Group Sequential Robust Designs in Genetic Studies

By

Lihan Yan

B.S. in Computer Science 1989, Shanghai Normal University, Shanghai, China

M.S. in Statistics 1993, Virginia Polytechnology Institute and State University,

Blacksburg, VA

A Dissertation Submitted to

The Faculty of
Columbian College of Arts and Sciences
of The George Washington University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

May 17, 2009

Dissertation Directed by

Zhaohai Li

Professor of Statistics and Biostatistics

Gang Zheng
Mathematical Statistician
National Heart, Lung and Blood Institute
National Institutes of Health

The Columbian College of Arts and Sciences of The George Washington University certifies that Lihan Yan has passed the Final Examination for the degree of Doctor of Philosophy as of March 9, 2009. This is the final and approved form of the dissertation.

Group Sequential Robust Designs in Genetic Studies

Lihan Yan

Dissertation Research Committee:

Zhaohai Li, Professor of Statistics and Biostatistics,
Dissertation Co-Director

Gang Zheng, Mathematical Statistician,
National Heart, Lung and Blood Institute, National Institutes of Health,
Dissertation Co-Director

Yinglei Lai, Assistant Professor of Statistics,
Committee Member

Heather Hoffman, Assistant Professor of Epidemiology and Biostatistics,
Committee Member

Dedication

To my two beautiful daughters Heather and Allison

Acknowledgments

First and foremost, I would like to thank my advisor Professor Zhaohai Li for his direction of this work. His encouragement had a significant impact on my decision of pursuing a Ph.D. degree in biostatistics seven years ago. He opened up the door to the genetic statistics area where I did not think I would go before. His enthusiasm and devotion inspired me and led me to a path filled with challenges and excitement, where the application of statistical methods can contribute a great deal in time to the modern health research involving genetic data.

I would like to thank my co-advisor Dr. Gang Zheng for his vision and enthusiasm for genetic statistics which have had an important impact on the success of my research work. I appreciate the enormous amount of time he has spent with me mentoring and directing my research topics. I would also like to thank Dr. Aiyi Liu for his constructive comments and introduction of research ideas during the development of my work. I learned a great deal from both of them along with Prof. Li, not only about statistics, but also how to conduct successful researches.

Each member of my committee contributed to this dissertation. Prof. Tapan Nayak, Prof. Yinglei Lai, Prof. Heather Hoffman, Prof. Subrata Kundu, and Dr. Kai Yu encouraged me to reach higher and go the extra distance. I deeply appreciate their interest and personal investment of time. Thanks also goes to faculty of the departments of statistics, biostatistics and epidemiology, especially Professor John Lachin and Professor Dante Verme for their advice during my pursuit of the Ph.D. degree.

I would like to express my gratitude to my employers, including the EMMES Corporation, and the Food and Drug Administration, for providing partial financial support for my doctoral program studies. I would especially like to thank my former boss, Dr. Mark Wolff, for having brought out the confidence and persistence I needed to achieve this goal. In addition, this dissertation and my manuscripts benefited enormously from the editorial comments made by my sister-in-laws, Kathleen Parker and Ann Pluchino, as well as my former colleagues, Jill Barrett and Jacqueline Fiore.

I am very grateful to my father Bicheng Yan and my mother Jingyao Ni who passed away in 2002. Their never-ending love, encouragement and support made me who I am and carried me to where I am. I also thank my brother Litao Yan and my sister-in-law Xiaojun Ren as well as my many friends for their emotional support during hardships of my life.

Finally, I am particularly indebted to my husband Tom Keesling and my two beautiful daughters, Heather and Allison, for their unconditional love and strong support, without which this dissertation would not have been possible.

Abstract of Dissertation

Group Sequential Robust Designs in Genetic Studies

With the development of modern technology in genetic testing, studies of association and linkage between genetic markers and diseases have emerged and become an important area in human health research. There are three major genetic study designs: population-based case-control study, family-based association study and linkage study. Each of these designs has its own strength and is most suitable for a particular setting. The statistical analysis method of the data collected in any of these studies is often a choice among typically two or three statistical tests. Each test is optimal under the specified inheritance mode or genetic model, such as the additive, dominant, recessive or sometimes multiplicative model. Unfortunately, the information about the underlying genetic model is often unknown for many complex diseases. Applying a statistical test that mis-specifies the genetic model can result in significant loss of power. To accommodate this problem, robust statistics have been proposed by researchers. One commonly referred test statistic, proposed by Gastwirth [1965], is called the "maximin efficient robust test" or MERT. This statistic "averages" the two extreme test statistics of a series of tests by taking into account their correlation with certain constraints. The other statistic is the maximum of all the test statistics, referred to as MAX [Freidlin *et al.* 2003, Zheng *et al.* 2002]. For genetics studies, The MERT and MAX show similar robustness when the genetic model is additive. When the genetic model is dominant or recessive and minor allele frequency is small, MAX has shown advantages in achieving higher power than MERT. MAX then becomes preferable as the robust test in genetic studies. However,

while the computation of MAX is simple and straightforward, it no longer follows asymptotic normal distribution like MERT.

From a different point of view in terms of study designs, group sequential designs have been widely used in clinical trials. Such designs allow early stopping of the trial according to pre-specified stopping rules for either efficacy, futility, or both. Application of group sequential designs in genetic studies can be beneficial in terms of tremendous saving of cost and time incurred during genetic testing and sample ascertainment. It is therefore helpful to develop a tool that can statistically guard an approach that combines the group sequential design and the robust approach of statistical testing in genetic studies. This dissertation presents the applications of such an approach in different genetic study designs and solutions to several challenges encountered from different aspects during the application. It describes methods of obtaining critical values for setting up boundaries in a two-stage group sequential design using MAX as the test statistic. Optimal designs varying in different designs and genetics parameters are studied. This dissertation also presents adaptive designs using two proposed genetic model selection methods as the alternative robust approaches that preserve power under an unknown genetic model.

Chapter 1 provides an introduction to the topic and a summary of the results from the dissertation.

Chapter 2 provides a detailed review of literature with respect to the applications of the robust test statistics in each of the genetic study designs and traditional group sequential design, as well as its application in genetic studies.

Chapter 3 presents applications of two-stage group sequential designs using robust test (MAX) in family-based genetic studies. These applications focus on controlling type I error when early stopping is allowed for significant findings during interim analyses. The calculation methods of the upper bound of the type I error due to simultaneous hypothesis testings proposed by Efron [1997] are extended for usage of determining critical values at each stage of the group sequential tests using MAX. Although these methods are based on one-sided tests, critical values for two-sided tests are obtained using the method of symmetry. Monte Carlo simulations are conducted for both one-sided and two-sided tests for various settings of the family-based association studies and confirm that these critical values adequately control overall type I error.

In Chapter 4, a two-stage sequential design is extended to linkage studies of affected sib pairs. In this type of genetic study, nonparametric statistical methods (the means tests and proportions tests) using the information of allele sharing identical-by-descent (IBD) are considered. Although the knowledge of genetic model is not required in these tests, it has been shown that each test shows advantages in power under the study of rare diseases. The MERT statistic is constructed by combining the means test and the proportions test whereas the MAX statistic is defined as the maximum of the means, proportions and MERT test statistics. Since the tests in linkage studies are directional, the methods developed in Chapter 3 can be readily applied to obtain critical value for each stage under the specified alpha spending function. Both type I errors and powers are examined and compared among different test statistics in this study setting.

Chapter 5 focuses on the population-based case-control study design, which might be the most commonly used design due to its simplicity of sample ascertainment and straightforward application of the traditional statistical methods. Average sample sizes (ASN) and powers, as the measures of study efficiency, are examined in a two-stage sequential study. Optimization between minimizing the average sample size and maximizing the power are explored by changing parameters, such as allocation fraction for samples in the first stage, and the alpha spending function, under different assumptions of allele frequencies and genetic models.

Chapter 6 presents an innovative two-stage group sequential design that uses adaptive statistics based on the examination of data obtained at the first stage. The motivation of this design also comes from the problem of the unknown genetic model. Two approaches for selecting the genetic model hence the test statistic are examined. Significance tests are based on Fisher's combination test proposed by Bauer and Kohne [1994]. Powers are compared among several non-adaptive approaches and the two adaptive approaches. The results show slight power advantage and efficiency for adaptive designs using the genetic model selection method based on the Hardy-Weinberg disequilibrium test.

Finally, Chapter 7 discusses the strength and limitations of proposed designs. Future research needs are also presented.

Contents

Dedication iii

Acknowledgements iv

Abstract of Dissertation vi

Table of Contents x

List of Figures xii

List of Tables xiii

1 Introduction and Summary of Results 1

1.1 Controlling type I error in two-stage group sequential robust tests in family based association 2

1.2 Application of group sequential designs for linkage studies 3

1.3 Optimization for two-stage sequential robust designs in population-based case-control association studies 3

1.4 Two-stage adaptive designs in population-based case-control association studies 4

2 Literature Review 6

2.1 Genetic models 6

2.2 Genetic study designs and statistical tests 7
2.3 Group Sequential Design 11

3 Two-Stage Group Sequential Robust Test in Family Based Association Studies: Controlling Type I Error 16

3.1 Conditional score tests 16
3.2 Asymptotic null correlations 18
3.3 Upper bound for the probability of the maximum of statistics 19
3.4 Controlling type I error in the robust two-stage design 23
3.5 Results 27
3.6 Summary 30

4 Two-Stage Designs for Linkage Studies 41

4.1 Introduction 41
4.2 Tests in linkage studies of affected relative pairs and their correlations 42
4.3 Simulation studies 46
4.4 Summary 48

5 Optimal Two-Stage Sequential Design for Case-Control Genetic Association Studies 51

5.1 Introduction 51
5.2 Background 52
5.3 Methods 53
5.4 Simulation studies and examples 57
5.5 Discussion 60

6 Two-Stage Adaptive Design for Case-Control Genetic Association

Studies **74**

6.1 Introduction 74

6.2 Methods 76

6.3 Simulation studies 79

6.4 Summary 82

7 Discussion and Future Research in Related Areas **86**

7.1 Discussion 86

7.2 Limitations and research needs 88

References **92**

Appendix: R Programs **100**

List of Figures

3.1 Event $\{T_{max} > c\} = E_1 \cup E_2$ in terms of the standard normal vector Z (Efron[1997]) 22

3.2 Power Comparisons for different sets of critical values in a two-stage family-based study 40

5.1 ASNs and sample sizes for one-stage and two-stage studies at 80% Power: γ_2=1.5: ASF = ASF_1. 69

5.2 ASNs and sample sizes for single-stage and two-stage studies at 80% Power: γ_2=1.5, ASF = ASF_2. 70

5.3 ASNs and sample sizes for single-stage and two-stage studies at 80% Power: γ_2=1.5, ASF = ASF_3. 71

List of Tables

3.1 Correlation matrix in a two-stage sequential study 32

3.2 Asymptotic null correlation matrices given allele frequencies (p) . . . 32

3.3 Critical values for one-sided tests at the overall type I error of $\alpha = 0.025$ in two-stage group sequential family-based association studies using different alpha spending functions (ASF) 33

3.4 Simulated type I errors for two-stage family-based association studies: allele frequency $p = 0.1$. 34

3.5 Simulated Type I errors for two-stage family-based association studies: allele frequency $p = 0.3$. 35

3.6 Simulated Type I errors for two-stage family-based association studies: allele frequency $p = 0.5$. 36

3.7 Simulated Type I errors with ranges for two-stage group sequential family-based association studies: allele frequency $p = 0.3$. 37

3.8 Simulated Type I errors with ranges for two-stage group sequential family-based association studies: allele frequency $p = 0.1$ in the trio setting ($r = 1$, $s = 0$). 38

3.9 Simulated Type I errors with ranges for two-stage group sequential family-based association studies: allele frequency $p = 0.5$ in the trio setting ($r = 1$, $s = 0$). 38

3.10 Simulated Type I errors for two-Stage family-based association studies at desired overall alpha level of .0001 using O'Brien-Fleming spending function (ASF=1): allele frequency $p = 0.3$ in the trio setting ($r = 1$, $s = 0$). 39

4.1 Probabilities that a pair of relative share 0, 1, 2 alleles IBD at an autosomal (non-sex chromosome) locus. 43

4.2 Critical values for one-side MAX test in linkage studies and simulation results 49

4.3 Critical values used in the power study reflexing O'Brien-Fleming bound 50

4.4 Powers of different statistics while type I error is maintained under the null 50

5.1 Population-based case-control study design. 53

5.2 Optimal allocation fractions π to achieve the minimum ASN or the maximum power given the target sample size for a single-stage design ($N = 1000$ per group): ASF = ASF_1. The alternatives are specified by the GRRs (γ_1, γ_2) under three genetic models with different MAFs p. 63

