}\mathbf{I})\tilde{\mathbf{M}} \geq 0)$. The p -value of $V_{\mathbf{Z}}^{H,c}$ and V_I^c can be computed similarly.

5.2 Simulations

We performed a Monte Carlo simulation to assess the finite-sample performance of (heterogeneity) weighted V tests on survival time that follow the additive hazards model, considering adjustment covariates and left-truncation time. Each individual potentially experiences two competing events with Cause 1 (primary event) and Cause 2. In all simulations, unless otherwise specified, two sample sizes $n = 600, 800$ and two SNP set sizes $p = 8, 12$ were considered. Genetic covariates, denoted as $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$, is a set of p SNPs (coded by 0, 1, or 2), unless otherwise specified in the scenario under confounding effect. The generation of \mathbf{G} follows a two-step procedure that considered potential LD structure, 1) sample n vectors independently from $MVN(\mathbf{0}, \Sigma)$, where $\Sigma = \{0.5^{|k-l|}\}_{p \times p}$, the kl -th element is the covariance between \mathbf{G}_k and \mathbf{G}_l ; 2) categorize each quantitative column into SNP with three levels 0, 1, and 2 by using predetermined quantile cutoffs: a^2 , $(1-a)^2$, and $2a(1-a)$ of standard normal distribution with $a \sim Beta(2, 5)$, to ensure the resulting population is in Hardy-Weinberg equilibrium (HWE). Two baseline covariates were adjusted $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$, one binary $Z_1 \sim Binomial(0.5)$ and one continuous $Z_2 \sim Uniform(3, 5)$. Unless otherwise specified, we assumed that there is no confounding effect, in other words, \mathbf{Z} is independent of \mathbf{G} . The failure of all competing events was assumed to follow the additive hazards model, and the observed survival times were generated following steps in section 3.2 of Beyersmann et al. (2009). Left truncation times were generated from $Uniform(0, 1)$ and the residual censoring times from $Exponential(0.3)$ to control the right censoring rate around 35%. In all simulations, 1000 Monte Carlo samples were generated and the significance level of a test was set at 0.05 unless otherwise specified in testing genetic association under the stringent

p -value threshold scenario.

5.2.1 Association test in the absence of genetic heterogeneity

In this series of simulation, we investigated the empirical size and power of weighted V statistic ($V_{\mathbf{Z}}$) in the absence of genetic heterogeneity. For subject i , both competing events were assumed to follow additive hazards model with cause specific hazard (CSH) function of the following form,

$$\lambda(t|\mathbf{Z}_i) = ct^2 + \boldsymbol{\beta}^T \mathbf{G}_i + \boldsymbol{\alpha}^T \mathbf{Z}_i, \quad (5.6)$$

where ct^2 is baseline hazard function. Coefficients, c , $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ vary depending on simulation scenario and cause of failure.

Empirical size and power under various n 's and p 's

In this simulation, we assessed the empirical size and power of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$. CSH function of both competing events for subject i ($i = 1, \dots, n$) are as follow,

$$\begin{cases} \lambda_1(t|\mathbf{Z}_i, \mathbf{G}_i) = 0.6t^2 + \sum_{j=1}^p \beta_j G_{ij} + \sum_{k=1}^2 0.5Z_{ik}, \\ \lambda_2(t|\mathbf{Z}_i, \mathbf{G}_i) = 0.3t^2 + \sum_{j=1}^p 0.05G_{ij} + \sum_{k=1}^2 0.1Z_{ik}, \end{cases}$$

where λ_1 is CSH of Cause 1 competing event. We set $\beta_1 = \dots = \beta_p$ to 0 and 0.2 for empirical size and power assessment, while CSH for Cause 2 competing event stays the same. IBS kernel was used to measure genetic similarity in both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$. Gaussian kernel was used to measure similarity of adjustment covariates \mathbf{Z} for test $V_{\mathbf{Z}}^H$. Table 5.1 shows that the Type I error of both tests are close to nominal level and their power are comparable under various n 's and p 's. This indicates that when the genetic effect is homogeneous, both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are valid and have similar statistical power.

Table 5.1 Empirical size and power comparison of test $V_{\mathbf{Z}}$ and test $V_{\mathbf{Z}}^H$ in testing genetic effects considering adjustment covariates and left truncation.

	Empirical Size (Power)	
	p=8, n=600	p=8, n=800
$V_{\mathbf{Z}}$	0.058 (0.584)	0.052 (0.667)
$V_{\mathbf{Z}}^H$	0.059 (0.572)	0.049 (0.674)
	p=12, n=600	p=12, n=800
$V_{\mathbf{Z}}$	0.050 (0.653)	0.054 (0.784)
$V_{\mathbf{Z}}^H$	0.052 (0.640)	0.052 (0.772)

Empirical size and power under quadratic confounding

In this simulation, we assessed the empirical size and power of $V_{\mathbf{Z}}$ in the absence of genetic heterogeneity. We assumed that \mathbf{Z} and \mathbf{G} are quadratically related, in other words, there is quadratic confounding effect. The genetic covariate was replaced by gene expression values generated by $\mathbf{G}_i = Z_{i1} + Z_{i2} + Z_{i1}^2 + Z_{i2}^2 + Z_{i1}Z_{i2} + \mathbf{e}_i$ ($i = 1, \dots, n$), where $\mathbf{e}_i \sim MVN(\mathbf{0}, \Sigma)$, where $\Sigma = \{0.5^{|k-l|}\}_{p \times p}$ the covariance between any pair of genetic variants $(\mathbf{G}_k, \mathbf{G}_l)$. CSH function for Cause 1 and Cause 2 competing events of subject i are as follow,

$$\begin{cases} \lambda_1(t|\mathbf{Z}_i, \mathbf{G}_i) = 9t^2 + \sum_{j=1}^p \beta_j G_{ij} + \sum_{k=1}^2 0.1 Z_{ik}, \\ \lambda_2(t|\mathbf{Z}_i, \mathbf{G}_i) = 3t^2 + \sum_{j=1}^p 0.05 G_{ij} + \sum_{k=1}^2 0.025 Z_{ik}, \end{cases}$$

where λ_1 is CSH for Cause 1 competing event, $\beta_1 = \dots = \beta_p$ are set to 0 and 0.1 for empirical size and power assessment, while CSH for Cause 2 stays the same. Gaussian kernel was used to measure gene expression similarity. Table 5.2 shows that Type I error of $V_{\mathbf{Z}}$ is close to nominal level and power increases with sample size. This indicates that weighted V statistic handles quadratic confounding effect properly.

Table 5.2 Empirical size and power of test $V_{\mathbf{Z}}$ in testing genetic effects under quadratic confounding and left truncation.

	Empirical Size (Power)	
	p=4, n=600	p=4, n=800
$V_{\mathbf{Z}}$	0.047 (0.212)	0.051 (0.320)
	p=6, n=600	p=6, n=800
	$V_{\mathbf{Z}}$	0.051 (0.241)

5.2.2 Association test in the presence of genetic heterogeneity

In this series of simulations, we investigated the performance of heterogeneity weighted V statistics, $V_{\mathbf{Z}}^H$, in the presence of genetic heterogeneity across 1) observable subpopulations, 2) two latent subpopulations, 3) twenty subpopulations, and 4) individual genome profile. The advantage of heterogeneity weighted V statistic $V_{\mathbf{Z}}^H$ over $V_{\mathbf{Z}}$ is obvious through the performance comparison shown below. For all simulations in this section, we assume that the effect of \mathbf{G} on survival time is heterogeneous for Cause 1 and homogeneous for Cause 2.

Genetic heterogeneity across observable subpopulation

In this simulation, we assessed the empirical size and power of heterogeneity weighted V statistic, $V_{\mathbf{Z}}^H$, in the presence of heterogeneity across two observable subpopulations. Therefore, two genetic covariates (each has sample size $n/2$) were generated using two different sets of MAFs ($\sim Beta(2, 5)$) and LD structure (measured by ρ in $\Sigma = \{\rho^{|k-l|}\}$) as follow,

MAFs = {0.16187719, 0.24027357, 0.29026717, 0.42216265, 0.10528255, 0.50324631, 0.13275845, 0.48439386, 0.07259317, 0.20828786, 0.22822443, 0.39021594} and $\rho = 0.5$ in Subpopulation 1;

MAFs = {0.17645756, 0.22244350, 0.08635246, 0.40581343, 0.49636636, 0.48130878, 0.25831106, 0.38090314, 0.14896296, 0.16374183, 0.29928099, 0.04836184} and $\rho = 0.3$

in Subpopulation 2.

The first p MAFs will be used. We used a one-dimensional vector $\mathbf{X} = (X_1, \dots, X_n) \sim \text{Binomial}(0.5)$ to infer two observable subpopulations (e.g., male and female). The CSH of Cause 1 and Cause 2 for subject i are as follow,

$$\begin{cases} \lambda_1(t|\mathbf{Z}_i, \mathbf{G}_i) = 9t^2 + \sum_{j=1}^p (\beta_0 + \beta_1 Z_i) G_{ij} + 0.1X_i, \\ \lambda_2(t|\mathbf{Z}_i, \mathbf{G}_i) = 3t^2 + \sum_{j=1}^p 0.001G_{ij} + 0.1X_i, \end{cases}$$

where $\beta_0 = \beta_1 = 0$ for empirical size assessment and $(\beta_0, \beta_1) = \{(0.002, 0.6), (0.002, 0.7)\}$ for power assessment under different genetic heterogeneity sizes ($\beta_1 = 0.6, 0.7$). IBS kernel was used to measure genetic similarity and identity kernel to measure subpopulation similarity. Performance of weighted V statistic, $V_{\mathbf{Z}}$, was also included. Table 5.3 shows that Type I error of both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are close to nominal level. Table 5.4 shows that $V_{\mathbf{Z}}^H$ is more powerful than $V_{\mathbf{Z}}$ by taking advantage of the subpopulation structure. Also, as size of heterogeneity increases, $V_{\mathbf{Z}}^H$ gains more power.

Table 5.3 Empirical sizes of test $V_{\mathbf{Z}}$ and test $V_{\mathbf{Z}}^H$ in testing genetic association under genetic heterogeneity across two observable subpopulations, considering adjustment covariates and left truncation.

	Empirical Size	
	p=8, n=600	p=8, n=800
$V_{\mathbf{Z}}^H$	0.059	0.054
$V_{\mathbf{Z}}$	0.058	0.052
	p=12, n=600	p=12, n=800
$V_{\mathbf{Z}}^H$	0.046	0.049
$V_{\mathbf{Z}}$	0.046	0.045

Table 5.4 Power of test $V_{\mathbf{Z}}$ and test $V_{\mathbf{Z}}^H$ in testing genetic association under genetic heterogeneity across two observable subpopulations, considering adjustment covariates and left truncation.