5.3 Optimal allocation fractions π to achieve the minimum ASN or the maximum power given the target sample size for a single-stage design ($N = 1000$ per group): ASF = ASF_2. The alternatives are specified by the GRRs (γ_1, γ_2) under three genetic models with different MAFs p. 64

5.4 Optimal allocation fractions π to achieve the minimum ASN or the maximum power given the target sample size for a single-stage design ($N = 1000$ per group): ASF = ASF_3. The alternatives are specified by the GRRs (γ_1, γ_2) under three genetic models with different MAFs p. 65

5.5 Sample size and optimal allocation fraction to achieve the power 80%: ASF = ASF_1. 66
5.6 Sample size and optimal allocation fraction to achieve the power 80%: ASF = ASF_2. 67
5.7 Sample size and optimal allocation fraction to achieve the power 80%: ASF = ASF_3. 68
5.8 Description of genotype distributions in 17 SNPs from 4 GWAS studies. 72
5.9 ASNs and powers for a two-stage design for the 17 SNPs 73

6.1 Distribution (in percentage) of model selection by the maximum and HWDTT approaches 83
6.2 Power comparisons among non-adaptive and adaptive designs under HWE (F=0) 84
6.3 Power comparisons among non-adaptive and adaptive designs without HWE (F=0.05) 85

Chapter 1

Introduction and Summary of Results

During the assessment of the association and linkage between a candidate gene and a disease in genetic studies, the statistical tests are genetic model specific. In another words, the most powerful or optimal test statistic is constructed according to the inheritance mode or genetic model such as dominant, recessive, additive or multiplicative model. However, the genetic model is often unknown. When the genetic model is mis-specified in the test statistic, there could be a substantial power loss. The test statistic under the additive model had been used since it is the most robust among the three and maintains decent power in many situations-but not all. Researchers have proposed two other simple robust test statistics: the maximin efficient robust statistic (MERT) and the maximum statistic (MAX). While the MERT statistic has its advantage of maintaining asymptotic normal distribution after averaging the two extreme statistics by taking into account their correlations, the MAX statistic has shown better power in more seen situations such as under the dominant or recessive model. However, its distribution is in complex form that involves multiple integration and no longer follows asymptotically normal distribution.

Group sequential designs have been widely used in clinical trials. Such designs

allow early stopping during interim look(s) of the study data and provide flexibility so that the efficiency of the study may be maximized. Applications of group sequential designs in genetic studies have also been studied by researchers. Statistical methods have been developed for these designs to deal with the control of the overall type I error because of multiplicity of tests. However, these methods are typically developed for a single test statistic with a known form of asymptotic distribution. When there is a need to use the MAX test statistic in these designs, the methods are not readily available and have not been studied. In this research, the applications of group sequential designs are studied in different types of genetic studies when using the robust statistic MAX. We propose methods of obtaining the critical values when MAX is used and recommend optimal designs after considering the impact of various designs and study parameters. We also present adaptive designs that use adaptive statistics based on the selected genetic model from the examination of the first stage data and compare results with the nonadaptive robust design.

1.1 Controlling type I error in two-stage group sequential robust tests in family based association

In family-based association studies, the test statistics are based on the score tests conditional on mating types [Li *et al.* 2005]. We considered using the maximum of three test statistics under the dominant, additive and recessive genetic models. For a two-stage sequential design with a pre-specified alpha spending function, the critical values were obtained through several approximation methods proposed by Efron [1997] for calculating the probabilities of simultaneous testing from correlated test statistics. Simulation studies were carried out for different allele frequencies and

for different family structures, i.e. different number of diseased and non-diseased offsprings in a family. The results for a two-stage sequential design showed good controls of type I error and trend of conservativeness among different methods and under different specified values of parameters.

1.2 Application of group sequential designs for linkage studies

As special type of linkage studies, studies of affected sib pairs (ASPs) have a broad application in the genetic research. In the literature, the means and proportions tests are the most commonly used non-parametric tests used in linkage studies of relatives pairs. While they are "model free", studies have shown one has advantage over the other in different scenarios. A "Minimax" statistic proposed by Whittmore and Tu [1995], is algebraically equivalent to the MERT statistic for ASP studies. In the present study, we extend the methods of obtaining critical values proposed earlier to linkage studies. We study not only control of type I errors but also the power of the designs for the three test statistic described above, along with the MAX statistic of them. Simulation results confirmed preservation of overall type I error using the approaches described in Chapter 3 and showed power improvement by using MAX.

1.3 Optimization for two-stage sequential robust designs in population-based case-control association studies

The present study investigates the cost of sample size versus the statistical power in a case-control genetic study with a two-stage sequential design using the robust test statistic MAX. We study various factors that can affect the goal of minimizing

the average sample size (ASN) or maximizing power of a genetic case-control study under the two-stage sequential design and make recommendations for the optimal solutions to different scenarios. The results show advantages of having a two-stage design in terms of savings on average sample size while maintaining the power in a slightly reduced rate. The results also indicated that the typical allocation fraction of half (0.5) often balances the needs between sample sizes and powers. Under strong alternatives, a smaller allocation fraction, e.g. 0.3, can result in a reduction of more than half of the sample size comparing to a single stage study. The choice of alpha spending function can impact the average sample size. Overly conservative alpha spending helps little in average sample size in a sequential design. Finally, we also apply these solutions in four existing genome-wide association studies of a total of 17 SNPs and confirm the savings.

1.4 Two-stage adaptive designs in population-based case-control association studies

There has been a great deal of recent interest in adaptive designs due to various uncertainties during design. It is also the case for genetic studies, for example, the uncertainty of the genetic model hence the employment of the test statistic. In present study, we propose to use two genetic model selection methods to select the optimal statistic for the selected model in a two-stage adaptive design. The first approach (the MAX method) uses the genetic model that results in the highest test statistic. The second approach, proposed by Zheng and Ng [2008], is based on the direction of magnitude of the Hardy-Weinberg disequilibrium test statistic (the HWDTT method). Simulation studies are performed to provide the operating characteristics and compare them among various choices of adaptive and non-adaptive designs. The results

show similar robustness between the non-adaptive robust design and the adaptive design using the MAX method. On the other hand, the adaptive design using the HWDTT method has slight better power than the non-adaptive robust design.

Chapter 2

Literature Review

2.1 Genetic models

Genetic model, or inheritance mode is a unique description for genetics data. It describes the manner in which a particular genetic trait or disorder is passed from one generation to the next. The types of models include, but not limited to, three major genetic models: the dominant, recessive and additive models. Denote the three genotypes AA, Aa, and aa and let A be the minor allele. The dominant model indicates that the risks of having the disease are the same as long as there is at least one minor allele A (either Aa or AA). Similarly, the recessive model indicates that the risks of having, or rather, not having the disease is the same for genotypes with at least one major allele (a) (Aa or aa). Under the additive model, the risk of having disease for the genotype with two minor alleles (AA) is twice as likely for the genotype with no minor alleles (aa) in a certain direction as is for the genotype with just one minor allele (Aa) when compared to the genotype with no minor alleles (aa).

These definitions can be typically expressed in terms of penetrance factors. Denote the penetrances for the genotypes aa, Aa, and AA by f_0, f_1, and f_2, respectively, where f_i=P(case $|i$ A alleles), $i = 0, 1, 2$. As A is the candidate high risk allele, we

assume $f_2 \geq f_1 \geq f_0 > 0$. Let $\gamma_i = f_i/f_0$ be the genotype relative risk, $i = 1, 2$. Then the three genetic models can be described as: $\gamma_1 = \gamma_2$ for the dominant model, $\gamma_1 = 1$ for the recessive model, and $2\gamma_1 = 1 + \gamma_2$ for the additive model. The multiplicative model $\gamma_1^2 = \gamma_2$ is almost equivalent to the additive model. This can be proved by $\gamma_1^2 \approx 1 + 2(\gamma_1 - 1) = 2\gamma_1 - 1$, using a first order Taylor approximation. Then $\gamma_1^2 = \gamma_2 \approx 2\gamma_1 - 1$. The equation of first two values indicates multiplicative model, whereas the approximation of the last two values also indicates additive model.

2.2 Genetic study designs and statistical tests

Genetic study designs are divided into two categories: association designs and linkage designs. Association designs can be further divided into population-based case-control studies and family-based designs. Population-based studies consist of independent cases (subjects with the disease) and controls (subjects without the disease) whereas family-based studies consist of sample population in units of families, in which the genotypes and disease status are available for both parents and their offsprings. On the other hand, linkage studies typically consist of genetically related subjects such as affected sib pairs. Each of these study designs has its own advantages and disadvantages. The population-based case-control study design is susceptible to population stratification. When allele frequencies and disease prevalence rates vary across subpopulations, using a case-control design may lead spurious estimates of the association [Cardon and Palmer 2003, Hattersley and McCarthy 2005, Wang *et al.* 2005, Evangelou *et al.* 2006]. However, it provides the higher power and may be the most efficient for sample ascertainment among the study designs. The family-based study design can eliminate the effect from population stratification, but then it requires more effort during sample collections and may be less powerful and are

susceptible to other biases. Finally, the linkage study design usually involves looking at large families where the disease affects individuals in several generations. The advantage of this design is that it is statistically robust and unlikely to give false positive results. The disadvantage is that it is less suitable for identification of gene responsible for complex traits and has low statistical power to detect gene with modest effect. The statistical tests used in these designs are described below.

2.2.1 Population-based case-control association study design

In a population-based case-control study of association between a diallelic genetic marker (with alleles a and A) and a underlying disease, there are three genotypes denoted by aa, Aa and AA. Assume allele A is the risk allele. There are two commonly used test statistics for population-based case-control studies. The first classifies cases and controls according to their alleles and corresponds to an allelic contingency table, where a Pearson's χ^2 test is used. This test is referred to as the allele-based test (ABT). Information of underlying genetic model (dominant, additive or recessive) is not needed in such test. The advantage of this test model is that the number of observations has been doubled. The disadvantage is that the genotypic-specific information, such as which alleles are paired together, is ignored. A further disadvantage of basic allele testing is that stratification correction through the principal components analysis method is not available for this model. Moreover, this test assumes Hardy-Weinberg equilibrium (HWE), i.e. both allele and genotype frequencies in a population remain constant. When HWE is not satisfied due to various reasons such as non-random mating, mutations, selection, limited population size, random genetic drift and gene flow, this test is biased and invalid [Sasini 1997].

The second test classifies cases and controls according to their genotypes which leads to a genotypic contingency table. The most commonly used test method is the

Cochran-Armitage trend test (CATT) [Armitage 1955, Cochran 1954]. This test has been shown to have more statistical power than the typical χ^2 test for a 2x3 table test. This test can be used even when HWE is not satisfied. It utilizes a set of scores that can be the efficient score for a logistic regression. Hence this test is a locally optimal test statistic for a set of given score, which is driven by the underlying genetic model [Freidlin *et al.* 2002, Schaid and Sommer 2001, Tarone and Gart 1980]. The disadvantage is that mis-specification of genetic model can result in substantial power loss. To avoid such situations, Zheng *et al.* [2003] proposed a robust test statistic that takes the maximum of the statistics under all three genetic models (dominant, additive and recessive), denoted as MAX. Note that under the null hypothesis, MAX no longer follows the asymptotic normal distribution as the individual test statistic. Its asymptotic distribution can be obtained by simulation [Zheng *et al.* 2005].

2.2.2 Family-based association study design

Although the information collected in the family-based association study is similar to that in the population-based study, the cases and controls may be correlated. Typical data collection in family-based studies requires genetic marker information from parents to their offsprings. A prime example of a family-based association test is the transmission disequilibrium test (TDT) [Spielman *et al.*, 1993]. The TDT test relies on data on transmission of marker alleles from heterozygous parents to affected offspring. In the special case of TDT using families of trios (heterozygous parents and an affected child), the test is an application of McNemar's test. Utilizing the fact that they are independent if conditioning on their parents' mating type, Li *et al.* [2005] proposed a more generalized genetic model-specific score test statistics conditional on parental mating types. Under additive or multiplicative models, the conditional score test is equivalent to the TDT. It is not the case for dominant or recessive models.

Zheng *et al.* [2003] compared the powers of the tests under their model assumptions and also found MAX robust against model misspecification.

2.2.3 Linkage study design

The principle of a linkage study is that the genetic markers run the same way as the disease in the family, assuming the gene that causes the disease is somewhere in the same area of the genome as the marker. Therefore, the linkage study design is the one that studies disease susceptibility genes using using a group of genetically related subjects, ideally generations of families. The study of affected sib-pairs (ASPs) is the most commonly used design. ASP analysis based upon allele sharing identical-by-descent (IBD) is a popular form of nonparametric linkage analysis. This method has frequently been used to study complex traits. Deviations of the observed IBD distribution in a sample of ASPs from the trinomial distribution expected under H_0 are indicative for linkage between the marker and the disease locus. Various statistics have been proposed to measure this deviation. The first test is the likelihood ratio test (also called the "maximum-lod-score test"), described by Risch [1990b]. This test requires the assumption of genetic model. The second test is called the means test. It is based on the total number of alleles IBD in the sample. The third test is called the proportions test and it is based on the number of affected sib-pairs sharing both alleles IBD. An explicit specification of the inheritance model is not required for performing nonparametric linkage analysis of ASPs. However, the absolute power of any ASP test as well as the relative performance of different ASP tests is influenced by the inheritance model and the genetic distance (i.e., the recombination fraction) between the marker and disease locus. There is no ASP test which is superior to all other ASP tests, i.e., there is no uniformly most powerful ASP test. Blackwelder and Elston [1985] compared the power of the means test,

the proportions test, and a χ^2 goodness-of-fit test of the observed and expected IBD distribution. They demonstrated superiority of the means test over the proportions test. On the other hand, Cox and Hinkley [1974] pointed that the means test and the proportions test are locally most powerful for the families of additive genetic models (i.e., those with no dominance variance component), and those of models with no or relatively small additive variance component in comparison with the dominant component, respectively. The latter situation describes a rare recessive allele. Schaid and Nick [1990] also compared the tests and a new statistic (the maximum of the means and proportions tests. They showed that this new statistic is robust against power loss among a range of true IBD probabilities. Whittemore and Tu [1998] introduced a robust test, called the "minmax" test and showed its robustness against misspecification of the true probabilities of IBD among the sibs. This test, when used in ASP studies is equivalent to the maximin efficient robust test (MERT) (Gastwirth [1966]) combining the means and proportions tests.

2.3 Group Sequential Design

Group sequential designs have been widely used in clinical trials. In these designs, interim analyses of accumulating data are conducted due to various reasons: ethical, administrative or economic. In trials involving human subjects, there is an ethical need to monitor results to ensure that individuals are not exposed to unsafe, inferior or ineffective treatment regimens. Even in negative trials where there appears to be no difference in the performance of two therapies, there is an ethical imperative to terminate a trial as soon as possible so that resources can be allocated to study the next most promising treatment waiting to be tested. Interim analyses are also needed for administrative reasons, for example, to ensure the trial is being executed as

planned, that the subjects or experimental units are from the correct population and satisfy eligibility criteria. Sequential statistical methods were originally developed in order to obtain economic benefits. For a trial with a positive result, early stopping means that a new product can be exploited sooner. If a negative result is indicated, early stopping ensures that resources are not wasted. Sequential methods typically lead to saving in sample size, time and cost when compared with standard fixed sample procedures. The key elements considered in a group sequential design include the planned statistical test, the number and spacing of interim looks and stopping rules/boundaries that are dependent on how the type I error is spent at each interim look and desired statistical power. Preservation of the type I error is one of the key areas of statistical researches. Below is a brief description of the literature on this aspect.

2.3.1 Alpha spending function and critical values

One element in a group sequential design is the alpha spending function, which allocates the type I error rate throughout a study. Because there are multiple looks, the alpha value at each look must be adjusted in order to preserve the overall type I error. Alpha spending functions establish these adjusted alpha values for each interim monitoring point, given the overall alpha. They hence constitute "stopping boundaries", which, when crossed, indicate that statistical significance has been established. Group sequential methods were introduced in the late 1970s. Pocock [1977] proposed even spending on alpha with pre-specified number of looks. O'Brien and Fleming [1979] introduced a more conservative spending for early stages of the study. That is, an early stopping occurs only when remarkable significance is observed. Both the Pocock and the O'Brien-Flemming methods require predetermination of the number of interim analyses and equal increments of information between looks. A more

flexible approach was introduced by Lan and DeMets [1983], which eliminated the constraints of the group sequential methods by introducing a spending function for alpha. Researchers also studied the characteristics of alpha spending functions and proposed methods of establishing optimal boundaries [Wang and Tsiatis 1987]. Essentially, the choice relies on the trade off between sample size and power.

2.3.2 Group sequential designs in genetic studies

The benefits of time and cost saving for a group sequential design are also important for the emerging genetic studies. A few researchers have recently studied and demonstrated the benefits of such a design. Konig *et al.* [2001, 2003] utilized the design to calculate sample size and stopping boundaries in linkage studies based on the means test and association studies on the TDT. They showed that the application of a group sequential design led to a maximal increase in sample size of 8% under the null hypothesis, compared with the fixed-sample design. In contrast, this provides savings of up to 20% in average sample sizes under the alternative hypothesis. A similar conclusion was also drawn in Konig and Ziegler [2003], which demonstrated the application of group sequential analysis for case-control studies. These savings affect the amounts of genotyping and phenotyping required for a study and therefore lead to a significant decrease in cost and time. However, the analyses of these studies were performed by employing a single test statistic with a asymptotically normal distribution. When a robust test such as MAX is desired, since it does not follow asymptotic normal distribution, the classical design methods of setting up boundaries for stopping no longer apply.

2.3.3 Adaptive designs

In recent years, a number of researchers have proposed adaptive methods that allows changes the one or more pre-defined statistical parameters through the examination of the interim data. The adaptation is designed to improve the precision of estimation and statistical efficiency or to "rescue" the study in the case the parameter was misspecified or misestimated during the planning of the study. Examples of the study parameters that can be adapted include: sample size reassessment, sampling or randomization plan, adding or dropping a treatment group, testing statistic, dose selection. In flexible designs of adaptive test statistics, the information for the interim data can be used to select the test procedures for the analysis of future study stages. Several researchers proposed this type of design for their targeted hypothesis. For example, Lang *et al.* [2000] used this feature for the situation of testing the equality of several normal means against an ordered alternative. Kieser *et al.* [2002] proposed an approach that selects the test statistic at interim via a bootstrap method for estimation of the power of the two-sample Wilcoxon test for shift alternatives. They showed that this method provided considerable gain of power especially when the initial assumption of the underlying distribution was wrong. Statistical methods of combining independent p-values such as the Fisher's combination test are often considered [Bauer and Kohne 1994]. For genetic studies, several researchers have studied two-stage or two-phase designs that involved adaption of test statistics in non-sequential settings. For example, Song *et al.* [2007] proposed an adaptive two-stage procedure that screens single-nucleotide polymorphisms (SNPs) using the Hardy-Weinberg disequilibrium trend test (HWDTT) in a first stage, and then tests a reduced number of SNPs that pass the screening step in a second stage using the CATT. Zheng and Ng [2008] presented a two-phase test combining the HWDTT test

and consequently selected genetic model specific CATT test based on the results of the HWDTT test. However, formal adaptive designs in group sequential settings have not been studied for genetic studies.

Therefore, it is critical to develop a method that can statistically guard a group sequential design using a robust or an adaptive approach in order to both promote savings and preserve powers for situations where the underlying genetic model is known. This dissertation is motivated by these practical needs and fills the gap between that application of group sequential design and robust/adaptive statistics in genetic studies. We propose a few computationally straightforward and expandable approaches that set up the stopping boundaries for group sequential designs using MAX. We also examine impacts on power from different study parameters in these designs and recommend optimal designs parameters such as allocation fraction and alpha spending function. Finally, we propose adaptive group sequential designs that provide inferences of genetic model for future stage analyses based on interim look(s). All these methods are computationally straightforward and have shown efficiencies, making group sequential robust/adaptive designs practically feasible in the next era.

Chapter 3

Two-Stage Group Sequential Robust Test in Family Based Association Studies: Controlling Type I Error

In this chapter, we describe several approaches for obtaining critical values that maintain the overall type I error level in a two-stage robust sequential family-based association study using the maximum of the three genetic-model specific test statistics. Simulation studies are conducted to compare their performance under the null hypothesis of no association.

3.1 Conditional score tests

Statistical methods have been developed during the past decade to analyze data collected in family-based association studies. The TDT of Spielman *et al.* [1993] and the score statistics of Schaid & Sommer [1993] can only be applied to trios. For families with various numbers of affected and unaffected offspring, Li *et al.* [2005] proposed score statistics based on conditional likelihood functions for three different genetic models: additive (or equivalently multiplicative), dominant, and recessive

models. The derivation for the multiplicative model at first order Taylor's expansion indicates the additive model expression. The conditional score statistics are obtained by conditioning on the parental mating type. Like the usual score statistics of Schaid and Sommer [1993], when both parents' genotypes are homozygous, the conditional distribution of the genotype of their offspring is degenerated to a single possibility and does not contribute any information to the likelihood function. Therefore, the conditional score statistics only use families that consist of at least one heterozygous parent, specifically, mating types 1: $AA \times Aa$, 2: $Aa \times Aa$, and 3: $Aa \times aa$, where A and a are the two alleles of a candidate marker. The counts of affected and unaffected offsprings may vary. For simplicity, assume we have families with r affected children and s unaffected children. A trio is a special case with $r = 1$ and $s = 0$. Define the penetrances as $f_2 = P(\text{affected}|AA)$, $f_1 = P(\text{affected}|Aa)$ and $f_0 = P(\text{affected}|aa)$. Then the null hypothesis is $H_0 : f_2 = f_1 = f_0$ and the alternative hypothesis is $H_1 : f_2 \neq f_1 \neq f_0$. Let $N_{ij,d}$ be the number of children from mating type i ($i = 1, 2, 3$) having j number of A alleles $j = (0, 1, 2)$ and disease status being d ($d = a$ if affected, $d = u$ if unaffected). Let $N_{i,d}$ be the total number of children for mating type i having disease status d and N_i be the total number of children from mating type i. Let τ_{ij} be the conditional probability of a child having j number of A alleles under mating type i, which, under Hardy-Weinberg equilibrium (HWE), are given in the following table:

	Offspring Genotype		
Mating Type (i)	**AA** (τ_{i2})	**Aa** (τ_{i1})	**aa** (τ_{i0})
1. $AA \times Aa$	1/2	1/2	0
2. $Aa \times Aa$	1/4	1/2	1/4
3. $Aa \times aa$	0	1/2	1/2

Li *et al.* [2005] derived the following conditional score function and its estimated variance under the null hypothesis, which can be written as:

$$\begin{aligned} U &= \sum_{i=1}^{3} \left\{ \frac{1}{f_0} \left[N_{i2,a} + xN_{i1,a} - r(\tau_{i2} + x\tau_{i1}) \right] \right\} \\ &\quad + \sum_{i=1}^{3} \left\{ \frac{1}{1-f_0} \left[N_{i2,u} + xN_{i1,u} - s(\tau_{i2} + x\tau_{i1}) \right] \right\}, \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} V^* &= \sum_{i=1}^{3} \left\{ \frac{N_i \cdot r}{f_0^2} \left[\tau_{i2} + x^2\tau_{i1} - (\tau_{i2} + x\tau_{i1})^2 \right] \right\} \\ &\quad + \sum_{i=1}^{3} \left\{ \frac{N_i \cdot s}{(1-f_0)^2} \left[\tau_{i2} + x^2\tau_{i1} - (\tau_{i2} + x\tau_{i1})^2 \right] \right\}, \end{aligned} \tag{3.2}$$

where $x = 1/2, 1, 0$ for the additive, dominant, and recessive models, respectively. Note that $x = 1/2$ can be approximately used for the multiplicative model. In (3.1) and (3.2), f_0 can be estimated by population disease risk and the tests are robust to the estimation of f_0 [Li *et al.* 2005].

The corresponding conditional score test statistics can be written as $T = U/\sqrt{V^*}$, which is denoted as T_A, T_R and T_D for the additive, recessive, and dominant models, respectively. These test statistics have asymptotic $N(0,1)$ distribution under the null hypothesis. Note that these conditional test statistics are different from the score statistics of Schaid and Sommer [1993] unless $r = 1$ and $s = 0$. The test statistic that we consider here is $\text{MAX} = \max(T_A, T_R, T_D)$, for one-sided alternatives or $\text{MAX} = \max(|T_A|, |T_R|, |T_D|)$ for two-sided alternatives.

3.2 Asymptotic null correlations

Li *et al.* [2005] derived the asymptotic null correlation among the three conditional score tests in terms of probabilities of parental mating types. Assuming HWE, these correlations can be simplified as $\rho_{AD} = corr_{H_0}(T_A, T_D) = \{4q/(3+q)\}^{1/2}$, $\rho_{AR} =$

$corr_{H_0}(T_A, T_R) = \{4p/(3+p)\}^{1/2}$, and $\rho_{DR} = corr_{H_0}(T_D, T_R) = [pq/\{(3+p)(3+q)\}]^{1/2}$, where $p = P(A)$, the allele frequency, and $q = 1 - p$. These correlations can be estimated under the null hypothesis as:

$$\widehat{\rho_{AD}} = \frac{2(N_2 + N_3)}{\sqrt{(N_1 + 2N_2 + N_3)(3N_2 + 4N_3)}},$$

$$\widehat{\rho_{AR}} = \frac{2(N_1 + N_2)}{\sqrt{(N_1 + 2N_2 + N_3)(3N_2 + 4N_1)}},$$

$$\widehat{\rho_{DR}} = \frac{N_2}{\sqrt{(3N_2 + 4N_1)(3N_2 + 4N_3)}},$$

where N_i is the number of children under mating type i, i=1,2,3.