	Power			
	p=8, n=600		p=8, n=800	
β_1	0.6	0.7	0.6	0.7
$V_{\mathbf{Z}}^H$	0.109	0.554	0.169	0.697
$V_{\mathbf{Z}}$	0.103	0.454	0.158	0.551
	p=12, n=600		p=12, n=800	
	0.6	0.7	0.6	0.7
β_1	0.6	0.7	0.6	0.7
$V_{\mathbf{Z}}^H$	0.135	0.594	0.192	0.777
$V_{\mathbf{Z}}$	0.114	0.472	0.176	0.633

Genetic heterogeneity across two latent subpopulations

In this simulation, we assessed the empirical size and power of $V_{\mathbf{Z}}^H$ in the presence of genetic heterogeneity across two latent subpopulations. A one-dim vector $\mathbf{X} = (X_1, \dots, X_n)$ was generated to infer two latent subpopulation structure. $X_i = a_i + 1 + e_i$ ($i = 1, \dots, n$), where $a_i \sim \text{Binomial}(0.5)$ and $e_i \sim \text{Normal}(0, 0.5)$. The CSH of Cause 1 and Cause 2 for subject i in subpopulation j are as follow

$$\begin{cases} \lambda_1(t|\mathbf{Z}_{ij}, \mathbf{G}_{ij}) = 9t^2 + \sum_{k=1}^p G_{ijk}\beta_{jk} + 0.5Z_{ij1} + 0.5Z_{ij2}, \\ \lambda_2(t|\mathbf{Z}_{ij}, \mathbf{G}_{ij}) = 3t^2 + \sum_{k=1}^p 0.0025G_{ik} + 0.1Z_{i1} + 0.1Z_{i2}, \end{cases} \quad (5.7)$$

where λ_1 is CSH of Cause 1, which is of our primary interest and λ_2 is CSH of the other competing risk. $\{\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp}) \text{ s.t. } \beta_{j1} = \dots = \beta_{jp}\}_{j=1}^2$ are genetic effects of p SNPs in two latent subpopulations. $\{\boldsymbol{\beta}_j\}_{j=1}^2$ were set to 0 for empirical size assessment and set to different values for power assessment. IBS kernel was used to measure genetic similarity and Gaussian kernel for subpopulation similarity. Four heterogeneity scenarios were considered

determined by the values of β_{1k} and β_{2k} . This includes the same effect size and the same effect direction (T1), identical sizes but opposite directions (T2), no effect in one subpopulation while positive effect in the other (T3), and different sizes but the same direction (T4). Table 5.5 shows that Type I error of both $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ are close to nominal level. Table 5.6 shows that when in the absence of genetic heterogeneity (T1), $V_{\mathbf{Z}}$ is more powerful, however, when in the presence of genetic heterogeneity, $V_{\mathbf{Z}}^H$ is more powerful and gains power as heterogeneity size increases.

Table 5.5 Empirical size of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across two latent subpopulations, with covariates adjustment and left truncation.

	Empirical Size	
	p=8, n=600	p=8, n=800
$V_{\mathbf{Z}}^H$	0.058	0.055
$V_{\mathbf{Z}}$	0.051	0.059
	p=12, n=600	p=12, n=800
$V_{\mathbf{Z}}^H$	0.052	0.058
$V_{\mathbf{Z}}$	0.054	0.052

Genetic heterogeneity across twenty latent subpopulation

In this simulation, we assessed empirical size and power of heterogeneity weighted V statistic in the presence of twenty latent subpopulations. The simulation setting is same to simulation above, except that we increased the number of latent subpopulation to twenty. We used $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{25})$ to infer subpopulation structure. $\mathbf{X}_k = \mathbf{a}_k + \mathbf{e}_k$ ($k = 1, \dots, 25$) where \mathbf{a}_k is a length n bootstrap sample from $\{1, \dots, 20\}$, representing subgroup assignments. Random error $\mathbf{e}_k = (e_{k1}, \dots, e_{kn}) \sim Normal(0, 0.5)$. Regression coefficients for j subpopulations, $\{\boldsymbol{\beta}_j = (\beta_{j1}, \dots, \beta_{jp}), s.t. \beta_{j1} = \dots = \beta_{jp}\}_{j=1}^{20}$, in CSH function 5.7 were all set to 0 for empirical size assessment. In power assessment, $\{\boldsymbol{\beta}_j\}_{j=1}^{20}$ were sampled from

Table 5.6 Power of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across two latent subpopulations, considering adjustment covariates and left truncation. Various heterogeneity scenarios were considered, determined by the values of β_{1k} and β_{2k} .

		Heterogeneity Scenario							
		T1		T2		T3		T4	
β_{1k}		0.2	0.3	0.25	0.35	0	0	0.05	0.1
β_{2k}		0.2	0.3	0.05	0.05	0.15	0.25	-0.05	-0.1
p=8,n=600	$V_{\mathbf{Z}}^H$	0.193	0.410	0.244	0.394	0.138	0.316	0.095	0.216
	$V_{\mathbf{Z}}$	0.448	0.724	0.243	0.362	0.103	0.176	0.048	0.049
p=8,n=800	$V_{\mathbf{Z}}^H$	0.258	0.512	0.341	0.529	0.187	0.377	0.103	0.252
	$V_{\mathbf{Z}}$	0.512	0.833	0.328	0.436	0.119	0.221	0.049	0.057
p=12,n=600	$V_{\mathbf{Z}}^H$	0.180	0.396	0.341	0.624	0.239	0.486	0.146	0.427
	$V_{\mathbf{Z}}$	0.566	0.819	0.270	0.392	0.088	0.172	0.047	0.044
p=12,n=800	$V_{\mathbf{Z}}^H$	0.261	0.529	0.499	0.759	0.311	0.642	0.171	0.538
	$V_{\mathbf{Z}}$	0.659	0.905	0.323	0.497	0.131	0.226	0.055	0.078

uniform distribution with mean μ_{β} and variance σ_{β}^2 . IBS kernel was used to measure genetic similarity and Gaussian kernel for background similarity. Table 5.7 shows that Type I error of both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are close to nominal level. Table 5.8 shows that $V_{\mathbf{Z}}^H$ is more powerful than $V_{\mathbf{Z}}$ and as heterogeneity size (σ_{β}) increases, $V_{\mathbf{Z}}^H$ gains more power.

Table 5.7 Empirical size of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ in testing genetic effects under genetic heterogeneity across twenty latent subpopulations considering adjustment covariates and left truncation.

	Empirical Size	
	p=8, n=600	p=8, n=800
$V_{\mathbf{Z}}^H$	0.050	0.057
$V_{\mathbf{Z}}$	0.052	0.053
	p=12, n=600	p=12, n=800
$V_{\mathbf{Z}}^H$	0.040	0.044
$V_{\mathbf{Z}}$	0.041	0.042

Table 5.8 Power of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ under genetic heterogeneity across twenty latent subpopulations, considering adjustment covariates and left truncation.

	Power			
	p=8, n=600		p=8, n=800	
	(0.15, 0.15)	(0.15, 0.25)	(0.15, 0.15)	(0.15, 0.25)
$(\mu_{\beta}, \sigma_{\beta})$				
$V_{\mathbf{Z}}^H$	0.211	0.434	0.261	0.532
$V_{\mathbf{Z}}$	0.177	0.139	0.185	0.167
	p=12, n=600		p=12, n=800	
	(0.15, 0.15)	(0.15, 0.25)	(0.15, 0.15)	(0.15, 0.25)
$(\mu_{\beta}, \sigma_{\beta})$				
$V_{\mathbf{Z}}^H$	0.353	0.691	0.449	0.844
$V_{\mathbf{Z}}$	0.172	0.149	0.218	0.183

Genetic heterogeneity across individual genome profile

In this simulation, we investigated the empirical size and power of heterogeneity weighted V statistic in the presence of heterogeneity across individual genome profile. In stead of considering two or twenty 'categorical' subpopulation structure, we now assume 'continuous' subpopulation structure. In other words, each individual is a subpopulation. We used a genome profile of 1,000 SNPs, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{1000})$, to infer subpopulation structure. \mathbf{X} was generated in three steps, 1) generate genetic effect sizes $\{\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{ip}), s.t. \beta_{i1} = \dots = \beta_{ip}\}_{i=1}^n$ from uniform distribution with mean μ_{β} and variance σ_{β}^2 , 2) generate n -dim vectors \mathbf{X}_d ($d = 1, \dots, 1000$) from $MVN(\mathbf{0}, \Sigma)$, where *Sigma* is a $n \times n$ identity matrix under null hypothesis and $\{\exp(-|\beta_{i1} - \beta_{j1}|/\sigma_{\beta})\}_{n \times n}$ under alternative hypothesis, 3) categorize \mathbf{X}_d ($d = 1, \dots, 1000$) into SNPs coded by 0, 1, or 2 by using predetermined quantile cutoffs: a^2 , $2a(1 - a)$, and $(1 - a)^2$ of a standard normal distribution with a is MAF generated from $Beta(2, 5)$, to ensure the resulting population is in HWE. IBS kernel was used to measure both genetic and subpopulation similarity. Table 5.9 shows that Type I error of both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are close to nominal level. Table 5.10 shows that $V_{\mathbf{Z}}^H$ is more powerful than $V_{\mathbf{Z}}$ and

for fixed mean effect size ($\mu_\beta = 0.15$), as heterogeneity size (σ_β) increases from 0.15 to 0.25, $V_{\mathbf{Z}}^H$ gains more power.

Table 5.9 Empirical size of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ in testing genetic effects under genetic heterogeneity across individual genome profiles, considering adjustment covariates and left truncation.

	Empirical Size	
	p=8, n=600	p=8, n=800
$V_{\mathbf{Z}}^H$	0.050	0.053
$V_{\mathbf{Z}}$	0.048	0.053
	p=12, n=600	p=12, n=800
$V_{\mathbf{Z}}^H$	0.048	0.052
$V_{\mathbf{Z}}$	0.052	0.051

Table 5.10 Power of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}$ under genetic heterogeneity across individual genome profiles, considering adjustment covariates and left truncation.

	Power			
	p=8, n=600		p=8, n=800	
$(\mu_\beta, \sigma_\beta)$	(0.15, 0.15)	(0.15, 0.25)	(0.15, 0.15)	(0.15, 0.25)
$V_{\mathbf{Z}}^H$	0.196	0.354	0.289	0.518
$V_{\mathbf{Z}}$	0.154	0.109	0.172	0.133
	p=12, n=600		p=12, n=800	
$(\mu_\beta, \sigma_\beta)$	(0.15, 0.15)	(0.15, 0.25)	(0.15, 0.15)	(0.15, 0.25)
$V_{\mathbf{Z}}^H$	0.404	0.720	0.541	0.892
$V_{\mathbf{Z}}$	0.163	0.106	0.192	0.078

5.2.3 Testing G-G/G-E interaction effect

In this simulation, we investigated the performance of weighted V statistic in testing interaction effect. The performance was also compared with Wald test. For testing G - G interaction effect between two genes (SNP set of size p and q): $\mathbf{G} = (\mathbf{G}_1, \dots, \mathbf{G}_p)$ and $\mathbf{H} = (\mathbf{H}_1, \dots, \mathbf{H}_q)$.

We assumed that two genes come from the same population, therefore have the same MAF

and LD structures. CSH of Cause 1 and Cause 2 for subject i is as follow,

$$\begin{cases} \lambda_1(t|\mathbf{G}_i, \mathbf{H}_i) = 6t^2 + \sum_{j=1}^p 0.2G_{ij} + \sum_{j=1}^q 0.25H_{ij} + \sum_{l=1}^{pq} \beta_l(GH)_l, \\ \lambda_2(t|\mathbf{G}_i, \mathbf{H}_i) = 3t^2 + \sum_{j=1}^p 0.1G_{ij} + \sum_{j=1}^q 0.15H_{ij}, \end{cases}$$

where λ_1 is CSH for Cause 1 competing event (primary event), $(GH)_l$ is l -th column of interaction between gene \mathbf{G} and \mathbf{H} . Two SNP-set sizes $p = q = 2$ and $p = q = 3$ were used. We set $\beta_l = 0$ under null hypothesis of no G - G interaction effect and $\beta_l = 0.2$ under alternative hypothesis, when G - G interaction effect exists only in Cause 1 (primary event).

Cross-product kernel was used to measure genetic similarity for both gene \mathbf{G} and \mathbf{H} .

For testing G - E interaction effect, the simulation setting is the same as testing G - G interaction effect except that \mathbf{H} is replaced by a one-dim vector binary environmental variable generated from $Binomial(0.5)$. Two SNP-set sizes $p = 4, q = 1$ and $p = 6, q = 1$ were considered. Identity kernel was used to measure environmental variable similarity.

Table 5.11 and 5.12 show that Type I error of both V_I and Wald test are close to nominal level. In all scenarios, V_I is more powerful than the Wald test since the asymptotic null distribution of V_I has a more effective degree of freedom than the Wald test. Furthermore, the advantage of V_I over the Wald test increases as the correlation between SNPs (LD) gets stronger.

5.2.4 Empirical size under stringent p-value thresholds

In case-control studies, there is usually a huge amount of genetic variants are tested. To account for the multiple testing issue, Bonferroni correction is used and results in a p -value threshold that is 5×10^{-3} or smaller. In this simulation, we investigated the performance of 1)

Table 5.11 Empirical sizes and powers of test V_I and Wald test in testing G-G interaction considering left truncation.

	Empirical Size (power)	
	p=2, q=2, n=600	p=2, q=2, n=800
V_I	0.048 (0.208)	0.058 (0.319)
Wald	0.058 (0.133)	0.052 (0.197)
	Empirical Size (power)	
	p=3, q=3, n=600	p=3, q=3, n=800
V_I	0.053 (0.397)	0.055 (0.562)
Wald	0.044 (0.184)	0.059 (0.265)

Table 5.12 Empirical sizes and powers of test V_I and Wald test in detecting G-E interaction considering left truncation.

	Empirical Size (power)	
	p=4, q=1, n=600	p=4, q=1, n=800
V_I	0.044 (0.506)	0.043 (0.673)
Wald	0.044 (0.358)	0.042 (0.485)
	Empirical Size (power)	
	p=6, q=1, n=600	p=6, q=1, n=800
V_I	0.043 (0.620)	0.048 (0.786)
Wald	0.042 (0.395)	0.042 (0.573)

$V_{\mathbf{Z}}$ in the absence of genetic heterogeneity and 2) $V_{\mathbf{Z}}^H$ in the presence of genetic heterogeneity across observable subpopulations under stringent p -value thresholds. The corresponding simulation settings for significance level 0.5 were adopted, except that only scenario $p = 8, n = 800$ was investigated and 500K Monte Carlo samples were generated and empirical size is the proportion of p -values less than the corresponding thresholds. Table 5.13 shows that Type I error of both $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ are close to nominal level, indicating the capability of $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ for large-scale GWAS.

5.2.5 Small-sample correction

So far, we have shown that $V_{\mathbf{Z}}$, $V_{\mathbf{Z}}^H$, and V_I performed well under relatively large sample sizes as compared to SNP-set size. However, we noticed that the proposed weighted V

Table 5.13 Empirical sizes of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^H$ under stringent p -value thresholds and left truncation.

Threshold	Empirical Size	
	$V_{\mathbf{Z}}$	$V_{\mathbf{Z}}^H$
0.05	0.050	0.051
0.005	0.0051	0.0050
0.0005	0.00053	0.00049
0.00005	0.000044	0.000047

statistics could be asymptotically wrong. In this series of simulations, we investigated the performance of small sample corrected (heterogeneity) weighted V statistic in the following three scenarios: 1) test $V_{\mathbf{Z}}^c$ in the absence of genetic heterogeneity, 2) test $V_{\mathbf{Z}}^{H,c}$ in the presence of genetic heterogeneity across observable subpopulations, and 3) test V_I^c in the presence of G - G and G - E interaction effect.

In the absence of genetic heterogeneity under small sample size

In this simulation, we assessed the empirical size and power of $V_{\mathbf{Z}}^c$ and $V_{\mathbf{Z}}$ in testing genetic association in the absence of genetic heterogeneity. The simulation settings were the same as that of the simulations shown above, except that we used sample size $n = 100$ and SNP-set size $p = 15$. Table 5.14 and Figure 5.1 show that small sample corrected weighted V statistic, $V_{\mathbf{Z}}^c$, is asymptotically correct without loss of power, while the unadjusted weighted V statistic, $V_{\mathbf{Z}}$, is not.

In the presence of observable subpopulation under small sample size

In this simulation, we assessed the empirical size and power of $V_{\mathbf{Z}}^{H,c}$ and $V_{\mathbf{Z}}^H$ in testing genetic association in the presence of genetic heterogeneity across observable subpopulations. The simulation settings were the same as simulations for unadjusted heterogeneity weighted V statistics, except that we used sample size $n = 200$ and SNP-set size $p = 10$. Table 5.15

and figure 5.2 show that small sample corrected weighted V statistic, $V_{\mathbf{Z}}^{H,c}$, is asymptotically correct without loss of power, while the unadjusted heterogeneity weighted V statistic, $V_{\mathbf{Z}}^H$, is not.

Table 5.14 Empirical size and power of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^c$ considering left truncation, with $n = 100, p = 15$.

	Empirical Size (power)
$V_{\mathbf{Z}}^c$	0.059 (0.141)
$V_{\mathbf{Z}}$	0.051 (0.134)

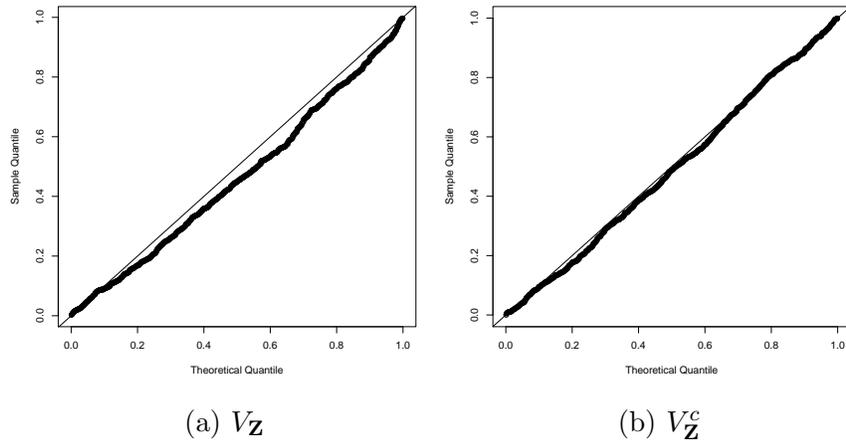


Figure 5.1 Comparison of uniform QQ plot of test $V_{\mathbf{Z}}$ and $V_{\mathbf{Z}}^c$ in detecting genetic association without considering genetic heterogeneity effect, with $n=100$ and $p=15$.

Table 5.15 Empirical size and power of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}^{H,c}$ considering genetic heterogeneity across two observable subpopulations and left truncation, with $n = 200, p = 10$.

	Empirical Size	Power	
β_1	0	0.6	0.7
$V_{\mathbf{Z}}^{H,c}$	0.046	0.211	0.249
$V_{\mathbf{Z}}^H$	0.043	0.196	0.237

Interaction test under small sample size

In this simulation, we assessed the empirical size and power of V_I^c and V_I in testing $G-G$ and $G-E$ interaction effects. The simulation settings were the same to simulation for weighted V

statistic without small-sample correction that test interaction effect. To mimic small sample scenario, we used smaller sample size and larger SNP-set size. Specifically, we used $n = 300$ and $p = 10$ for testing $G-G$ interaction effect and $n = 200$ and $p = 10$ for testing $G-E$ interaction effect. Table 5.16 and 5.17 show that type I error of small-sample corrected test statistic V_I^c is close to nominal level while that of V_I is slightly inflated. Figure 5.3, 5.4 show that small sample corrected weighted V statistic, V_I^c , is asymptotically correct without loss of testing power, while V_I has slightly inflated Type I error under small sample size.

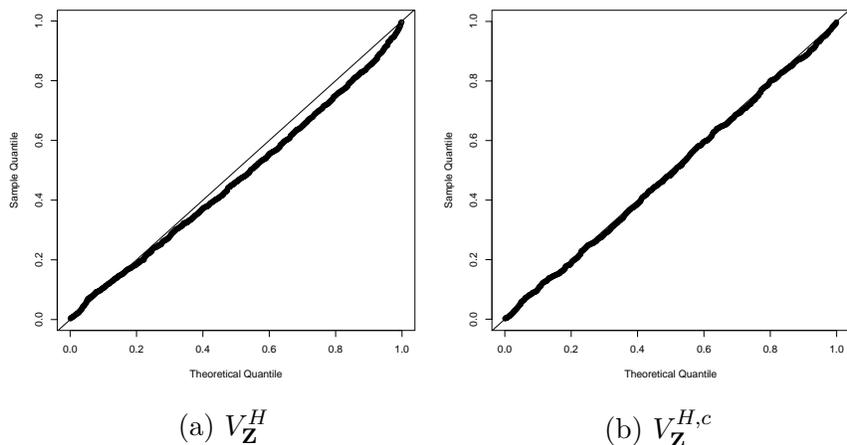


Figure 5.2 Comparison of uniform QQ plot of test $V_{\mathbf{Z}}^H$ and $V_{\mathbf{Z}}^{H,c}$, considering genetic heterogeneity across two observable subpopulations, with $n = 200$ and $p = 10$.

Table 5.16 Empirical size and power of test V_I and V_I^c in testing G-G interaction effect, considering left truncation, with $n = 300$, $p = q = 10$.

	Empirical Size (power)
V_I^c	0.044 (0.134)
V_I	0.059 (0.198)

5.3 A Real Application

We applied the developed small sample adjusted (heterogeneity) weighted V test statistics, $V_{\mathbf{Z}}^c$ and $V_{\mathbf{Z}}^{H,c}$, on ROSMAP by assuming the age-to-onset of AD follows the additive hazards

Table 5.17 Empirical size and power of test V_I and V_I^c in testing G-E interaction effect, considering left truncation, with $n = 200$, $p = 10$.

	Empirical Size (power)
V_I^c	0.055 (0.152)
V_I	0.069 (0.177)

model. ROSMAP is a GWAS that includes dementia exam data from two large longitudinal studies of aging and dementia: 1) the Religious Orders Study (ROS) and the Rush Memory and Aging Project (MAP) (Bennett et al., 2018). The ROSMAP genotype dataset includes 1,679 subjects, each has 750,173 SNPs. We first performed quality control on genotype dataset 1) at SNP level by removing SNPs with $MAF > 0.01$, HWE test's p -value $> 10^{-6}$, or missing rate $> 2\%$, then 2) at the subject level by removing subjects with missing genotype rate $> 2\%$. This results in a quality-controlled genotype dataset with 1,618 subjects each having 619,061 SNPs. There were still $\sim 0.3\%$ missing genotypes, which were then imputed with random samples generated from $Binomial(2, MAF)$, where MAF was estimated from the genotype dataset. PCA was performed using Plink software (version 1.9) to obtain the first 6 principal components that represent the underlying ancestry information. After that, we grouped all available SNPs into 21,285 different genes by using human genome annotation (GRCh38/hg38) obtained from UCSC Genome Browser. We also formed a new gene APOE4 which was coded as the count of the APOE4- $\epsilon 4$ allele.

The ROSMAP phenotype dataset includes information from 2,543 subjects. Each subject has 1) baseline covariates: race, gender, years of education, and binary study cohort indicator (ROS or MAP) and 2) follow-up information: subject's age and clinical cognitive diagnosis result at each examination center visit including at the study registry. For our analysis, we include only subjects who are disease-free in the study registry. Age at baseline served as

left-truncation time, the age at first AD diagnosis served as survival time, and the age at study termination served as right censoring time if the subject is disease-free throughout the study follow-up period.

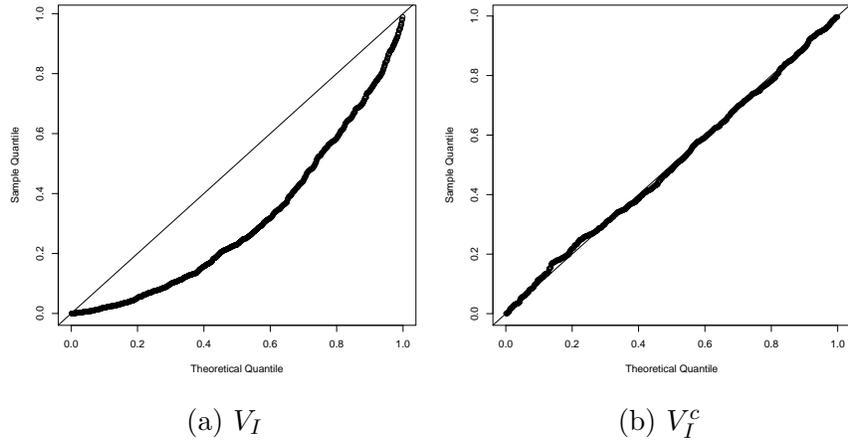


Figure 5.3 Comparison of uniform QQ plot of test V_I and V_I^c in testing G-G interaction effect, with $n=300$, $p=q=10$.