In a two-stage sequential association study, samples in stage 2 contain samples in stage 1. Denote sample sizes at stages 1 and 2 as n_1 and n_2, respectively. Let T_{Ai}, T_{Di}, and T_{Ri} be the statistics for the three genetic models at stage i, $i = 1, 2$. Then, for each genetic model, the correlation between the two statistics at stage 1 and stage 2 is a fraction ($\sqrt{n_1/n_2}$) of their correlation at stage 1. The correlation matrix for different genetic models and the two stages is given in Table 3.1. Within each stage, Table 3.2 shows examples of null correlations among the three test statistics given allele frequency.

3.3 Upper bound for the probability of the maximum of statistics

A group sequential design is developed based on repeated significance testing. Many methods have been developed to adjust for simultaneous significance tests in order to control for overall type I error level. The rejection rule can be typically written as "reject H_0 if any of the tests are significant". In a group sequential design that allows early stopping for significance only, stage 2 is performed only if stage 1 is not significant. Let there be J looks ($J - 1$ interim analyses and one final analysis) in the study and denote T_j as the statistics used at the jth stage and c_j

as the corresponding critical value. Let $\text{MAX} = \max(T_1, T_2, \cdots, T_J)$ for the J test statistics, and $E_j = \{T_j > c_j\}$ for $j = 1, \ldots, J$. Using sequential partitioning, we have

$$P(\text{MAX} > c) = P(\bigcup_j E_j) = P(E_1) + P(\bar{E}_1 E_2) + \ldots + P(\bar{E}_1 \bar{E}_2 \ldots \bar{E}_{J-2} \bar{E}_{J-1} E_J),$$

which can be approximated by

$$\begin{aligned} P(\text{MAX} > c) &\leq P(E_1) + P(\bar{E}_1 E_2) + P(\bar{E}_1 \bar{E}_2 E_3) + \ldots + P(\bar{E}_{J-2} \bar{E}_{J-1} E_J) \quad (3.3) \\ &\leq P(E_1) + P(\bar{E}_1 E_2) + P(\bar{E}_2 E_3) + \ldots + P(\bar{E}_{J-1} E_J). \quad (3.4) \end{aligned}$$

The above inequalities (3.3) and (3.4) give upper bounds of the probability $P(\text{MAX} > c)$. In (3.4), the equality holds if $\bar{E}_{j-1} E_j = \bar{E}_1 \bar{E}_2 \ldots \bar{E}_{j-1} E_j$ for all j. Under the assumption that the test statistics $(T_1, T_2, \ldots, T_J)$ follow standard multivariate normal distributions under the null hypothesis H_0, Efron [1997] summarized three approaches that determine the upper bounds of the two-point probability formula (3.4) in terms of the null correlations among the J statistics. These bounds usually beat the Bonferroni bound and can be good approximations to the actual value of $P(\text{MAX} > c)$ in favorable circumstances. The three approaches of calculating the term $P(\bar{E}_{j-1} E_j)$ are summarized below.

(i) *Two-point formula via numerical integration*: This formula can be derived through numerical integrating the bivariate normal distribution. Since

$$\begin{aligned} P(\bar{E}_{j-1} E_j) &= P(T_{j-1} \leq c_{j-1} \ \ and \ \ T_j > c_j) \\ &= \int_{-\infty}^{c_{j-1}} \int_{c_j}^{\infty} \frac{1}{2\pi\sqrt{1-\rho_j^2}} \cdot e^{\frac{t_{j-1}^2 - 2\rho_j t_{j-1} t_j + t_j^2}{2(1-\rho_j^2)}} dt_j \cdot dt_{j-1} \\ &= \int_{-\infty}^{c_{j-1}} \int_{c_j}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho_j^2}} \cdot e^{\frac{(t_j - \rho_j t_{j-1})^2}{2(1-\rho_j^2)}} dt_j \frac{1}{\sqrt{2\pi}} e^{-\frac{t_{j-1}^2}{2}} dt_{j-1} \\ &= \int_{-\infty}^{c_{j-1}} \bar{\Phi}\{\frac{c_j - \rho_j t}{(1-\rho_j^2)^{\frac{1}{2}}}\} \phi(t) dt \quad (3.5) \end{aligned}$$

where $\rho_j = cor(T_{j-1}, T_j)$, the correlation between T_j and T_{j-1}, ϕ and Φ are the density and distribution functions of $N(0,1)$, respectively, and $\bar{\Phi} = 1 - \Phi$.

(ii) *Length formula*: It further reduces the two-point probability $P(\bar{E}_{j-1}\bar{E}_j)$ into a one-dimensional funtion after transforming the bivariate normal distribution function into polar coordinates. Let the length be defined as $L_j = \arccos(\rho_j)$. Then,

$$P(\bar{E}_{j-1}E_j) \leq e^{-c^2/2}\frac{L_j}{2\pi} \tag{3.6}$$

Below is a brief proof:

Proof: Considering the case $J = 2$, where we observe test statistics T_1 and T_2 each $N(0,1)$ under H_0 and with null hypothesis correlation ρ. The Length is thus $L = arccos(\rho)$.

Let $Z \sim N_2(0, I)$. We can represent T_1 and T_2 as inner products, $T_j = \gamma_j' Z (j = 1, 2)$, where γ_1 and γ_2 are unit vectors with $\gamma_1'\gamma_2 = cos(L) = \rho$. For a fixed value of c, let

$$E_j = \{T_j > c\},\ \bar{E}_j = \{T_j \leq c\}$$

$$T = \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = \begin{pmatrix} \gamma_1' Z \\ \gamma_2' Z \end{pmatrix}$$

$$E(T) \ = \ \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$Var(T) \ = \ \begin{pmatrix} 1 & \gamma_1'\gamma_2 \\ \gamma_2'\gamma_1 & 1 \end{pmatrix}$$

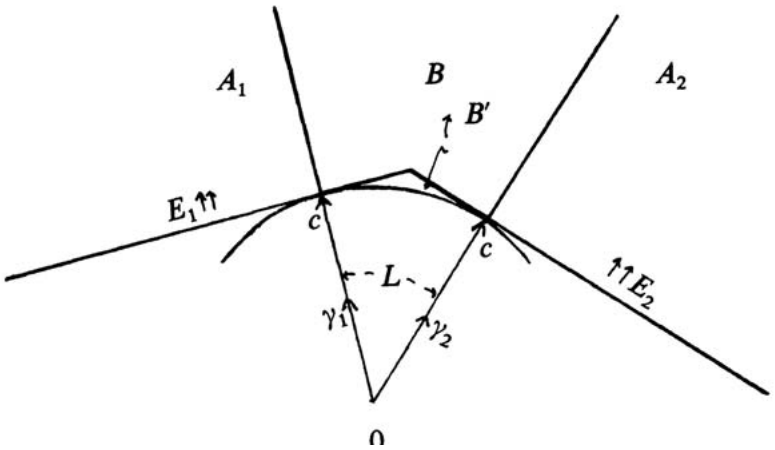
In terms of the vecter Z,

$$\begin{pmatrix} Z_1 \cos\theta_1 + Z_2 \sin\theta_1 \\ Z_1 \cos\theta_2 + Z_2 \sin\theta_2 \end{pmatrix}$$

Given θ_i, in terms of the vecter Z, E_j is a half space with the boundary perpendicular to γ_i, intersecting the ray $R_j := \{x\gamma_j, x > 0\}$ at distance c from the origin. The

event of interest $\{T_{max} > c\} = E_1 \cup E_2$ (Figure 3.1). We can decompose $\{T_{max} > c\}$ into 3 disjoint sets,

Figure 3.1: Event $\{T_{max} > c\} = E_1 \cup E_2$ in terms of the standard normal vector Z (Efron[1997])



$$\{T_{max} > c\} = A_1 \cup B \cup A_2,$$

where A_1 and A_2 are quadrants bounded by R_1 and R_2 respectively, and B is a W-shaped region lying between R_1 and R_2. Notice that $pr(A_1) = pr(A_2) = \bar{\Phi}(c)/2$, so

$$pr(\{T_{max} > c\} = \bar{\Phi}(c) + pr(B).$$

The W-shaped region B is containd in B', the portion of the disc-complement $\{\|Z\| > c\}$ lying between R_1 and R_2. Thus

$$pr(B) \leq pr(B') = e^{-c^2/2}\frac{L}{2\pi}$$

The probability $e^{-c^2/2}\frac{L}{2\pi}$ for B' is obtained by expressing Z in polar coordinates, since $\|Z\|^2$ has a ${\chi_2}^2$ distribution independent of the uniformly distributed direction vector $Z/\|Z\|$. This can be proved by Basu's theorem or by definition of independence.

Imagine that the $\chi^2 > c^2$ is the area outside the circle and then $L/2\pi$ restricts to sector considering the pr. outside the circle is uniform.

(iii) *W formula*: It improves the approximation given in (3.6) with another geometric inequality. The term $P(\bar{E}_{j-1}E_j)$ in (3.6) can be conveniently evaluated by numeically integrating the formula

$$P(\bar{E}_{j-1}E_j) = \frac{1}{\pi}\int_0^{L_j/2} e^{-\frac{1}{2}c^2 sec\theta} d\theta \tag{3.7}$$

Since $sec^2\theta \geq 1 + \theta^2$, (3.7) gives

$$P(\bar{E}_{j-1}E_j) \leq \frac{e^{-c^2/2}}{\pi}\int_0^{L_j/2} e^{-\frac{1}{2}c^2 sec\theta} d\theta = \phi(c)\frac{\Phi(cL_j/2) - \frac{1}{2}}{c/2}. \tag{3.8}$$

where $L_j = \arccos(\rho_j)$.

The probability of simultaneous hypothesis testing in (3.4) can be approximated accordingly. These methods can be easily adapted in the testing procedure in a group sequential design. The only difference is that the critical value varies for different stages.

3.4 Controlling type I error in the robust two-stage design

3.4.1 The design

In the two-stage robust group sequential design, $\text{MAX}_1 = \max(T_{A1}, T_{D1}, T_{R1})$ and $\text{MAX}_2 = \max(T_{A2}, T_{D2}, T_{R2})$ are applied to the two stages, respectively. In the subscripts, A, D and R stand for the additive, dominant, and recessive model, respectively; 1 and 2 stands for the stage. We can apply an appropriate alpha spending function to determine the levels α_1 and α_2 for the two stages, and then apply the upper

bounds given in (3.5)-(3.8) to determine the corresponding critical values c_1 and c_2 for the MAX statistics in the two stages. Then, in stage 1,

$$\begin{aligned} P(\text{MAX}_1 > c_1) &\leq P(T_{A1} > c_1) + P(T_{A1} \leq c_1, T_{D1} > c_1) + P(T_{D1} \leq c_1, T_{R1} > c_1) \\ &= \alpha_1, \end{aligned} \tag{3.9}$$

and in stage 2,

$$\begin{aligned} P(\text{MAX}_1 \leq c_1,\, \text{MAX}_2 > c_2) &\leq P(T_{R1} \leq c_1, T_{A2} > c_2) + P(T_{A2} \leq c_2, T_{D2} > c_2) \\ &\quad + P(T_{D2} \leq c_2, T_{R2} > c_2) = \alpha_2. \end{aligned} \tag{3.10}$$

It is apparent that the order of the test statistics $(A_i, R_i, D_i)(i = 1, 2)$, can affect the approximations in (3.9) and (3.10). While the formulas are presented in the order (A_i, R_i, D_i), that is not necessarily the case in the approximation process. In the simulation described in a later section, we chose the best order of the three test statistics (T_A, T_D, T_R) at each stage based on the case in which the upper bound is closest to the actual value, i.e., when the minimum $L = \sum_{j=2}^{J} \arccos(corr(T_{j-1}, T_j))$ is achieved, where j denotes the jth statistic in MAX.