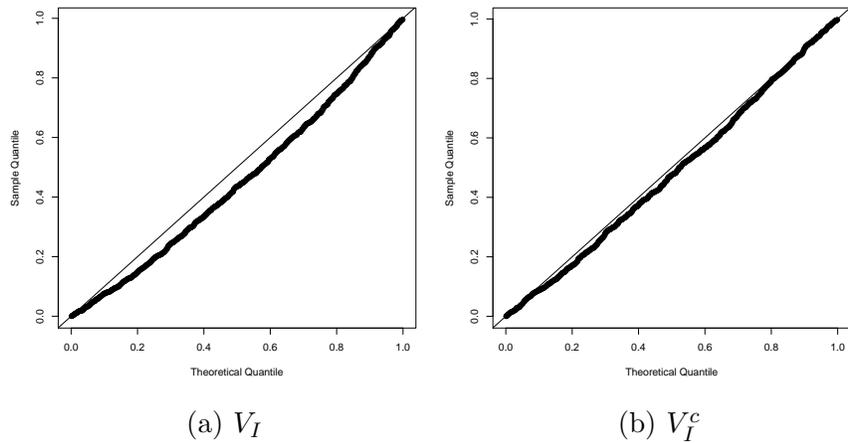


Figure 5.4 Comparison of uniform QQ plot of test V_I and V_I^c in testing G-E interaction effect, with $n=200$, $p=10$, $q=1$.

We merged preprocessed genotype and phenotype datasets by keeping only the overlapping subjects to generate the analysis-ready dataset, which includes 1,440 subjects, with a censoring rate of 62.5%. We are interested in 1) testing genetic covariates that are associated

with the time-to-onset of first AD diagnosis and 2) detecting potential genetic heterogeneity across subpopulations inferred by either baseline covariates or individual genome profiles (represented by a random sample of 200K SNPs), in the presence of left-truncation time and adjust for gender, year of education, study cohort indicator (ROS served as baseline), and first 6 principal components. Note that race is not adjusted because 1,439 out of 1,440 subjects are white. IBS kernel was used to measure genetic similarity, while the choice of similarity kernel for subpopulation similarity depends on the source of heterogeneity, which is elaborated in Table 5.18.

Table 5.18 shows the results of the ROSMAP dataset analysis. To account for multiple testing issues, the p -value thresholds were obtained by controlling the FDR under 10% using the Benjamini-Hochberg procedure Benjamini and Hochberg (1995). Four different kernels were considered to measure genetic similarity: cross-product (CP), IBS, laplacian (Lap), and quadratic (Quad). When in the absence of genetic heterogeneity (scenario S1), APOE4 appeared to be the most significant gene followed by APOC1. When in the presence of genetic heterogeneity (scenarios S2-S4), even though APOE4 and APOC1 were still the most significant genes, their p -values varies across scenarios, with APOE4 having a notably smaller p -value when the Quadratic kernel was used under scenario S3 that considered genetic heterogeneity across gender. This indicates that the effect of APOE4 on the age-to-onset of AD varies between genders.

5.4 Discussion

We developed a suite of multi-marker genetic association and interaction tests based on the additive hazards model. The proposed tests have three major advantages: 1) they fill the research gap that no available genetic association and interaction test for survival times

Table 5.18 Top five genes discovered by $V_{\mathbf{Z}}^c$ and $V_{\mathbf{Z}}^{H,c}$ from ROSMAP dataset by assuming age-to-onset of AD follows additive hazards model. CP, IBS, Lap and Quad are IBS, cross-product, Laplacian and Quadratic kernel, respectively, which measure genetic similarity. Various types of heterogeneity were considered, including no genetic heterogeneity (S1), heterogeneity across education categories (S2), heterogeneity between sexes (S3), and heterogeneity across genetic backgrounds (S4).

Kernel	Scenario	Genes and p-values				
CP	S1	APOE4	APOC1	IGSF23	GRIP1	MIR12117
		3.55E-15	7.26E-10	6.05E-05	6.99E-05	1.06E-04
	S2	APOE4	APOC1	TBCC	IGSF23	MIR12117
		1.99E-14	2.84E-09	6.75E-05	7.16E-05	8.42E-05
	S3	APOE4	APOC1	GRIP1	IGSF23	MIR12117
		3.10E-15	2.44E-09	6.12E-05	1.26E-04	1.37E-04
	S4	APOE4	APOC1	IGSF23	TBCC	HSBP1
		7.63E-14	1.50E-09	7.09E-05	7.26E-05	1.22E-04
IBS	S1	APOE4	APOC1	IGSF23	MIR12117	GRIP1
		7.97E-14	4.59E-09	8.21E-05	8.78E-05	9.65E-05
	S2	APOE4	APOC1	IGSF23	EFEMP2	DCAF12
		3.71E-11	2.50E-08	3.03E-04	3.42E-04	7.19E-04
	S3	APOE4	APOC1	GRIP1	IGSF9B	PLEKHG5
		1.42E-13	9.74E-09	8.82E-05	9.83E-05	1.20E-04
	S4	APOE4	APOC1	IGSF23	MIR12117	GRIP1
		7.53E-14	4.61E-09	8.25E-05	8.79E-05	9.66E-05
Lap	S1	APOE4	APOC1	GRIP1	IGSF23	MIR12117
		4.73E-13	1.52E-08	6.27E-05	7.78E-05	8.72E-04
	S2	APOE4	APOC1	IGSF23	DCAF12	EFEMP2
		4.64E-12	3.86E-08	1.15E-04	2.27E-04	3.43E-04
	S3	APOE4	APOC1	GRIP1	IGSF9B	MIR12117
		8.44E-13	2.77E-08	5.53E-05	1.32E-04	1.39E-04
	S4	APOE4	APOC1	GRIP1	IGSF23	MIR12117
		4.69E-13	1.53E-08	6.27E-05	7.81E-05	8.74E-05
Quad	S1	APOE4	APOC1	GRIP1	MIR12117	GABBR1
		1.11E-15	6.35E-10	6.66E-05	1.12E-04	1.14E-04
	S2	APOE4	APOC1	GRIP1	MIR12117	GABBR1
		6.43E-15	2.78E-09	8.38E-05	8.84E-05	1.10E-04
	S3	APOE4	APOC1	GRIP1	EHHADH-AS1	GABBR1
		2.22E-16	3.10E-09	5.87E-05	7.55E-05	1.33E-04
	S4	APOE4	APOC1	GRIP1	MIR12117	GABBR1
		1.77E-15	6.24E-10	6.64E-05	1.11E-04	1.13E-04
p-value threshold		4.46E-07	8.91E-07	1.34E-06	1.78E-06	2.23E-06

that follow the additive hazards model, 2) they are powerful when the genetic effect is heterogeneous across subpopulations, and 3) they are robust to quadratic confounding effect and can handle competing risk.

The asymptotic null distribution of weighted V statistic can be well approximated by a rescaled χ^2 distribution, $c_0\chi_{d_0}^2$, using the Satterthwaite method (Liu, Lin, and Ghosh, 2007). c_0 is the scaled parameter and d_0 is the χ^2 distribution's degree of freedom estimated by

matching the mean and variance of weighted V statistics and $c_0\chi_{d_0}^2$. In testing $G-G$ and $G-E$ interaction effect, the scaled χ^2 distribution of weighted V statistic has d_0 degree of freedom, which is smaller than pq , the number of $G-G$ interaction terms, when there is a weak or strong correlation between SNPs. Therefore the weighted V statistic is more powerful than regular tests (e.g., LRT, Wald test).

CHAPTER 6 Concluding remarks

6.1 Remarks regarding selection of survival model

In this dissertation, we developed four suites of novel multi-marker association and interaction tests to detect genetic association and interaction with survival outcomes. Three popular survival models were studied: (conventional) Cox PH model, additive hazard model, and accelerated failure time (AFT) model. In a real application, the true survival model underlying the observed data is rarely known, therefore, it would be helpful to determine which model best fits the dataset and the goal of the analysis. Cox PH model is the most popular model that models the hazard rate as a function of predictor variables in a multiplicative manner as follows,

$$\lambda(t|X) = \lambda_0(t) \exp(\boldsymbol{\beta}^T X),$$

where $\lambda_0(t)$ is the non-parametric baseline hazard function and X is the predictor variable. One critical assumption of the Cox PH model is that the predictor variables analyzed should satisfy the PH assumption. In other words, for any two predictor variable values X and X^* , the corresponding hazard ratio is constant over time, as follows

$$\frac{\lambda(t|X)}{\lambda(t|X^*)} = \frac{\lambda_0(t) \exp(\boldsymbol{\beta}^T X)}{\lambda_0(t) \exp(\boldsymbol{\beta}^T X^*)} = \exp(\boldsymbol{\beta}^T (X - X^*)).$$

Violation of PH assumption can be tested based on Schoenfeld residuals. When the PH assumption is violated, the additive hazard model and AFT model are alternatives to the Cox PH model.

Additive hazard model assumes that the predictor affects hazard function in an additive

manner and also assumes constant hazard difference, instead of hazard ratio in the Cox PH model.

$$\lambda(t|X) = \lambda_0(t) + \boldsymbol{\beta}^T X,$$

AFT model models the direct effect of predictors on logarithmic survival time through a linear model,

$$\log(T) = \boldsymbol{\beta}^T X + \epsilon,$$

where ϵ is an independent and identically distributed random variable and is independent of predictor X . A negative $\boldsymbol{\beta}$ indicates that an increase in X would accelerate the onset of the event of interest and a negative value prolongs the event onset. Two important assumptions of the AFT model are the linearity assumption and the constant time ratio assumption. Linearity assumption, similar to that in linear regression models, assumes that logarithmic survival time has a linear relationship with covariates. The constant time ratio assumption assumes that the survival time ratio is constant over time for any given two sets of covariates X and X^* as follows,

$$\frac{T}{T^*} = \frac{\exp(\boldsymbol{\beta}^T X)e^\epsilon}{\exp(\boldsymbol{\beta}^T X^*)e^\epsilon} = \exp(\boldsymbol{\beta}^T (X - X^*)) \Rightarrow T = T^* e^{(\boldsymbol{\beta}^T (X - X^*))}.$$

The appropriateness of the AFT model can be assessed based on Cox-Snell residual plot.

When the survival time is assumed to follow a class of semiparametric transformation model as introduced in Chapter 3. We have the flexibility to choose from different transformation functions controlled by r ($r \geq 0$). For example, we can vary r from 0 to 1.5 in steps of 0.05 (Zeng, Mao, and Lin, 2016), and the transformation function corresponds to the r

value that achieves the highest likelihood is selected. One can also choose from $r = 0$ and $r = 1$, whichever is the closest to the optimal r to achieve better model interpretability.

6.2 Future work

Throughout this dissertation, the proposed four suites of novel association and interaction tests all assume independence between study subjects. This assumption is violated in datasets such as time-to-carries data of second molars in preschool child participants of DDHP (Detroit Dental Health Project) dataset studied in Pak, Li, and Todem (2019), where four tooth-level time-to-ECC were considered for each child participant. So far, a multi-marker genetic association and interaction test is lacking for multivariate survival data. However, the estimation methods for such multivariate survival data were developed in Lin (1994) for the multivariate Cox PH model, and in Zeng, Gao, and Lin (2017) for multivariate interval-censored data, which contribute to the development of the multi-marker test.

In many applications (Kelly and Lim, 2000; Amorim and Cai, 2014), we deal with the recurrent event, where the event of interest might occur repeatedly, instead of once as assumed in this dissertation. A recurrent event subject to interval censoring is panel count data, which has been studied (Zeng and Lin, 2020; Sun and Wei, 2000; Wang, Ma, and Yan, 2013). It is worth developing a multi-marker association and interaction test for such panel count data since this type of data is frequently encountered in complex human diseases such as bladder tumor study (Byar, 1980) and skin cancer (Bailey et al., 2010).

APPENDICES

APPENDIX A: Proofs for Chapter 2

Proof of Theorem 1

We prove Theorem 1 following the proof of Theorem 3 in Wei and Lu (2017). By the expression (2), we have

$$nV^* = \sum_{t=1}^{\infty} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_t^*(\mathbf{G}_i) M_i \right)^2. \quad (6.1)$$

Under the null hypothesis that T is independent of \mathbf{G} ,

$$E[\psi_t^*(\mathbf{G}_i) M_i] = E[\psi_t^*(\mathbf{G}_i)] E M_i = 0 \quad (6.2)$$

and

$$E[\psi_t^*(\mathbf{G}_i) M_i \psi_s^*(\mathbf{G}_i) M_i] = E[\psi_t^*(\mathbf{G}_i) \psi_s^*(\mathbf{G}_i)] E(M_i^2) = \begin{cases} \eta_t \lambda & \text{if } t = s \\ 0 & \text{otherwise} \end{cases} \quad (6.3)$$

These results, in conjunction with the multivariate central limit theorem, imply that for any finite index set $\mathcal{I} = \{i_1, \dots, i_K\}$, the multivariate random variable $\{\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_t^*(\mathbf{G}_i) M_i\}_{t \in \mathcal{I}}$ converges in distribution to a zero-mean multivariate normal random variable with a covariance matrix whose lk -th element is $\lambda \eta_l I(l = k)$ ($l = 1, \dots, K; k = 1, \dots, K$).

In addition, we have

$$\begin{aligned} \sum_{t=1}^{\infty} E[\{\psi_t^*(\mathbf{G}_1) M_1\}^2] &= \lambda \sum_{t=1}^{\infty} \eta_t \\ &= \lambda E[\tilde{f}(\mathbf{G}_i, \mathbf{G}_j)] < \infty. \end{aligned}$$

By Theorem 2.13.1 in van der Vaart and Wellner (1996), the countably infinite sequence of functions $\{\psi_t^*(\cdot) M(\cdot)\}$, where $M(N(\cdot), Y(\cdot)) \equiv N(\infty) - \int_0^\infty Y(s) d\Lambda(s)$, is a Donsker

class. Therefore, the empirical process, $\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_t^*(\mathbf{G}_i)M_i$, converges weakly to a zero-mean Gaussian process Z_t with covariance function, $cov(Z_t, Z_s) = \eta_t \lambda I(t = s)$. By continuous mapping theorem, we have

$$nV^* \rightsquigarrow \sum_{t=1}^{\infty} Z_t^2 = \lambda \sum_{t=1}^{\infty} \eta_t \chi_{1t}^2,$$

where χ_{1t}^2 's are independent chi-square random variables with degree 1.

Proof of Theorem 2

Similar to (6.1), $nV_{\mathbf{Z}}^*$ can be re-written as

$$nV_{\mathbf{Z}}^* = \sum_{t=1}^{\infty} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) M_{\mathbf{Z},i} \right)^2, \quad (6.4)$$

where $\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) = \nu_t^{0.5} \phi_t(\mathbf{G}_i, \mathbf{Z}_i)$. Under the null hypothesis that T is independent of \mathbf{G} given \mathbf{Z} ,

$$\begin{aligned} E[\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) M_{\mathbf{Z},i}] &= E[\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) E\{M_{\mathbf{Z},i} | \mathbf{G}_i, \mathbf{Z}_i\}] \\ &= E[\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) E\{M_{\mathbf{Z},i} | \mathbf{Z}_i\}] \\ &= 0 \end{aligned}$$

We are then going to prove that $E[\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) M_{\mathbf{Z},i} \phi_s^*(\mathbf{G}_i, \mathbf{Z}_i) M_{\mathbf{Z},i}] = \xi \nu_t I(t = s)$. Indeed, when \mathbf{G} is independent of \mathbf{Z} , it is easy to show that $\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j) = f(\mathbf{G}_i, \mathbf{G}_j) - E[f(\mathbf{G}_i, \mathbf{G}_j) | \mathbf{G}_i] - E[f(\mathbf{G}_i, \mathbf{G}_j) | \mathbf{G}_j] + E[f(\mathbf{G}_i, \mathbf{G}_j)]$. When $\mathbf{G} = \mathbf{a} + \mathbf{B}^T \mathbf{Z} + \mathbf{e}$ and $f(\mathbf{G}_i, \mathbf{G}_j) =$

$$\mathbf{G}_i^T \mathbf{G}_j,$$

$$\begin{aligned} \tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j) &= \mathbf{G}_i^T \mathbf{G}_j - \mathbf{G}_i^T E(\mathbf{G}_j | \mathbf{Z}_j) - E(\mathbf{G}_i^T | \mathbf{Z}_i) \mathbf{G}_j + E(\mathbf{G}_i^T | \mathbf{Z}_i) E(\mathbf{G}_j | \mathbf{Z}_j) \\ &= \mathbf{e}_i^T \mathbf{e}_j. \end{aligned}$$

In both cases, $\tilde{f}_{\mathbf{Z}}(\mathbf{G}_i, \mathbf{G}_j)$ is independent of $(\mathbf{Z}_i, \mathbf{Z}_j)$. As a result, $\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i)$ is also independent of \mathbf{Z}_i . Thus,

$$\begin{aligned} E[\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) M_{\mathbf{Z},i} \phi_s^*(\mathbf{G}_i, \mathbf{Z}_i) M_{\mathbf{Z},i}] &= E[\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) \phi_s^*(\mathbf{G}_i, \mathbf{Z}_i) E\{M_{\mathbf{Z},i}^2 | \mathbf{G}_i, \mathbf{Z}_i\}] \\ &= E[\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) \phi_s^*(\mathbf{G}_i, \mathbf{Z}_i) E\{M_{\mathbf{Z},i}^2 | \mathbf{Z}_i\}] \\ &= E[\phi_t^*(\mathbf{G}_i, \mathbf{Z}_i) \phi_s^*(\mathbf{G}_i, \mathbf{Z}_i)] E[E\{M_{\mathbf{Z},i}^2 | \mathbf{Z}_i\}] \\ &= \nu_t I(t = s) E\{M_{\mathbf{Z},i}^2\} \\ &= \xi \nu_t I(t = s). \end{aligned}$$

The rest of the proof follows the same arguments as in the proof of Theorem 1.

APPENDIX B: Proofs for Chapter 4

Derivation of the large-sample null distribution of weighted V statistic based on AFT model

Write \mathbf{E}_1 as $\mathbf{E}_1 = (\mathbf{E}_{11}, \dots, \mathbf{E}_{1n})$. Then

$$\begin{aligned}
 \mathbf{E}_1 \tilde{\mathbf{M}} &= \sum_{i=1}^n \mathbf{E}_{1i} \tilde{M}_i \\
 &= \sum_{i=1}^n \mathbf{E}_{1i} \int_{-\infty}^{\infty} \left\{ dN_i(\hat{\boldsymbol{\beta}}, t) - \frac{\nu_j(\hat{\boldsymbol{\beta}}, t) Y_j(\hat{\boldsymbol{\beta}}, t)}{\sum_{j=1}^n \nu_j(\hat{\boldsymbol{\beta}}, t) Y_j(\hat{\boldsymbol{\beta}}, t)} \sum_{j=1}^n dN_j(\hat{\boldsymbol{\beta}}, t) \right\} \\
 &= \sum_{i=1}^n \int_{-\infty}^{\infty} \left\{ \mathbf{E}_{1i} - \frac{\sum_{j=1}^n \mathbf{E}_{1j} \nu_j(\hat{\boldsymbol{\beta}}, t) Y_j(\hat{\boldsymbol{\beta}}, t)}{\sum_{j=1}^n \nu_j(\hat{\boldsymbol{\beta}}, t) Y_j(\hat{\boldsymbol{\beta}}, t)} \right\} dN_i(\hat{\boldsymbol{\beta}}, t) \\
 &\equiv Q_n(\hat{\boldsymbol{\beta}}).
 \end{aligned}$$

Under the null, $\hat{\boldsymbol{\beta}}$ is $n^{1/2}$ -consistent for $\boldsymbol{\beta}$ (Lai and Ying, 1991). Using similar arguments to the proof of Theorem 1(ii) in Lai and Ying (1991), it can be shown that

$$n^{-1/2} \mathbf{E}_1 (\tilde{\mathbf{M}} - \hat{\mathbf{M}}) = n^{-1/2} \{Q_n(\hat{\boldsymbol{\beta}}) - Q_n(\boldsymbol{\beta})\} = B n^{1/2} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + o_p(1 + n^{1/2} \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|), \quad (6.5)$$

where B is the asymptotic slope matrix of $n^{-1} Q_n(\boldsymbol{\beta})$. Following Zeng and Lin (2006), we use a least squares method to estimate B . Specifically, let $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + n^{-1/2} \mathbf{W}$, where $\mathbf{W} \sim N(\mathbf{0}, I_q)$ and I_q is a $q \times q$ identity matrix. (6.5) implies that

$$n^{-1/2} \{Q_n(\tilde{\boldsymbol{\beta}}) - Q_n(\hat{\boldsymbol{\beta}})\} = B n^{1/2} (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) + o_p(1) = B \mathbf{W} + o_p(1).$$

The least squares method has the following steps:

Step 1: Generate L , say 10,000, independent realizations of \mathbf{W} , denoted by $\mathbf{W}_1, \dots, \mathbf{W}_L$.

Step 2: Calculate $n^{-1/2}\{Q_n(\hat{\boldsymbol{\beta}} + n^{-1/2}\mathbf{W}_l) - Q_n(\hat{\boldsymbol{\beta}})\}$ ($l = 1, \dots, L$).

Step 3: For $j = 1, \dots, m$, regress $n^{-1/2}\{Q_n(\hat{\boldsymbol{\beta}} + n^{-1/2}\mathbf{W}_l) - Q_n(\hat{\boldsymbol{\beta}})\}$ onto \mathbf{W}_l ($l = 1, \dots, L$) to estimate B_j , the j -th row of B , using the least squares estimation.

Denote the estimator of B by \hat{B} . Recall that $\hat{\boldsymbol{\beta}}$ is obtained from Eq. (5) in Chiou and Xu (2017) with log-rank weights,

$$U_n(\boldsymbol{\beta}) \equiv \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} \left\{ \mathbf{z}_i - \frac{\sum_{j=1}^n \mathbf{z}_j \nu_j(\boldsymbol{\beta}, t) Y_j(\boldsymbol{\beta}, t)}{\sum_{j=1}^n \nu_j(\boldsymbol{\beta}, t) Y_j(\boldsymbol{\beta}, t)} \right\} dN_i(\boldsymbol{\beta}, t) = \mathbf{0}.$$

Chiou and Xu (2017) showed that

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -A^{-1}n^{1/2}U_n(\boldsymbol{\beta}) + o_p(1), \quad (6.6)$$

where A is the asymptotic slope matrix of $U_n(\boldsymbol{\beta})$. A can be estimated using a similar method to that for estimating B . Specifically, we carry out the following steps:

Step I: Generate \tilde{L} , say 10,000, independent realizations of $\mathbf{S} \sim N(\mathbf{0}, I_q)$, denoted by $\mathbf{S}_1, \dots, \mathbf{S}_{\tilde{L}}$.

Step II: Calculate $n^{1/2}U_n(\hat{\boldsymbol{\beta}} + n^{-1/2}\mathbf{S}_l)$ ($l = 1, \dots, \tilde{L}$).