An alpha spending function can be used to allocate the overall type I error α into two stages with α_1 and α_2. Columns 2 to 4 in Table 3.3 lists a selection of alpha spending functions [Betensky 1998] and their corresponding α_1, α_2 levels given that the overall α is 0.025 and the information fraction is half at Stage 1. Note that the first two alpha functions, $\alpha_1(\tau)$ and $\alpha_2(\tau)$, are continuous forms of the alpha spending values proposed by O'Brien and Fleming [1979] and Pocock [1977], respectively, which are discretized at equal increments of information. The O'Brien-Fleming type of spending function is more commonly used in conservative situations because it allows very small spending of alpha in the beginning of the study, i.e. very early termination requires a dramatic difference, compared to aggressive spending at earlier stages in the Pocock type of spending function. The characteristics of different alpha spending functions and choice of one have been studied by researchers such as Kim and DeMets

[2002] and Selwyn and Fish [2004]; the trade off between early stopping and overall power as well as minimization of the sample sizes are taken into consideration when making a choice. Regardless of the choice of alpha spending function, for a given α_1 and α_2, the critical values c_1 and c_2 in (3.9) and (3.10) can be obtained by the Newton-Raphson method described in next section. Since the critical values resulting from a two-sided approach are more complicated and yet have minimal impact on the results (see the results section), the critical values that are computed based on the right tail one-sided test at the 0.025 level are also used for two-sided tests.

3.4.2 Newton-Raphson methods for computing critical vaules from the two-point formula

The Newton-Raphson method uses an iterative process to approach one root of a function. Given an equation $f(x) = 0$ and initial value x_1, the root can be located through iterative steps $x_{n+1} = x_n - f(x_n)/f'(x_n)$ until convergence.

Consider the two-point equation (3.5)

$$\begin{aligned} f(c_j) &= P(T_{j-1} \le c_{j-1}, T_j \ge c_j) \\ &= \int_{-\infty}^{c_{j-1}} \bar{\Phi}\left(\frac{c_j - \rho_j t}{(1-\rho_j^2)^{1/2}}\right)\phi(t)dt \end{aligned}$$

where c_j is the critical value and ρ_j is the correlative between T_{j-1} and T_j statistics in MAX. The first derivatives can be derived as below considering two scenarios:

(1) $c_{j-1} \neq c_j$ (c_{j-1} is considered as a constant)

$$\begin{aligned} f'(c_j) &= \frac{d}{dc_j}\int_{-\infty}^{c_{j-1}} \bar{\Phi}(\frac{c_j - \rho_j t}{(1-\rho_j^2)^{1/2}})\phi(t)dt \\ &= \int_{-\infty}^{c_{j-1}} \frac{d}{dc_j}\bar{\Phi}\left(\frac{c_j - \rho_j t}{(1-\rho_j^2)^{1/2}}\right)\phi(t)dt \\ &= \int_{-\infty}^{c_{j-1}} -\phi\left(\frac{c_j - \rho_j t}{(1-\rho_j^2)^{1/2}}\right)\frac{1}{(1-\rho_j^2)^{1/2}}\phi(t)dt \end{aligned}$$

(2) $c_{j-1} = c_j$, then $f(c_j) = \int_{-\infty}^{c_j} \bar{\Phi}\left(\frac{c_j - \rho_j t}{(1-\rho_j^2)^{1/2}}\right)\phi(t)dt$, taking derivative with regard

to c_j, we have:

$$f'(c_j) = \bar{\Phi}\left(\frac{c_j - \rho_j t}{(1-\rho_j^2)^{1/2}}\right)\phi(c_j) + \int_{-\infty}^{c_j} -\phi\left(\frac{c_j - \rho_j t}{(1-\rho_j^2)^{1/2}}\right)\frac{1}{(1-\rho_j^2)^{1/2}}\phi(t)dt$$

Applying the above results with more appropriate notations in a two-stage study, the function and the first derivative at each stage can be summarized as the following:

MAX_i: MAX statistic for Stage i: $\text{MAX}_i = max(T_{i1}, T_{i2}, T_{i3})$, $i = 1, 2$

T_{ij}: jth order statistic in MAX at Stage i, $i = 1, 2; j = 1, 2, 3$

c_i : Critical value at Stage i, $i = 1, 2$

α_i: Desired type I error rate at Stage i, $i = 1, 2$

$g(c_i)$: Function of c_i, the quantity is set to α_i to solve for c_i

ρ_{ij}: Correlation coefficient between T_{ij} and $T_{i,j-1}$, $i = 1, 2$, $j = 2, 3$. As a special notation, $\rho_{13,21}$ is the correlation coefficient between T_{13} and T_{21}

Φ: Cumulative distribution function for the standard normal distribution

ϕ: Probability density function for the standard normal distribution

Stage 1:

$$\begin{aligned} P(\text{MAX}_1 > c_1) &\leq P(T_{11} > c_1) + P(T_{11} \leq c_1, T_{12} > c_1) + P(T_{12} \leq c_1, T_{13} > c_1) \\ &= \bar{\Phi}(c_1) + \sum_{j=2}^{3} \int_{-\infty}^{c_1} \bar{\Phi}\left(\frac{c_1 - \rho_{1j} t}{(1-\rho_{1j}^2)^{1/2}}\right)\phi(t)dt \\ &= g(c_1) = \alpha_1, \end{aligned}$$

and

$$\begin{aligned} g'(c_1) &= -\phi(c_1) \\ &+ \sum_{j=2}^{3}\left\{\bar{\Phi}\left(\frac{c_1 - \rho_{1j} t}{(1-\rho_{1j}^2)^{1/2}}\right)\phi(c_1) + \int_{-\infty}^{c_1} -\phi\left(\frac{c_1 - \rho_{1j} t}{(1-\rho_{1j}^2)^{1/2}}\right)\frac{1}{(1-\rho_{1j}^2)^{1/2}}\phi(t)dt\right\}. \end{aligned}$$

Stage 2:

$$
\begin{aligned}
P(\mathrm{MAX}_1 \le c_1 \quad , \quad & \mathrm{MAX}_2 > c_2) \\
&\le P(T_{13} \le c_1, T_{21} > c_2) + P(T_{21} \le c_2, T_{22} > c_2) + P(T_{22} \le c_2, T_{23} > c_2) \\
&= \int_{-\infty}^{c_1} \bar{\Phi}\left(\frac{c_2 - \rho_{13,21}t}{(1-\rho_{13,21}^2)^{1/2}}\right)\phi(t)dt + \sum_{j=2}^{3}\int_{-\infty}^{c_2} \bar{\Phi}\left(\frac{c_2 - \rho_{2j}t}{(1-\rho_{2j}^2)^{1/2}}\right)\phi(t)dt \\
&= g(c_2) = \alpha_2,
\end{aligned}
$$

and

$$
\begin{aligned}
g'(c_2) &= \int_{-\infty}^{c_1} -\phi\left(\frac{c_2 - \rho_{13,21}t}{(1-\rho_{13,21}^2)^{1/2}}\right)\frac{1}{(1-\rho_{13,21}^2)^{1/2}}\phi(t)dt \\
&+ \sum_{j=2}^{3}\left\{\bar{\Phi}\left(\frac{c_2 - \rho_{2j}t}{(1-\rho_{2j}^2)^{1/2}}\right)\phi(c_2) + \int_{-\infty}^{c_2} -\phi\left(\frac{c_2 - \rho_{2j}t}{(1-\rho_{2j}^2)^{1/2}}\right)\frac{1}{(1-\rho_{2j}^2)^{1/2}}\phi(t)dt\right\}.
\end{aligned}
$$

c_1 and c_2 can be then obtained through Newton-Raphson iterative method based on the above function and first derivative formula at each stage. The R program of this procedure is provided in the appendix.

3.5 Results

3.5.1 Critical values

Simulations were carried out under the assumption that HWE holds in the population. As discussed in earlier sections, the null correlations among different test statistics depend only on allele frequency p (see Table 3.2). The table also shows that the null correlation between the test statistics for the recessive and dominant models are the lowest, from 0.086 to 0.143 for various allele frequencies, among all the pair-wise correlations because these two are considered 'extreme' genetic models

[Freidlin *et al.* 2002, Zheng *et al.* 2002]. Based on six different choices of (α_1, α_2), Table 3.3 presents the critical values for two-stage robust group sequential studies based on the three approximation methods in (3.5)-(3.8) at allele frequency $p = 0.1$, 0.3 and 0.5. At stage 1, as expected, the Length method gives the largest critical values c_1 (the most conservative approach), whereas the two-point method gives the smallest critical values c_1 (the least conservative approach). The critical values at stage 2 (c_2) are affected by those used at stage 1 and do not necessarily show such patterns.

3.5.2 Simulation studies

In simulation studies, we considered the following four family settings: (i) One affected child only ($r = 1$, $s = 0$), i.e., trios; (ii) affected sib-pair ($r = 2$, $s = 0$); (iii) one affected and one unaffected sib-pair ($r = 1$, $s = 1$); and (iv) two affected and one unaffected sib-triple ($r = 2$, $s = 1$). For each setting, we first generated the parental genotypes for a given allele frequency p, and selected mating types that have at least one heterozygous parent. Without loss of generality, a baseline penetrance $f_0 = 0.1$ was used and, under the null hypothesis, $f_1 = f_2 = f_0$. The offspring genotypes and disease status were then generated for each mating type. This process was repeated until the targeted sample size was reached for each family setting. The total sample size used in the simulation is 200 families, and stage 1 was set at half of the sample size with 100 families. The simulated alpha levels (α_1, α_2) and the cumulative α level (type I error) were calculated based on 10,000 replicates. The results based on the three most common alpha spending functions in the first three in Table 3.3 are presented in Table 3.4 through Table 3.6 for $p = 0.1, 0.3$, and 0.5, respectivley. The columns under α are the cumulative type I errors for the one-sided or two-sided tests accordingly.

The simulation results demonstrate good control of the desired overall α levels us-

ing the methods described earlier. In contrast, the type I errors would be inflated if we used the conventional boundaries based on a single statistic with asymptotic normal distribution in a sequential setting. For example, in the family setting where $r = 1$, $s = 0$, when the critical values based on the O'Brien-Fleming method ($c_1 = 2.8029$, $c_2 = 1.9817$) were used, the cumulative type I errors for one-sided and two-sided were 0.055 and 0.098, respectively, about double of the desired rates. Similarly, using the critical values $c_1 = c_2 = 2.178$ by the Pocock method results in the cumulative type I errors of 0.058 and 0.097 for one-sided and two-sided tests, respectively. In both cases, they were about twice as much as the desired type I error (0.025 for one-sided tests and 0.05 for two-sided tests). When the results from different methods are compared, the Length method provides the most conservative critical values, whereas the two-point (3.4) method provides the critical values that lead to type I error closest to the nominal level. Since the null correlations and the "Length" are functions of the allele frequency p, the results for the Length approach also vary for different values of p. The higher the correlation, the better the approximation using the "Length".

The results also show that when more affected offspring are ascertained, the α levels are more conservatively controlled. For example, the observed α levels are lower in the family of $r = 2, s = 0$ than those in the family of $r = 1, s = 0$. Similar findings are obtained when we compare families of $r = 2, s = 1$ with families of $r = 1, s = 1$ or $r = 1, s = 0$. Moreover, when $p = 0.5$, the observed two-sided type I errors approximately double the corresponding one-sided type I errors, indicating an approximate symmetric distribution for MAX. When $p = 0.1$ or $p = 0.3$, the observed type I errors are less than double the corresponding one-sided type I errors, indicating the null distribution of MAX is skewed slightly to the left. Likewise, when $p > 0.5$, the null distribution of MAX would be skewed slightly to the right. In addition, when c_1 and c_2 are used, some type I errors in Tables 3.4 and 3.5 appear to be conservative. Thus, we also examined the potential power loss due to using conservative c_1 and c_2. A plot of power functions with different c_1 and c_2 corresponding to different

cumulative type I error at the alternatives of different genotype relative risks (GRRs) for having disease for AA compared with aa (f_2/f_0) is presented in Figure 3.6. The loss of power due to a conservative test (e.g. $c_1 = 2.6060$, $c_2 = 2.5938$ for ASF=3 using length formula) appears to be minimal when these values are shifted slighted towards getting the desired alpha level.

To access the sensitivity of the simulation results, Table 3.7 further reports the simulated type I errors with the 5th and 95th percentiles of the estimated type I errors ($\alpha(5^{th},95^{th})$) for $p = 0.3$ when the simulations are repeated 100 times. Due to excessive simulation time, we present the results for only two settings: 1) $r = 1$ and $s = 0$, the trio setting; and 2) $r = 2$ and $s = 1$. Results for $p = 0.1$ and 0.5 in the trio setting ($r = 1$ and $s = 0$) are reported in Table 3.8 and Table 3.9, respectively. These results show that the one-sided type I error levels are between .020 and .028 whereas the two-sided levels are between .035 and .053. While the one-sided levels appear to be more liberal than the two-sided ones due to the skewness of the distributions, the majority of simulated results are slightly below the desired level indicating that the approximations of the tails of MAX statistics are reasonably accurate.