Step III: For $j = 1, \dots, q$, regress $n^{1/2}U_n(\hat{\boldsymbol{\beta}} + n^{-1/2}\mathbf{S}_l)$ onto \mathbf{S}_l ($l = 1, \dots, \tilde{L}$) to estimate A_j , the j -th row of A , using the least squares estimation.

Denote the estimator of A by \hat{A} . Combining (6.5) and (6.6), we have

$$\mathbf{E}_1(\tilde{\mathbf{M}} - \hat{\mathbf{M}}) = -BA^{-1}nU_n(\boldsymbol{\beta}) + o_p(n^{1/2}). \quad (6.7)$$

Let $M_i(t) = N_i(\boldsymbol{\beta}, t) - \int_{-\infty}^t \nu_i(\boldsymbol{\beta}, u) Y_i(\boldsymbol{\beta}, u) d\Lambda_\varepsilon(u)$, which a martingale under the null. It

is easy to show that

$$\begin{aligned} U_n(\boldsymbol{\beta}) &= \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} \left\{ \mathbf{Z}_i - \frac{\sum_{j=1}^n \mathbf{Z}_j \nu_j(\boldsymbol{\beta}, t) Y_j(\boldsymbol{\beta}, t)}{\sum_{j=1}^n \nu_j(\boldsymbol{\beta}, t) Y_j(\boldsymbol{\beta}, t)} \right\} dM_i(\boldsymbol{\beta}, t) \\ &= \frac{1}{n} Z^T \int_{-\infty}^{\infty} (I_n - V(\boldsymbol{\beta}, t) \mathbf{1}^T) d\mathbf{M}(t), \end{aligned} \quad (6.8)$$

where $\mathbf{1}$ is a n -dimension vector of 1's, $Z = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^T$, and $V(\boldsymbol{\beta}, t)$ is a $n \times 1$ vector $V(\boldsymbol{\beta}, t) = (\nu_1(\boldsymbol{\beta}, t) Y_1(\boldsymbol{\beta}, t), \dots, \nu_n(\boldsymbol{\beta}, t) Y_n(\boldsymbol{\beta}, t))^T \{\sum_{j=1}^n \nu_j(\boldsymbol{\beta}, t) Y_j(\boldsymbol{\beta}, t)\}^{-1}$, and $\mathbf{M}(t) = (M_1(t), \dots, M_n(t))^T$. It is also easy to show that

$$\hat{\mathbf{M}} = \int_{-\infty}^{\infty} (I_n - V(\boldsymbol{\beta}, t) \mathbf{1}^T) d\mathbf{M}(t). \quad (6.9)$$

Combining (6.7), (6.8) and (6.9), we have

$$\begin{aligned} \mathbf{E}_1 \tilde{\mathbf{M}} &= \mathbf{E}_1 \hat{\mathbf{M}} + \mathbf{E}_1 (\tilde{\mathbf{M}} - \hat{\mathbf{M}}) \\ &= \int_{-\infty}^{\infty} (\mathbf{E}_1 - BA^{-1} Z^T) (I_n - V(\boldsymbol{\beta}, t) \mathbf{1}^T) d\mathbf{M}(t) + o_p(n^{1/2}) \end{aligned}$$

Since $\mathbf{M}(t)$ is a vector of independent martingales under the null, whose compensators are $\boldsymbol{\Lambda}(t) = \int_{-\infty}^t d\boldsymbol{\Lambda}(t) \equiv (\int_{-\infty}^t d\Lambda_1(t), \dots, \int_{-\infty}^t d\Lambda_n(t))^T$, where $d\Lambda_i(t) = \nu_i(\boldsymbol{\beta}, t) Y_i(\boldsymbol{\beta}, t) d\Lambda_\varepsilon(t)$ ($i = 1, \dots, n$), we have by the martingale theory (Fleming and Harrington, 1991, Chapters 2 and 5) that

$$Cov(\mathbf{E}_1 \tilde{\mathbf{M}}) \approx \int_{-\infty}^{\infty} (\mathbf{E}_1 - BA^{-1} Z^T) (I_n - V(\boldsymbol{\beta}, t) \mathbf{1}^T) \text{diag}(d\boldsymbol{\Lambda}(t)) (I_n - \mathbf{1} V(\boldsymbol{\beta}, t)^T) (\mathbf{E}_1^T - Z A^{-T} B^T),$$

where $\text{diag}(\mathbf{a})$ represents a diagonal matrix with \mathbf{a} as the diagonal, and that $\mathbf{E}_1\tilde{\mathbf{M}}$ is approximately multivariate normal with mean zero. An estimator of $Cov(\mathbf{E}_1\tilde{\mathbf{M}})$ is

$$\hat{C}ov(\mathbf{E}_1\tilde{\mathbf{M}}) = \int_{-\infty}^{\infty} (\mathbf{E}_1 - \hat{B}\hat{A}^{-1}Z^T)(I_n - V(\hat{\boldsymbol{\beta}}, t)\mathbf{1}^T) \text{diag}(d\tilde{\boldsymbol{\Lambda}}(t))(I_n - \mathbf{1}V(\hat{\boldsymbol{\beta}}, t)^T)(\mathbf{E}_1^T - Z\hat{A}^{-T}\hat{B}^T),$$

where $d\tilde{\boldsymbol{\Lambda}}(t) = (d\tilde{\Lambda}_1(t), \dots, d\tilde{\Lambda}_n(t))^T$ with each element as follows,

$$d\tilde{\Lambda}_i(t) = \nu_i(\hat{\boldsymbol{\beta}}, t)Y_i(\hat{\boldsymbol{\beta}}, t) \sum_{j=1}^n dN_j(\hat{\boldsymbol{\beta}}, t) \left\{ \sum_{j=1}^n \nu_j(\hat{\boldsymbol{\beta}}, t)Y_j(\hat{\boldsymbol{\beta}}, t) \right\}^{-1}. \quad (6.10)$$

By algebra, we can simplify $\hat{C}ov(\mathbf{E}_1\tilde{\mathbf{M}})$ as

$$\hat{C}ov(\mathbf{E}_1\tilde{\mathbf{M}}) = \int_{-\infty}^{\infty} (\mathbf{E}_1 - \hat{B}\hat{A}^{-1}Z^T) \{ \text{diag}(d\tilde{\boldsymbol{\Lambda}}(t)) - V(\hat{\boldsymbol{\beta}}, t)\mathbf{1}^T d\mathbf{N}(\hat{\boldsymbol{\beta}}, t)V(\hat{\boldsymbol{\beta}}, t)^T \} (\mathbf{E}_1^T - Z\hat{A}^{-T}\hat{B}^T), \quad (6.11)$$

where $\mathbf{N}(\hat{\boldsymbol{\beta}}, t) = (N_1(\hat{\boldsymbol{\beta}}, t), \dots, N_n(\hat{\boldsymbol{\beta}}, t))^T$. Take an eigendecomposition of $\hat{C}ov(\mathbf{E}_1\tilde{\mathbf{M}})$,

$$\hat{C}ov(\mathbf{E}_1\tilde{\mathbf{M}}) = \Gamma \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix} \Gamma^T,$$

where Γ is a $m \times m$ orthogonal matrix. Together with the asymptotic normality of $\mathbf{E}_1\tilde{\mathbf{M}}$,

we have

$$\mathbf{E}_1\tilde{\mathbf{M}} \stackrel{d}{\approx} \Gamma \begin{bmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_m^{1/2} \end{bmatrix} \boldsymbol{\chi},$$

where $\boldsymbol{\chi} \equiv (\chi_{11}, \dots, \chi_{1m})^T \sim N(\mathbf{0}, I_m)$. Thus,

$$V_{\mathbf{Z}} = \tilde{\mathbf{M}}^T \mathbf{K} \tilde{\mathbf{M}} = \tilde{\mathbf{M}}^T \mathbf{E}_1^T \mathbf{E}_1 \tilde{\mathbf{M}} \stackrel{d}{\approx} \sum_{j=1}^m \lambda_j \chi_{1j}^2,$$

where χ_{1j}^2 's are independent chi-square variables with degree 1.

APPENDIX C: Proofs for Chapter 5

Derivation of the large-sample null distribution of weighted V statistic based on additive hazard model

Consider a general scenario where the hazard function of survival time T that is associated with a possibly time-varying d -dimensional covariate $\mathbf{Z}(t)$ is as follows,

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) + \boldsymbol{\beta}^T \mathbf{Z}(t).$$

The proposed weighted V test statistic is $V_{\mathbf{Z}} = \tilde{\mathbf{M}}^T \mathbf{F} \tilde{\mathbf{M}}$, where $\tilde{\mathbf{M}} = (\tilde{M}_1, \dots, \tilde{M}_n)^T$ and $\tilde{M}_i = M_i(\hat{\Lambda}_0(\hat{\boldsymbol{\beta}}, t), \hat{\boldsymbol{\beta}})$ ($i = 1, \dots, n$) as follows,

$$\tilde{M}_i = N_i(\infty) - \int_0^\infty Y_i(t) \left\{ \frac{\sum_{j=1}^n dN_j(t)}{\sum_{j=1}^n Y_j(t)} - \hat{\boldsymbol{\beta}}^T \bar{\mathbf{Z}}(t) dt + \hat{\boldsymbol{\beta}}^T \mathbf{Z}_i(t) dt \right\},$$

where $\bar{\mathbf{Z}}(t) = \sum_{j=1}^n Y_j(t) \mathbf{Z}_j(t) / \sum_{j=1}^n Y_j(t) = \mathbf{Z}^T(t) V(t)$, with $V(t) = Y(t) (\mathbf{1}^T Y(t))^{-1}$. $V(t) = (V_1(t), \dots, V_n(t))^T$ and $Y(t) = (Y_1(t), \dots, Y_n(t))^T$. $\hat{\boldsymbol{\beta}}$ is $n^{1/2}$ -consistent estimator of $\boldsymbol{\beta}$ as shown in equation (2.8) in Lin and Ying (1994). By first-order Taylor expansion of $\tilde{\mathbf{M}}$ around $\boldsymbol{\beta}_0$, true parameter value of $\boldsymbol{\beta}$, we have

$$\tilde{\mathbf{M}} \approx \hat{\mathbf{M}} + \frac{\partial \tilde{\mathbf{M}}}{\partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \quad (6.12)$$

The influence function of $\hat{\boldsymbol{\beta}}$ is

$$\begin{aligned}
(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= D^{-1}U(\boldsymbol{\beta}_0) + o_p\left(\frac{1}{\sqrt{n}}\right) \\
&\approx D^{-1} \sum_{i=1}^n \int_0^\infty \{\mathbf{Z}_i(t) - \bar{\mathbf{Z}}(t)\} dM_i(t) \\
&= D^{-1} \int_0^\infty \{\mathbf{Z}(t)^T - \bar{\mathbf{Z}}(t)\mathbf{1}_n^T\} d\mathbf{M}(t), \tag{6.13}
\end{aligned}$$

where $\mathbf{Z}(t) = (\mathbf{Z}_1(t), \dots, \mathbf{Z}_n(t))^T$ is a $n \times d$ covariate matrix at time t , $\bar{\mathbf{Z}}(t)$ is a d vector at time t , $\mathbf{1}_n$ is $n \times 1$ vector with each element equals 1, and $\mathbf{M}(t) = (M_1(t), \dots, M_n(t))$. $U(\boldsymbol{\beta}_0)$ is estimating function of $\boldsymbol{\beta}_0$ according to equation (2.7) in Lin and Ying (1994). When covariate is time-independent, $\mathbf{Z}(t) = \mathbf{Z}$, we have

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = D^{-1}\mathbf{Z}^T \int_0^\infty (I_n - V(t)\mathbf{1}^T) d\mathbf{M}(t).$$

Matrix D has the following form

$$\begin{aligned}
D &= \sum_{j=1}^n \int_0^\infty Y_j(t) \{\mathbf{Z}_j(t) - \bar{\mathbf{Z}}(t)\}^{\otimes 2} dt \\
&= \int_0^\infty \mathbf{Z}(t)^T \left(\text{diag}(Y(t)) - Y(t)V(t)^T \right) \mathbf{Z}(t) dt, \tag{6.14}
\end{aligned}$$

where $a^{\otimes 2} = aa^T$. When covariate is time-independent, $\mathbf{Z}(t) = \mathbf{Z}$, we have

$$\begin{aligned}
D &= \mathbf{Z}^T \left[\text{diag} \left(\int_0^\infty Y(t) dt \right) - \int_0^\infty Y(t)V(t)^T dt \right] \mathbf{Z} \\
&= \mathbf{Z}^T \mathbf{W} \mathbf{Z},
\end{aligned}$$

where $\mathbf{W} = \text{diag}(\int_0^\infty Y(t)dt) - \int_0^\infty Y(t)V(t)^T dt$.