In genome wide associatation studies of hundreds of genes, it is often desired to use a smaller alpha level due to multiplicity of tests. To examine the performance of approximation for a smaller α, we conducted a simulation when the nominal level was $\alpha = 0.0001$ for $p = 0.3$ and trios based on 1,000,000 replicates (Table 3.10). None of the simulated samples reached significance level at the first stage. The cumulative type I errors, when the W formula and two-point approximation were used, are 0.000104 for the two-sided tests.

3.6 Summary

With the emerging needs of implementing group sequential designs in genetic studies, in particular for large association studies, and the sensitivity of choosing

genetic model-based statistical tests, we present tools for setting up critical values in conducting two-stage robust group sequential tests in family-based association candidate-gene studies using MAX. When the MAX statistics are used sequentially in a study, we are in fact dealing with multiple statistics at multiple stages. The problem of an upper bound for type I error can be imagined as a "nested" test where the multiple statistics are nested within the sequential tests. Previous work of approximation of multivariate normal distribution [Efron 1997] is utilized not only to deal with multiple test statistics at one time point, but also to deal with the same set of statistics at sequential time points. Given alpha levels at each time point and null correlations among the statistics, the critical values are back-calculated. Te results have shown that the use of approximation methods works well in this nested situation. Having these tools, we establish a realistic setting of test criteria for conducting family-based genetic studies in the group sequential manner with MAX.

Our approaches are presented for computing group sequential boundaries that allow early stopping for significance while the type I error is controlled. These approaches can be extended to studies that allow early stopping for futility or insignificance, or those that allow stopping for both significance and insignificance. In those cases, type II error or both type I and type II errors need to be controlled. Similar to alpha spending function, beta spending functions may be introduced to allocate type II errors in the interim analyses. Since power perservation and robust properties of MAX compared to optimal genetic model specific tests, as single stages tests, have been demonstrated in a series of research [Gastwirth 1985, Schaid and Nick 1990, Gastwirth and Freidlin 2000, and Freidlin *et al.* 2002]. We expect to observe the same preservation of power using MAX in the group sequential setting as in a one-time test, i.e., non-sequential setting [Zheng *et al.* 2002] when type II errors are controlled at different levels in interim analyses.

Table 3.1: Correlation matrix in a two-stage sequential study

	T_{1A}	T_{1D}	T_{1R}	T_{2A}	T_{2D}	T_{2R}
T_{1A}	1	ρ_{AD}	ρ_{AR}	$\sqrt{n_1/n_2}$	$\rho_{AD}\sqrt{n_1/n_2}$	$\rho_{AR}\sqrt{n_1/n_2}$
T_{1D}		1	ρ_{DR}	$\rho_{AD}\sqrt{n_1/n_2}$	$\sqrt{n_1/n_2}$	$\rho_{DR}\sqrt{n_1/n_2}$
T_{1R}			1	$\rho_{AR}\sqrt{n_1/n_2}$	$\rho_{DR}\sqrt{n_1/n_2}$	$\sqrt{n_1/n_2}$
T_{2A}				1	ρ_{AD}	ρ_{AR}
T_{2D}					1	ρ_{DR}
T_{2R}						1

Table 3.2: Asymptotic null correlation matrices given allele frequencies (p)

$p = 0.1$

	T_D	T_R	T_A
T_D	1	0.0863	0.9608
T_R		1	0.3592
T_A			1

$p = 0.3$

	T_D	T_R	T_A
T_D	1	0.1311	0.8699
T_R		1	0.6030
T_A			1

$p = 0.5$

	T_D	T_R	T_A
T_D	1	0.1429	0.7559
T_R		1	0.7559
T_A			1

T_D: Test under dominant model
T_R: Test under recessive model
T_A: Test under additive model

Table 3.3: Critical values for one-sided tests at the overall type I error of $\alpha = 0.025$ in two-stage group sequential family-based association studies using different alpha spending functions (ASF)

ASF	Formula	α_1	α_2	Method	$p = 0.1$		$p = 0.3$		$p = 0.5$	
					c_1	c_2	c_1	c_2	c_1	c_2
1	$\alpha_1(\tau) = 2\left[1 - \Phi(Z_{\alpha/2}\tau^{-1/2})\right]$	.0015	.0235	Length	3.2957	2.3382	3.2896	2.3296	3.2880	2.3273
				W	3.2230	2.2578	3.2434	2.2743	3.2487	2.2783
				2pt	3.2176	2.2984	3.2410	2.3170	3.2470	2.3215
2	$\alpha_2(\tau) = \alpha log\left[1 + (e-1)\tau\right]$	.0155	.0095	Length	2.5257	2.6976	2.5184	2.6902	2.5165	2.6882
				W	2.4682	2.6083	2.4841	2.6276	2.4880	2.6325
				2pt	2.4624	2.5903	2.4819	2.6141	2.4867	2.6200
3	$\alpha_3(\tau) = \alpha\tau$	.0125	.0125	Length	2.6060	2.5938	2.5988	2.5860	2.5969	2.5840
				W	2.5466	2.5069	2.5631	2.5255	2.5673	2.5302
				2pt	2.5408	2.5019	2.5609	2.5245	2.5659	2.5300
4	$\alpha_4(\tau) = \alpha\tau^{3/2}$	.0088	.0162	Length	2.7307	2.4928	2.7237	2.4847	2.7219	2.4826
				W	2.6686	2.4084	2.6860	2.4262	2.6904	2.4306
				2pt	2.6627	2.4184	2.6837	2.4396	2.6890	2.4448
5	$\alpha_5(\tau) = \alpha\tau^2$	.0063	.0188	Length	2.8503	2.4324	2.8436	2.4242	2.8418	2.4220
				W	2.7858	2.3496	2.8040	2.3669	2.8086	2.3712
				2pt	2.7800	2.3701	2.8017	2.3903	2.8072	2.3953
6	$\alpha_6(\tau) = \begin{cases} 0.8\alpha\tau & if\ 0 \le \tau \le 0.75 \\ 1.6\alpha\tau - 0.6\alpha & if\ 0.75 \le \tau \le 1 \end{cases}$	.0100	.0150	Length	2.6869	2.5225	2.6798	2.5145	2.6780	2.5124
				W	2.6257	2.4374	2.6429	2.4554	2.6472	2.4599
				2pt	2.6199	2.4427	2.6406	2.4643	2.6458	2.4696

Table 3.4: Simulated type I errors for two-stage family-based association studies: allele frequency $p = 0.1$.

			One-sided			Two-sided		
Family	ASF	Method	α_1	α_2	α	α_1	α_2	α
$r = 1,\ s = 0$	1	Length	.0016	.0234	.0250	.0018	.0383	.0401
		W	.0024	.0272	.0296	.0029	.0461	.0490
		2pt	.0024	.0252	.0276	.0029	.0413	.0442
	2	Length	.0169	.0062	.0231	.0222	.0093	.0315
		W	.0192	.0071	.0263	.0271	.0113	.0384
		2pt	.0199	.0075	.0274	.0278	.0119	.0397
	3	Length	.0144	.0080	.0224	.0186	.0128	.0314
		W	.0163	.0108	.0271	.0215	.0171	.0386
		2pt	.0163	.0108	.0271	.0215	.0171	.0386
$r = 1,\ s = 1$	1	Length	.0016	.0240	.0256	.0021	.0390	.0411
		W	.0020	.0290	.0310	.0025	.0486	.0511
		2pt	.0020	.0263	.0283	.0025	.0436	.0461
	2	Length	.0166	.0070	.0236	.0242	.0102	.0344
		W	.0193	.0095	.0288	.0275	.0138	.0413
		2pt	.0196	.0096	.0292	.0281	.0142	.0423
	3	Length	.0135	.0103	.0238	.0197	.0150	.0347
		W	.0158	.0125	.0283	.0230	.0187	.0417
		2pt	.0161	.0127	.0288	.0233	.0191	.0424
$r = 2,\ s = 0$	1	Length	.0016	.0225	.0241	.0019	.0419	.0438
		W	.0019	.0263	.0282	.0025	.0506	.0531
		2pt	.0019	.0241	.0260	.0025	.0459	.0484
	2	Length	.0179	.0055	.0234	.0280	.0101	.0381
		W	.0204	.0066	.0270	.0319	.0125	.0444
		2pt	.0205	.0068	.0273	.0320	.0131	.0451
	3	Length	.0141	.0076	.0217	.0219	.0145	.0364
		W	.0169	.0093	.0262	.0262	.0188	.0450
		2pt	.0176	.0091	.0267	.0275	.0184	.0459
$r = 2,\ s = 1$	1	Length	.0019	.0237	.0256	.0025	.0422	.0447
		W	.0024	.0286	.0310	.0032	.0529	.0561
		2pt	.0024	.0256	.0280	.0032	.0472	.0504
	2	Length	.0159	.0064	.0223	.0258	.0109	.0367
		W	.0184	.0080	.0264	.0302	.0145	.0447
		2pt	.0186	.0084	.0270	.0307	.0152	.0459
	3	Length	.0131	.0091	.0222	.0203	.0163	.0366
		W	.0153	.0110	.0263	.0246	.0205	.0451
		2pt	.0154	.0114	.0268	.0248	.0212	.0460

ASF = Alpha Spending Function.
Note: The trio setting is a special case with $r = 1$ and $s = 0$.

Table 3.5: Simulated Type I errors for two-stage family-based association studies: allele frequency $p = 0.3$.

			One-sided			Two-sided		
Family	ASF	Method	α_1	α_2	α	α_1	α_2	α
r=1, s=0	1	Length	.0012	.0237	.0249	.0023	.0463	.0486
		W	.0014	.0272	.0286	.0026	.0525	.0551
		2pt	.0015	.0241	.0256	.0027	.0473	.0500
	2	Length	.0135	.0072	.0207	.0265	.0131	.0396
		W	.0153	.0088	.0241	.0291	.0159	.0450
		2pt	.0153	.0091	.0244	.0291	.0164	.0455
	3	Length	.0105	.0104	.0209	.0213	.0191	.0404
		W	.0123	.0121	.0244	.0238	.0226	.0464
		2pt	.0123	.0121	.0244	.0238	.0226	.0464
r=1, s=1	1	Length	.0009	.0237	.0246	.0015	.0441	.0456
		W	.0011	.0272	.0283	.0019	.0508	.0527
		2pt	.0011	.0242	.0253	.0019	.0450	.0469
	2	Length	.0155	.0064	.0219	.0261	.0123	.0384
		W	.0170	.0076	.0246	.0285	.0148	.0433
		2pt	.0170	.0079	.0249	.0285	.0153	.0438
	3	Length	.0120	.0093	.0213	.0204	.0176	.0380
		W	.0134	.0110	.0244	.0225	.0205	.0430
		2pt	.0136	.0111	.0247	.0227	.0207	.0434
r=2, s=0	1	Length	.0012	.0203	.0215	.0024	.0420	.0444
		W	.0015	.0245	.0260	.0029	.0493	.0522
		2pt	.0015	.0210	.0225	.0029	.0429	.0458
	2	Length	.0131	.0053	.0184	.0271	.0108	.0379
		W	.0148	.0060	.0208	.0309	.0132	.0441
		2pt	.0148	.0062	.0210	.0310	.0134	.0444
	3	Length	.0100	.0082	.0182	.0204	.0169	.0373
		W	.0110	.0094	.0204	.0225	.0200	.0425
		2pt	.0110	.0094	.0204	.0226	.0199	.0425
r=2, s=1	1	Length	.0016	.0193	.0209	.0029	.0405	.0434
		W	.0018	.0223	.0241	.0034	.0457	.0491
		2pt	.0019	.0200	.0219	.0035	.0416	.0451
	2	Length	.0151	.0053	.0204	.0278	.0126	.0404
		W	.0158	.0070	.0228	.0304	.0153	.0457
		2pt	.0158	.0070	.0228	.0304	.0154	.0458
	3	Length	.0118	.0080	.0198	.0218	.0179	.0397
		W	.0132	.0093	.0225	.0241	.0201	.0442
		2pt	.0132	.0093	.0225	.0241	.0201	.0442

ASF = Alpha Spending Function.
Note: The trio setting is a special case with $r = 1$ and $s = 0$.

Table 3.6: Simulated Type I errors for two-stage family-based association studies: allele frequency $p = 0.5$.