$\hat{\mathbf{M}}$ in equation 6.12 can be written as

$$\begin{aligned}\hat{\mathbf{M}} &= \int_0^\infty d\hat{\mathbf{M}}(t) \\ &= \left\{ \int_0^\infty dM_i(t) - Y_i(t) \frac{\sum_{j=1}^n dM_j(t)}{\sum_{j=1}^n Y_j(t)} \right\}_{i=1}^n \\ &= \int_0^\infty (I_n - V(t)\mathbf{1}^T) d\mathbf{M}(t)\end{aligned}$$

The first-order derivative of $\tilde{\mathbf{M}}$ around β_0 ($\partial\tilde{\mathbf{M}}/\partial\beta^T|_{\beta=\beta_0}$) in equation 6.12 can be written as

$$\begin{aligned}\frac{\partial\tilde{\mathbf{M}}}{\partial\beta^T}|_{\beta=\beta_0} &= \left\{ \int_0^\infty Y_i(t)(\bar{\mathbf{Z}}(t)^T - \mathbf{Z}_i(t)^T) dt \right\}_{i=1}^n \\ &= - \int_0^\infty (\text{diag}(Y(t) - Y(t)V(t)^T) \mathbf{Z}(t) dt \\ &= -\mathcal{G}.\end{aligned}\tag{6.15}$$

When covariate is time-independent, we have

$$\begin{aligned}\frac{\partial\tilde{\mathbf{M}}}{\partial\beta^T}|_{\beta=\beta_0} &= - \int_0^\infty (\text{diag}(Y(t)) - Y(t)V(t)^T) dt \mathbf{Z} \\ &= -\mathbf{WZ}\end{aligned}$$

By plugging 6.13, 6.14, and 6.15 into equation 6.12, we can approximate $\tilde{\mathbf{M}}$ as follows,

$$\begin{aligned}\tilde{\mathbf{M}} &\approx \int_0^\infty (I_n - V(t)\mathbf{1}^T) d\mathbf{M}(t) - \mathcal{G}D^{-1} \int_0^\infty \mathbf{Z}(t)^T (I_n - V(t)\mathbf{1}^T) d\mathbf{M}(t) \\ &= \int_0^\infty (I_n - \mathcal{G}D^{-1})(I_n - V(t)\mathbf{1}^T) d\mathbf{M}(t).\end{aligned}$$

When covariate is time-independent, $\mathbf{Z}(t) = \mathbf{Z}$, the approximation of $\tilde{\mathbf{M}}$ can be simplified as

$$\tilde{\mathbf{M}} \approx (I_n - \mathbf{WZ}(\mathbf{Z}^T \mathbf{WZ})^{-1}) \int_0^\infty (I_n - V(t)\mathbf{1}^T) d\mathbf{M}(t).$$

Since $\mathbf{M}(t)$ is a vector of independent martingales under the null hypothesis, whose compensators are $\mathbf{\Lambda}(t) = \int_0^t d\mathbf{\Lambda}(t) = (\int_0^t d\Lambda_1(t), \dots, \int_0^t d\Lambda_n(t))^T$, where $d\Lambda_i(t) = Y_i(t)d\Lambda(t)$ ($i = 1, \dots, n$). By martingale theory (Fleming and Harrington, 1991, Chapters 2 and 5), $\tilde{\mathbf{M}}$ is approximately $MVN(\mathbf{0}, Cov(\tilde{\mathbf{M}}))$, where $Cov(\tilde{\mathbf{M}})$ is approximated as follows,

$$Cov(\tilde{\mathbf{M}}) \approx \int_0^\infty (I_n - \mathcal{G}D^{-1}\mathbf{Z}(t)^T)(I_n - V(t)\mathbf{1}^T) \text{diag}(d\mathbf{\Lambda}(t))(I_n - \mathbf{1}V(t)^T)(I_n - \mathbf{Z}(t)D^{-1}\mathcal{G}^T),$$

where $d\mathbf{\Lambda}(t)$ is approximates $Cov(d\mathbf{M}(t))$. When $\mathbf{Z}(t)$ is independent of time, $\mathbf{Z}(t) = \mathbf{Z}$, we have

$$\begin{aligned} Cov(\tilde{\mathbf{M}}) &\approx (I_n - \mathbf{WZ}(\mathbf{Z}^T \mathbf{WZ})^{-1}\mathbf{Z}^T) \left[\int_0^\infty (I_n - V(t)\mathbf{1}^T) \text{diag}(d\mathbf{\Lambda}(t))(I_n - \mathbf{1}V(t)^T) \right] \\ &\quad (I_n - \mathbf{Z}(\mathbf{Z}^T \mathbf{WZ})^{-1}\mathbf{Z}^T \mathbf{W}^T), \end{aligned}$$

where $d\mathbf{\Lambda}(t)$ is a $n \times 1$ vector with i -th element equals $\{Y_i(t)(d\Lambda_0(t) + \beta_0^T \mathbf{Z}_i(t)dt)\}$ ($i = 1, \dots, n$). Replace β_0 with its estimator $\hat{\beta}$, we can obtain an estimator of $Cov(\tilde{\mathbf{M}})$ as follows,

$$\begin{aligned} \hat{Cov}(\tilde{\mathbf{M}}) &= (I_n - \mathbf{WZ}(\mathbf{Z}^T \mathbf{WZ})^{-1}\mathbf{Z}^T) \left[\int_0^\infty (I_n - V(t)\mathbf{1}^T) \text{diag}(d\hat{\mathbf{\Lambda}}(t))(I_n - \mathbf{1}V(t)^T) \right] \\ &\quad (I_n - \mathbf{Z}(\mathbf{Z}^T \mathbf{WZ})^{-1}\mathbf{Z}^T \mathbf{W}^T), \end{aligned}$$

where $d\hat{\Lambda}(t) = Y(t)V(t)^T dN(t) + (\text{diag}(Y(t)) - Y(t)V(t)^T)\mathbf{Z}(t)\hat{\boldsymbol{\beta}}dt$.

Write $\mathbf{F} = \mathbf{E}_1^T \mathbf{E}_1$ for some $m \times n$ matrix \mathbf{E}_1 , where m ($1 \leq m \leq n$) is the rank of \mathbf{F} . Then the weighted V statistic can be written as $V_{\mathbf{Z}} = (\mathbf{E}_1 \tilde{\mathbf{M}})^T (\mathbf{E}_1 \tilde{\mathbf{M}})$, with $\mathbf{E}_1 \tilde{\mathbf{M}} \sim N(\mathbf{0}, \hat{Cov}(\mathbf{E}_1 \tilde{\mathbf{M}}))$. Take an eigendecomposition of $\hat{Cov}(\mathbf{E}_1 \tilde{\mathbf{M}})$,

$$\hat{Cov}(\mathbf{E}_1 \tilde{\mathbf{M}}) = \Gamma \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix} \Gamma^T,$$

where $\{\lambda_i\}_{i=1}^m$ are eigenvalues and Γ is a $m \times m$ orthogonal matrix. Together with the asymptotic normality of $\mathbf{E}_1 \tilde{\mathbf{M}}$, we have

$$\mathbf{E}_1 \tilde{\mathbf{M}} \stackrel{d}{\approx} \Gamma \begin{bmatrix} \lambda_1^{1/2} & & \\ & \ddots & \\ & & \lambda_m^{1/2} \end{bmatrix} \boldsymbol{\chi},$$

where $\boldsymbol{\chi} \equiv (\chi_{11}, \dots, \chi_{1m})^T \sim N(\mathbf{0}, I_m)$. Thus,

$$V_{\mathbf{Z}} = \tilde{\mathbf{M}}^T \mathbf{F} \tilde{\mathbf{M}} = (\mathbf{E}_1 \tilde{\mathbf{M}})^T \mathbf{E}_1 \tilde{\mathbf{M}} \stackrel{d}{\approx} \sum_{j=1}^m \lambda_j \chi_{1j}^2,$$

where χ_{1j}^2 's are independent chi-square variables with degree 1.

BIBLIOGRAPHY

BIBLIOGRAPHY

- Amorim, L.D., and J. Cai. 2014. “Modelling recurrent events: a tutorial for analysis in epidemiology.” *International Journal of Epidemiology* 44:324–333.
- Andersen, P.K., and R.D. Gill. 1982. “Cox’s Regression Model for Counting Processes: A Large Sample Study.” *The Annals of Statistics* 10:1100 – 1120.
- Azzato, E.M., P.D. Pharoah, P. Harrington, D.F. Easton, D. Greenberg, N.E. Caporaso, S.J. Chanock, R.N. Hoover, G. Thomas, D.J. Hunter, and P. Kraft. 2010. “A Genome-Wide Association Study of Prognosis in Breast Cancer.” *Cancer Epidemiology and Prevention Biomarkers* 19:1140–1143.
- Bailey, H.H., K. Kim, A.K. Verma, K. Sielaff, P.O. Larson, S. Snow, T. Lenaghan, J.L. Viner, J. Douglas, N.E. Dreckschmidt, M. Hamielec, M. Pomplun, H.H. Sharata, D. Puchalsky, E.R. Berg, T.C. Havighurst, and P.P. Carbone. 2010. “A Randomized, Double-Blind, Placebo-Controlled Phase 3 Skin Cancer Prevention Study of alpha-Difluoromethylornithine in Subjects with Previous History of Skin Cancer.” *Cancer Prevention Research* 3:35–47.
- Beer, D., S. Kardia, C.C. Huang, T. Giordano, A. Levin, D. Misek, L. Lin, G. Chen, G. Tarek, D. Thomas, M. Lizyness, R. Kuick, S. Hayasaka, J. Taylor, M. Iannettoni, M. Orringer, and S. Hanash. 2002. “Gene-Expression Profiles Predict Survival of Patients with Lung Adenocarcinoma.” *Nature medicine* 8:816–24.
- Benjamini, Y., and Y. Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B (Methodological)* 57:289–300.
- Bennett, D.A., A.S. Buchman, P.A. Boyle, L.L. Barnes, R.S. Wilson, and J.A. Schneider. 2018. “Religious Orders Study and Rush Memory and Aging Project.” *Journal of Alzheimer’s disease : JAD* 64:S161—S189.
- Beyene, J., D. Tritchler, J.L. Asimit, and J.S. Hamid. 2009. “Gene- or region-based analysis of genome-wide association studies.” *Genetic Epidemiology* 33:S105–S110.
- Beyersmann, J., A. Latouche, A. Buchholz, and M. Schumacher. 2009. “Simulating competing risks data in survival analysis.” *Statistics in Medicine* 28:956–971.
- Byar, D.P. 1980. *The Veterans Administration Study of Chemoprophylaxis for Recurrent Stage I Bladder Tumours: Comparisons of Placebo, Pyridoxine and Topical Thiotepa*, Boston, MA: Springer US. pp. 363–370.