			One-sided			Two-sided		
Family	ASF	Method	α_1	α_2	α	α_1	α_2	α
r=1, s=0	1	Length	.0009	.0256	.0265	.0022	.0467	.0489
		W	.0011	.0284	.0295	.0025	.0525	.0550
		2pt	.0011	.0259	.0270	.0025	.0471	.0496
	2	Length	.0145	.0063	.0208	.0272	.0121	.0393
		W	.0158	.0081	.0239	.0300	.0148	.0448
		2pt	.0158	.0084	.0242	.0300	.0153	.0453
	3	Length	.0115	.0102	.0217	.0219	.0189	.0408
		W	.0125	.0122	.0247	.0238	.0223	.0461
		2pt	.0125	.0122	.0247	.0240	.0222	.0462
r=1, s=1	1	Length	.0008	.0239	.0247	.0029	.0464	.0493
		W	.0010	.0263	.0273	.0031	.0519	.0550
		2pt	.0010	.0242	.0252	.0031	.0470	.0501
	2	Length	.0129	.0083	.0212	.0273	.0147	.0420
		W	.0141	.0096	.0237	.0297	.0168	.0465
		2pt	.0141	.0098	.0239	.0297	.0173	.0470
	3	Length	.0097	.0116	.0213	.0210	.0205	.0415
		W	.0103	.0140	.0243	.0227	.0248	.0475
		2pt	.0105	.0140	.0245	.0230	.0247	.0477
r=2, s=0	1	Length	.0010	.0211	.0221	.0026	.0478	.0504
		W	.0011	.0236	.0247	.0028	.0530	.0558
		2pt	.0011	.0215	.0226	.0028	.0483	.0511
	2	Length	.0144	.0057	.0201	.0294	.0124	.0418
		W	.0159	.0064	.0223	.0321	.0139	.0460
		2pt	.0159	.0069	.0228	.0321	.0150	.0471
	3	Length	.0108	.0094	.0202	.0226	.0192	.0418
		W	.0117	.0101	.0218	.0245	.0210	.0455
		2pt	.0119	.0100	.0219	.0248	.0209	.0457
r=2, s=1	1	Length	.0008	.0211	.0219	.0020	.0462	.0482
		W	.0011	.0242	.0253	.0024	.0524	.0548
		2pt	.0012	.0211	.0223	.0025	.0464	.0489
	2	Length	.0136	.0056	.0192	.0283	.0109	.0392
		W	.0147	.0071	.0218	.0307	.0137	.0444
		2pt	.0152	.0070	.0222	.0315	.0140	.0455
	3	Length	.0110	.0085	.0195	.0229	.0175	.0404
		W	.0118	.0096	.0214	.0245	.0204	.0449
		2pt	.0118	.0096	.0214	.0246	.0203	.0449

ASF = Alpha Spending Function.
Note: The trio setting is a special case with $r = 1$ and $s = 0$.

Table 3.7: Simulated Type I errors with ranges for two-stage group sequential family-based association studies: allele frequency $p = 0.3$.

		One-sided			Two-sided		
ASF	Method	α_1	α_2	$\alpha\ (5^{th}, 95^{th})$	α_1	α_2	$\alpha\ (5^{th}, 95^{th})$
				r=1, s=0			
1	Length	.00142	.02294	.02436 (.02235, .02710)	.00242	.04383	.04625 (.04325, .04910)
	W	.00166	.02627	.02793 (.02545, .03095)	.00285	.05026	.05310 (.04970, .05680)
	2pt	.00168	.02366	.02534 (.02320, .02810)	.00288	.04523	.04811 (.04520, .05125)
2	Length	.01466	.00597	.02063 (.01825, .02330)	.02687	.01133	.03819 (.03515, .04195)
	W	.01620	.00713	.02332 (.02095, .02620)	.02976	.01362	.04338 (.04025, .04745)
	2pt	.01625	.00748	.02374 (.02140, .02660)	.02988	.01431	.04419 (.04075, .04825)
3	Length	.01191	.00889	.02080 (.01855, .02355)	.02161	.01692	.03853 (.03530, .04225)
	W	.01313	.01057	.02369 (.02125, .02675)	.02392	.02004	.04396 (.04065, .04780)
	2pt	.01318	.01056	.02374 (.02130, .02675)	.02401	.02002	.04403 (.04070, .04795)
				r=2, s=1			
1	Length	.00160	.02336	.02495 (.02290, .02700)	.00261	.04505	.04766 (.04440, .05025)
	W	.00187	.02653	.02840 (.02650, .03065)	.00306	.05145	.05451 (.05090, .05720)
	2pt	.00189	.02392	.02580 (.02375, .02790)	.00308	.04624	.04932 (.04575, .05190)
2	Length	.01533	.00628	.02161 (.01930, .02365)	.02840	.01198	.04039 (.03705, .04390)
	W	.01675	.00748	.02422 (.02210, .02650)	.03124	.01434	.04558 (.04255, .04930)
	2pt	.01596	.00725	.02321 (.02015, .02570)	.02992	.01408	.04399 (.04125, .04740)
3	Length	.01163	.00850	.02012 (.01735, .02245)	.02173	.01644	.03817 (.03585, .04080)
	W	.01283	.01006	.02289 (.01960, .02525)	.02395	.01954	.04349 (.04075, .04645)
	2pt	.01305	.01057	.02363 (.02125, .02625)	.02479	.02035	.04514 (.04160, .04855)

ASF = Alpha Spending Function.

Table 3.8: Simulated Type I errors with ranges for two-stage group sequential family-based association studies: allele frequency $p = 0.1$ in the trio setting ($r = 1$, $s = 0$).

		One-sided			Two-sided		
ASF	Method	α_1	α_2	$\alpha\ (5^{th}, 95^{th})$	α_1	α_2	$\alpha\ (5^{th}, 95^{th})$
1	Length	.00185	.02455	.02640 (.02410, .02965)	.00245	.04082	.04327 (.04025, .04685)
	W	.00237	.02957	.03195 (.02905, .03600)	.00316	.05019	.05335 (.04910, .05710)
	2pt	.00241	.02657	.02897 (.02645, .03280)	.00321	.04486	.04807 (.04495, .05155)
2	Length	.01713	.00681	.02394 (.02175, .02625)	.02474	.01062	.03536 (.03240, .03855)
	W	.01897	.00878	.02775 (.02540, .03050)	.02812	.01411	.04223 (.03905, .04540)
	2pt	.01969	.00928	.02896 (.02660, .03190)	.02888	.01489	.04377 (.04075, .04740)
3	Length	.01347	.00997	.02343 (.02130, .02590)	.01944	.01596	.03540 (.03270, .03845)
	W	.01606	.01247	.02854 (.02645, .03140)	.02314	.02042	.04355 (.04005, .04675)
	2pt	.01626	.01253	.02879 (.02650, .03160)	.02345	.02052	.04398 (.04065, .04705)

ASF = Alpha Spending Function.

Table 3.9: Simulated Type I errors with ranges for two-stage group sequential family-based association studies: allele frequency $p = 0.5$ in the trio setting ($r = 1$, $s = 0$).

		One-sided			Two-sided		
ASF	Method	α_1	α_2	$\alpha\ (5^{th}, 95^{th})$	α_1	α_2	$\alpha\ (5^{th}, 95^{th})$
1	Length	.00119	.02271	.02390 (.02155, .02650)	.00248	.04509	.04757 (.04460, .05135)
	W	.00139	.02563	.02702 (.02460, .02965)	.0029	.05088	.05378 (.05045, .05785)
	2pt	.00139	.02285	.02424 (.02210, .02670)	.0029	.04535	.04825 (.04515, .05205)
2	Length	.01394	.00587	.01981 (.01765, .02195)	.02782	.01191	.03974 (.03610, .04240)
	W	.01523	.007	.02224 (.02000, .02440)	.03055	.01386	.04440 (.04115, .04775)
	2pt	.01523	.00736	.02259 (.02015, .02470)	.03055	.01459	.04513 (.04175, .04850)
3	Length	.01118	.00887	.02005 (.01790, .02235)	.02241	.01761	.04001 (.03665, .04375)
	W	.01219	.01024	.02243 (.01985, .02495)	.0245	.02029	.04479 (.04100, .04885)
	2pt	.01225	.01023	.02248 (.01985, .02500)	.02464	.02025	.04489 (.04110, .04885)

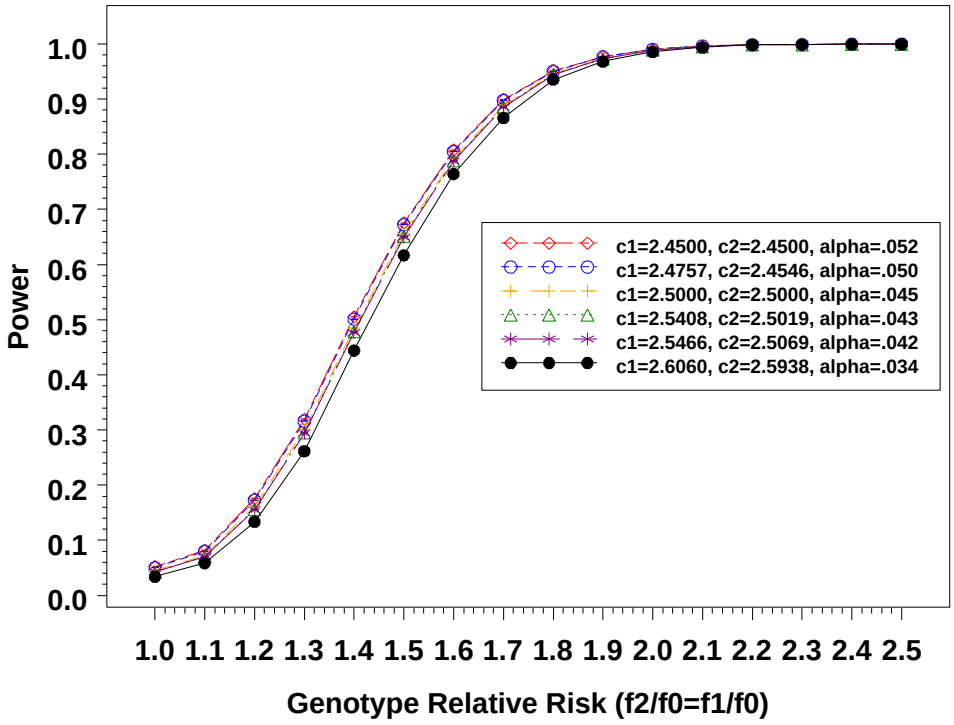
ASF = Alpha Spending Function.

Table 3.10: Simulated Type I errors for two-Stage family-based association studies at desired overall alpha level of .0001 using O'Brien-Fleming spending function (ASF=1): allele frequency $p = 0.3$ in the trio setting ($r = 1$, $s = 0$).

	No. of replicates=1,000,000, Total Sample Size=200 ($n_1 = n_2 = 100$)							
Method	**Critical Value**		**One-sided Test**			**Two-sided Test**		
	c_1	c_2	α_1^*	α_2	α	α_1^*	α_2	α
Length	**5.86528**	**4.21094**	0	.000046	.000046	0	.000065	.000065
W	**5.79399**	**4.12424**	0	.000076	.000076	0	.000104	.000104
2pt	**5.79226**	**4.12663**	0	.000076	.000076	0	.000104	.000104

*No rejection was observed in the simulations at stage 1.
ASF=Alpha Spending Function

Figure 3.2: Power Comparisons for different sets of critical values in a two-stage family-based study



Chapter 4

Two-Stage Designs for Linkage Studies

4.1 Introduction

One of the shortcomings of the tools proposed in Chapter 3 for obtaining critical values in a two-stage group sequential study is that they are based on right-tailed one-sided tests. Because the distribution of MAX can be skewed, the precision in a two-sided test may be compromised. In contrast, for linkage studies, since one-sided tests are suggested as an appropriate means of statistically testing the hypothesis that the recombination fraction is significantly less than 1/2, it is suitable to apply the approximation methods described in Chapter 3.

In this chapter, we focus on linkage studies that consist of data on affected relative pairs. These designs have been shown to be efficient and powerful in detecting linkage between the disease and suspected genetic marker. We also focus on the statistical tests based on allele shared identical-by-descent (IBD) on a single locus among relative pairs. These tests do not require knowledge of the genetic inheritance model, i.e., dominant, recessive, additive or multiplicative. The basic idea of ASP linkage analysis is to test whether the inheritance pattern of a marker deviates from Mendelian control distribution of the number of marker alleles shared IBD. The common statistical test

for linkage is a one-degree-of-freedom score based test on the weighted sum of the estimated proportion of allele shared IBD among the sibling pairs. There are two tests that have been quoted in the literature: the proportions test and the means test. The proportions test tests a single proportion of the IBD sharing, i.e. IBD=2, whereas the means test combines the IBD sharing of one and two alleles. Although the tests are not dependent on genetic models, there is not an optimal test for all underlying specifications of the parameters including genetic models. A "maxmin" test was introduced by Whittemore and Tu [1998] to minimize the power loss with an optimal weight. This test is equivalent to MERT in the ASP setting,which average the means and the proportions test statistics taking into account their null correlations.

With the similar idea described earlier in Chapter 3, we introduce the MAX test, which is the maximum across the three test statistics (means test, proportions test, and MERT) and investigate its application and characteristics under a group sequential design. Therefore, this chapter extends the methods described in the earlier chapter for obtaining critical values for a group sequential design to linkage studies, where affected relative pairs are used. In addition to confirming the appropriateness of the application of the methods that preserve the type I errors for linkage analysis, we compare the statistical power among different models using different test statistics.