- Cai, T., G. Tonini, and X. Lin. 2011. “Kernel Machine Approach to Testing the Significance of Multiple Genetic Markers for Risk Prediction.” *Biometrics* 67:975–986.
- Chen, H., T. Lumley, J. Brody, N.L. Heard-Costa, C.S. Fox, L.A. Cupples, and J. Dupuis. 2014. “Sequence Kernel Association Test for Survival Traits.” *Genetic Epidemiology* 38:191–197.
- Chen, H.Y., S.L. Yu, C.H. Chen, G.C. Chang, C.Y. Chen, A. Yuan, C.L. Cheng, C.H. Wang, H.J. Terng, S.F. Kao, W.K. Chan, H.N. Li, C.C. Liu, S. Singh, W.J. Chen, J.J. Chen, and P.C. Yang. 2007. “A Five-Gene Signature and Clinical Outcome in Non-Small-Cell Lung Cancer.” *New England Journal of Medicine* 356:11–20.
- Chen, J., W. Chen, N. Zhao, M.C. Wu, and D.J. Schaid. 2016. “Small Sample Kernel Association Tests for Human Genetic and Microbiome Association Studies.” *Genetic Epidemiology* 40:5–19.
- Chiou, S.H., and G. Xu. 2017. “Rank-Based Estimation for Semiparametric Accelerated Failure Time Model under Length-Biased Sampling.” *Statistics and Computing* 27:483–500.
- Cristianini, N., and J. Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.
- Davies, R.B. 1980. “Algorithm AS 155: The Distribution of a Linear Combination of χ^2 Random Variables.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29:323–333.
- Emond, M.J., T. Louie, J. Emerson, W. Zhao, R.A. Mathias, M.R. Knowles, F.A. Wright, M.J. Rieder, H.K. Tabor, D.A. Nickerson, K.C. Barnes, R.L. Gibson, M.J. Bamshad, L. National Heart, B.I.N.G.E.S. Project, and L. GO. 2012. “Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic *Pseudomonas aeruginosa* infection in cystic fibrosis.” *Nature Genetics* 44:886–889.
- Fleming, T.R., and D.P. Harrington. 1991. *Counting Processes and Survival Analysis*. John Wiley & Sons.
- Galvan, A., J.P. Ioannidis, and T.A. Dragani. 2010. “Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer.” *Trends in Genetics* 26:132 – 141.
- Goeman, J.J., J. Oosting, A.M. Cleton-Jansen, J.K. Anninga, and H.C. van Houwelingen. 2005. “Testing association of a pathway with survival using gene expression data.” *Bioinformatics* 21:1950–1957.
- Hofmann, T., B. Schölkopf, and A.J. Smola. 2008. “Kernel methods in machine learning.” *The Annals of Statistics* 36:1171 – 1220.

- Huang, K.L., E. Marcora, A. Pimenova, A. Narzo, M. Kapoor, S.C. Jin, O. Harari, S. Bertelsen, B. Fairfax, J. Czajkowski, V. Chouraki, B. Grenier-Boley, C. Bellenguez, Y. Deming, A. McKenzie, T. Raj, A. Renton, J. Budde, A. Smith, and A. Goate. 2017. “A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer’s disease.” *Nature neuroscience* 20.
- Huang, Y.T., R.S. Heist, L.R. Chirieac, X. Lin, V. Skaug, S. Zienolddiny, A. Haugen, M.C. Wu, Z. Wang, L. Su, K. Asomaning, and D.C. Christiani. 2009. “Genome-Wide Analysis of Survival in Early-Stage Non-Small-Cell Lung Cancer.” *Journal of Clinical Oncology* 27:2660–2667, PMID: 19414679.
- Hudgens, M.G., C. Li, and J.P. Fine. 2014. “Parametric likelihood inference for interval censored competing risks data.” *Biometrics* 70:1–9.
- Hughey, J.J., S.D. Rhoades, D.Y. Fu, L. Bastarache, J.C. Denny, and Q. Chen. 2019. “Cox regression increases power to detect genotype-phenotype associations in genomic studies using the electronic health record.” *BMC genomics* 20:805.
- Jabbari, E., S. Koga, R.R. Valentino, R.H. Reynolds, R. Ferrari, M.M.X. Tan, J.B. Rowe, C.L. Dalgard, S.W. Scholz, D.W. Dickson, T.T. Warner, T. Revesz, G.U. Höglinger, O.A. Ross, M. Ryten, J. Hardy, M. Shoai, H.R. Morris, K.Y. Mok, D.P. Murphy, S. Al-Sarraj, C. Troakes, S.M. Gentleman, K.S. Allinson, Z. Jaunmuktane, J.L. Holton, A.J. Lees, C.M. Morris, Y. Compta, E. Gelpi, J.C. van Swieten, A. Rajput, L. Ferguson, M.R. Cookson, J.R. Gibbs, C. Blauwendraat, J. Ding, R. Chia, B.J. Traynor, A. Pantelyat, C. Violett, B.J. Traynor, O. Pletnikova, J.C. Troncoso, L.S. Rosenthal, A.L. Boxer, G. Respondek, T. Arzberger, S. Roeber, A. Giese, D.J. Burn, N. Pavese, A. Gerhard, C. Kobylecki, P.N. Leigh, A. Church, and M. T.M. Hu. 2021. “Genetic determinants of survival in progressive supranuclear palsy: a genome-wide association study.” *The Lancet Neurology* 20:107–116.
- Kapoor, M., J.C. Wang, L. Wetherill, N. Le, S. Bertelsen, A.L. Hinrichs, J. Budde, A. Agrawal, L. Almasy, K. Bucholz, D.M. Dick, O. Harari, X. Xiaoling, V. Hesselbrock, J. Kramer, J.I. Nurnberger, J. Rice, M. Schuckit, J. Tischfield, B. Porjesz, H.J. Edenberg, L. Bierut, T. Foroud, and A. Goate. 2014. “Genome-wide survival analysis of age at onset of alcohol dependence in extended high-risk COGA families.” *Drug and Alcohol Dependence* 142:56–62.
- Kelly, P.J., and L.L.Y. Lim. 2000. “Survival analysis for recurrent event data: an application to childhood infectious diseases.” *Statistics in Medicine* 19:13–33.
- Kim, H.Y., J.M. Williamson, and H.M. Lin. 2016. “Power and sample size calculations for interval-censored survival analysis.” *Statistics in Medicine* 35:1390–1400.
- Lai, T.L., and Z. Ying. 1991. “Rank Regression Methods for Left-Truncated and Right-Censored Data.” *The Annals of Statistics* 19:531 – 556.

- Lee, S., G.R. Abecasis, M. Boehnke, and X. Lin. 2014. “Rare-variant association analysis: study designs and statistical tests.” *Am J Hum Genet* 95:5–23.
- Lee, S., M.J. Emond, M.J. Bamshad, K.C. Barnes, M.J. Rieder, D.A. Nickerson, D.C. Christiani, M.M. Wurfel, and X. Lin. 2012. “Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies.” *The American Journal of Human Genetics* 91:224–237.
- Li, C., D. Pak, and D. Todem. 2020. “Adaptive lasso for the Cox regression with interval censored and possibly left truncated data.” *Statistical Methods in Medical Research* 29:1243–1255, PMID: 31203741.
- Lin, D.Y. 1994. “Cox regression analysis of multivariate failure time data: The marginal approach.” *Statistics in Medicine* 13:2233–2247.
- Lin, D.Y., and L.J. Wei. 1989. “The Robust Inference for the Cox Proportional Hazards Model.” *Journal of the American Statistical Association* 84:1074–1078.
- Lin, D.Y., and Z. Ying. 1994. “Semiparametric analysis of the additive risk model.” *Biometrika* 81:61–71.
- Liu, D., X. Lin, and D. Ghosh. 2007. “Semiparametric Regression of Multidimensional Genetic Pathway Data: Least-Squares Kernel Machines and Linear Mixed Models.” *Biometrics* 63:1079–1088.
- McClellan, J., and M.C. King. 2010. “Genetic Heterogeneity in Human Disease.” *Cell* 141:210–217.
- Nagy, Á., G. Munkácsy, and B. Györffy. 2021. “Pancancer survival analysis of cancer hallmark genes.” *Scientific Reports* 11:6047.
- Pak, D., C. Li, and D. Todem. 2019. “Semiparametric analysis of correlated and interval-censored event-history data.” *Statistical Methods in Medical Research* 28:2754–2767.
- Shaffer, J., X. Wang, E. Feingold, M.K. Lee, F. Begum, D. Weeks, K. Cuenco, M. Barmada, S. Wendell, D. Crosslin, C. Laurie, K. Doheny, E. Pugh, Q. Zhang, B. Feenstra, F. Geller, H. Boyd, H. Zhang, M. Melbye, and M. Marazita. 2011. “Genome-wide Association Scan for Childhood Caries Implicates Novel Genes.” *Journal of dental research* 90:1457–62.
- Shu, X.O., J. Long, W. Lu, C. Li, W.Y. Chen, R. Delahanty, J. Cheng, H. Cai, Y. Zheng, J. Shi, K. Gu, W.J. Wang, P. Kraft, Y.T. Gao, Q. Cai, and W. Zheng. 2012. “Novel Genetic Markers of Breast Cancer Survival Identified by a Genome-Wide Association Study.” *Cancer Research* 72:1182–1189.
- Sinnott, J.A., and T. Cai. 2013. “Omnibus Risk Assessment via Accelerated Failure Time Kernel Machine Modeling.” *Biometrics* 69:861–873.

- Staley, J.R., E. Jones, S. Kaptoge, A.S. Butterworth, M.J. Sweeting, A.M. Wood, and J.M.M. Howson. 2017. “A comparison of Cox and logistic regression for use in genome-wide association studies of cohort and case-cohort design.” *European journal of human genetics : EJHG* 25:854–862.
- Subramanian, A., P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, and J.P. Mesirov. 2005. “Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles.” *Proceedings of the National Academy of Sciences* 102:15545–15550.
- Sun, J., and L.J. Wei. 2000. “Regression Analysis of Panel Count Data with Covariate-Dependent Observation and Censoring Times.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 62:293–302.
- Sun, T., and Y. Ding. 2019. “Copula-based semiparametric regression method for bivariate data under general interval censoring.” *Biostatistics* 22:315–330.
- van der Net, J.B., A.C.J.W. Janssens, M.J.C. Eijkemans, J.J.P. Kastelein, E.J.G. Sijbrands, and E.W. Steyerberg. 2008. “Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies.” *European journal of human genetics : EJHG* 16:1111–1116.
- van der Vaart, A.W., and J. Wellner. 1996. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics, Springer.
- Wang, X., S. Ma, and J. Yan. 2013. “AUGMENTED ESTIMATING EQUATIONS FOR SEMIPARAMETRIC PANEL COUNT REGRESSION WITH INFORMATIVE OBSERVATION TIMES AND CENSORING TIME.” *Statistica Sinica* 23:359–381.
- Wei, C., and Q. Lu. 2017. “A generalized association test based on U statistics.” *Bioinformatics* 33:1963–1971.
- Yu, S.L., H.Y. Chen, G.C. Chang, C.Y. Chen, H.W. Chen, S. Singh, C.L. Cheng, C.J. Yu, Y.C. Lee, H.S. Chen, T.J. Su, C.C. Chiang, H.N. Li, Q.S. Hong, H.Y. Su, C.C. Chen, W.J. Chen, C.C. Liu, W.K. Chan, W.J. Chen, K.C. Li, J.J. Chen, and P.C. Yang. 2008. “MicroRNA Signature Predicts Survival and Relapse in Lung Cancer.” *Cancer Cell* 13:48–57.
- Zeng, D., F. Gao, and D. Lin. 2017. “Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data.” *Biometrika* 104:505–525.
- Zeng, D., and D. Lin. 2020. “Maximum likelihood estimation for semiparametric regression models with panel count data.” *Biometrika* 108:947–963.
- Zeng, D., and D.Y. Lin. 2006. “Efficient estimation of semiparametric transformation models for counting processes.” *Biometrika* 93:627–640.

Zeng, D., L. Mao, and D.Y. Lin. 2016. "Maximum likelihood estimation for semiparametric transformation models with interval-censored data." *Biometrika* 103:253–271.