4.2 Tests in linkage studies of affected relative pairs and their correlations

When affected relative pairs are ascertained in the study, the observed numbers of pairs n_0, n_1, and n_2, respectively, sharing 0, 1 and 2 alleles on the suspected locus are distributed as trinomial distribution $Mul(N; p_0, p_1, p_2)$, where $N = n_0+n_1+n_2$. Under the null hypothesis of no linkage, (p_0, p_1, p_2) have the value (a_0, a_1, a_2), $a_0+a_1+a_2 = 1$. (a_0, a_1, a_2) are constants under the Mendalian laws of inheritance according to the relationship of the relative pairs (see Table 4.1) when there is no linkage. For example,

Table 4.1: Probabilities that a pair of relative share 0, 1, 2 alleles IBD at an autosomal (non-sex chromosome) locus.

	Probability of sharing IBD alleles		
Type of Relative Pairs	0 (a_0)	1 ($a_1 v$)	2 (a_2)
Monozygotic twins	0	0	1
Full sibs	1/4	1/2	1/4
Parent-offspring	0	1	0
First cousin	3/4	1/4	0
Double first cousin	13/16	1/8	1/16
Grandparent-grandchild	1/2	1/2	0
Half sibs	1/2	1/2	0
Avuncular (aunt/uncle-niece/nephew)	1/2	1/2	0

(a_0, a_1, a_2) are $(1/4, 1/2, 1/4)$ for the full sib pairs.

When there is an indication of linkage, the observed values of p_1 and p_2 are expected to be larger than a_1 and a_2. Therefore, the alternative hypothesis is one-sided, i.e., $p_1 > a_1$ or $p_2 > a_2$. The proportions test uses the proportions of two alleles IBD as the statistic and tests the null H_0: $p_2 = a_2$ versus the alternative H_1: $p_2 > a_2$. Likewise, the means test combines information from p_1 and p_2 and tests the null H_0: $p_2 + \frac{1}{2}p_1 = a_2 + \frac{1}{2}a_1$ against the alternative H_1: $p_2 + \frac{1}{2}p_1 > a_2 + \frac{1}{2}a_1$. The test statistics from both the proportions test and the means test, which are in forms of linear combinations of proportions, can be viewed to have asymptotically normal distribution and can be tested using standardized Z test statistic. The means and the proportions tests are equivalent for majority types of relative pairs where there are either 0% or 100% probability of sharing 2 alleles IBD. These statistical tests are different for ASP studies, the most popular linkage study design. In most scenarios, the means test has been preferred as it's shown to be locally optimal (Blackwelder and Elston [1985], Knapp *et al.* [1994], Li and Gastwirth [2003]) when comparing to the proportions test. However, this is not always true. The linear combination form

of the tests can be reparametized based on the constraints in the trinomial setting following:

$$(p_0, p_1, p_2) = \lambda(0, a, 1-a) + (1-\lambda)(a_0, a_1, a_2); 0 \leq \lambda \leq 1. \tag{4.1}$$

Under the null hypothesis, H_0: $\lambda = 0$, the choice of a can be anywhere from 0 and 0.5 for practical reason, and λ is between 0 and 1. The test statistic is score based and the general form is derived in the following:

The individual proportions are $p_0 = a_0(1-\lambda)$, $p_1 = \lambda a + a_1(1-\lambda)$, and $p_2 = \lambda(1-a) + a_2(1-\lambda)$. Then the likelihood function is $L \propto p_0^{n_0} p_1^{n_1} p_2^{n_2}$ and the log likelihood is:

$$l(\lambda) \propto n_0 log(1-\lambda) + n_1 log[(\lambda a + a_1(1-\lambda)] + n_2 log[\lambda(1-a) + a_2(1-\lambda)].$$

Taking the first and second derivatives and set λ=0 (null),

$$l'(\lambda) = \frac{n_0}{1-\lambda}(-1) + \frac{n_1}{\lambda a + a_1(1-\lambda)}(a - a_1) + \frac{n_2}{\lambda(1-a) + a_2(1-\lambda)}(1 - a - a_2)$$

$$l'(0) = -n_0 + n_1\frac{a}{a_1} - n_1 + n_2\frac{1-a}{a_2} - n_2$$

$$l''(\lambda) = \frac{n_0}{(1-\lambda)^2}(-1) + \frac{n_1(-1)}{(\lambda a + a_1(1-\lambda))^2}(a - a_1)^2 + \frac{n_2(-1)}{(\lambda(1-a) + a_2(1-\lambda))^2}(1 - a - a_2)^2$$

$$l''(0) = (-1)[n_0 + n_1(\frac{a}{a_1} - 1)^2 + n_2(\frac{1-a}{a_2} - 1)^2]$$

Since $E(n_0) = a_0 n$, $E(n_1) = a_1 n$, and $E(n_2) = a_2 n$, then

$$E[-l''(0)] = n[a_0 + a_1(\frac{a}{a_1} - 1)^2 + a_2(\frac{1-a}{a_2} - 1)^{)}]$$

Let $\hat{p_0} = n_0/n$, $\hat{p_1} = n_1/n$, and $\hat{p_2} = n_2/n$, then the efficient score statistic is then

$$T(a) = \frac{l'(0)}{\sqrt{E[-l''(0)]}} = \frac{\sqrt{n}[-\hat{p_0} + (\frac{a}{a_1} - 1)\hat{p_1} + (\frac{1-a}{a_2} - 1)\hat{p_2}]}{\sqrt{a_0 + a_1(\frac{a}{a_1} - 1)^2 + a_2(\frac{1-a}{a_2} - 1)^2}} \tag{4.2}$$

For ASP studies where $(a_0, a_1, a_2) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$, the test statistic is:

$$T_{ASP}(a) = \frac{\sqrt{n}[4(a-1)\hat{p_0} + (6a-4)\hat{p_1} + (3-4a)]}{\sqrt{6a^2 - 8a + 3}} \tag{4.3}$$

When $a = 0$, the test is equivalent to the proportions test; when $a = 0.5$, it is equivalent to the means test. Whittemore and Tu [1998] demonstrated that in the ASP setting, there is a minimax test, when a is set at 0.355. This statistic, for the test of ASPs, is equivalent to the maximum efficient robust test (MERT) proposed in Gastwirth [1966]. Denote T_1 for the proportions test, T_2 for the means test, and T_3 for the MERT test between T_1 and T_2. The test statistics of these three tests in general forms and then their specific forms for ASP studies are written as below:

$$\begin{aligned}
T_1 &= T(a=0) = \frac{\sqrt{n}(\hat{p_2} - a_2)}{\sqrt{a_2(1-a_2)}} = \frac{\sqrt{n}(4\hat{p_2} - 1)}{\sqrt{3}} \\
T_2 &= T(a=.5) = \frac{\sqrt{n}(\frac{1}{2a_1}\hat{p_1} + \frac{1}{2a_2}\hat{p_2} - 1)}{\sqrt{\frac{1}{4a_1} + \frac{1}{4a_2} - 1}} = \frac{\sqrt{n}(-2\hat{p_0} - \hat{p_1} + 1)}{\sqrt{1/2}} \\
T_3 &= T(a=.355) = \frac{T_1 + T_2}{\sqrt{2(1+\rho)}}
\end{aligned}$$

where ρ is the correlation between T_1 and T_2 under the null hypothesis, i.e., $\rho = \frac{1/2 - a_2}{\sqrt{a_2(1-a_2)(\frac{1}{4a_1} + \frac{1}{4a_2} - 1)}}$. For ASP studies, $\rho = \sqrt{2/3} \approx 0.816$. The correlation between T_1 and T_3 and that between T_2 and T_3 are the same, i.e., $corr(T_1, T_3) = corr(T_2, T_3) = \sqrt{\frac{1+\rho}{2}} \approx 0.953$. Note that the null correlations are fixed constants for linkage studies with the specified type of relative pairs.

The parameters for a group sequential design for a linkage study follow the traditional ones, similar to those described in Chapter 3. Let's denote τ_j the information fraction that is obtained at jth stage, $j = 1...k$ and k is the total number of looks planned in the study. This fraction is typically determined by the sample size ratio, n_j/n_k, where n_j is the sample size at Stage j and $n_k = N$ is the total sample size. Let T_{1j}, T_{2j}, and T_{3j} be the statistics for the means, the proportions and the MERT at stage j, respectively. The correlation between the two statistics at Stage i and

Stage j is a fraction multiplier ($\sqrt{\tau_i/\tau_j}$) of their correlation described above, i.e., $corr(T_{xi}, T_{yj}) = \sqrt{\tau_i/\tau_j} \cdot corr(T_x, T_y)$, where $x = 1, 2, 3, y = 1, 2, 3$.

The maximum of the means and the proportions test (MAX2) was initially studied by Schaid and Nick [1990]. Gastwirth and Freidlin [2000] compared the power properties of MAX2 and MERT. They found that both statistics provide similar protection against loss of power relative to the optimal test when the model is uncertain. However, MAX2 showed to be more optimal for extreme models. Moving one step further and without any loss, we propose a MAX type of statistic of the means, the proportions and the MERT statistic. i.e., MAX = $\max(T_1, T_2, T_3)$. In the following sections, we conduct simulations in with restricted attention to the ASP study setting. Simulated type I errors under the null hypothesis and powers under various alternatives are presented and compared among the statistics. As stated earlier, the test is one-sided and the desired overall type I error is set at .025.

4.3 Simulation studies

Simulations were carried out for two-stage linkage studies where trinomially distributed data were generated in terms of numbers of affected sibling pairs sharing 0,1, and 2 alleles IBD. The interim look at the first stage data was performed at half of the projected sample size. The parameters of the distribution are based on the λ's and a's described in (4.1). The null hypothesis is $H_0 : \lambda = 0 \Leftrightarrow (p_0, p_1, p_2) = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Two hundred (200) pairs of affected siblings were generated in each sample and 100,000 replicates of the sample were used in estimating the Types I errors under the null and powers under various alternatives. Since the tests are one-sided, the overall desired type I error is set at .025.

4.3.1 Type I error

Table 4.2 presents the derived critical values according to the spending functions when using the MAX test statistics in a linkage study of ASPs. Similar to the alpha

spending functions described in Table 3.3: the first two spending functions are the O'Brien-Fleming (ASF=1) and the Pocock (ASF=2) types of spending functions. The latter three represent spending functions from a power family in the form of $\alpha\tau^{\Phi}$ (ASF=3, 4, and 5 in order), where α is the desired type I error level, τ is the information fraction, and Φ is the power exponent. Simulated type I errors are presented in the last three columns when simulated data were generated following the null distribution. The overall simulated type I errors ranged between .0225 and .0327, a reasonable range for the targeted .025 level. These findings demonstrated acceptable control of the desired type I error level rate. The different approximation methods, the two-point (2pt), Length and W, appeared to be mildly sensitive to the choice of spending function. As expected, the two-point method provided the most accurate estimates. The Length and the W methods are on the conservative and aggressive sides of the two-point method, respectively. The results from these two methods are also reasonable and can be appealing because of their straightforward formula.

As comparisons, in the bottom of table, we also present the simulated type I errors using (in order): the O'Brien-Fleming method, the Lan-DeMets methods simulating O'Brien-Fleming spending, the Pocock method for single asymptotically normally distributed statistics in a two-stage even information group sequential design, and lastly the unadjusted standard normal test. As expected, when the MAX statistics are used, these methods are no longer applicable and largely inflate the type I error by 50% to 150%.

4.3.2 Power

We further investigated the powers of MAX and individual statistics in the group sequential design setting. Without losing generality, we used the O'Brien-Fleming type of spending function. For the MAX statistic, we used the critical values from the more accurate two-point calculation method. For single statistics (the means, pro-

portions and MERT), the Lan-DeMets software (Lan and DeMets [1983], download can be found on the web site: *http://www.biostat.wisc.edu/landemets*) that approximates the O'Brien-Fleming spending are used for obtaining the critical values. Table 4.3 demonstrated that these critical values are appropriate in controlling type I error. In Table 4.4, the powers are presented for various alternatives. The results verified that the proportions test $T(a = 0)$ and the means test $T(a = 0.5)$ for the locally most powerful the alternative proportion follows the same pattern. The MAX statistics showed to have similar robustness against power loss to the MERT statistic and a tendency of slightly better protection when the alternative departs further from the null.

4.4 Summary

Choosing from the wide spectrum of linkage studies, this chapter focused on the most common linkage study, the study of ASPs and applied the most common non-parametric test that is based on IBD and used for diseases that do not follow a clear Mendelian pattern of inheritance. The simulation results confirmed the feasibility of conducting a group sequential study while applying a maximum type of statistics. The results showed that type I errors are preserved under the calculated critical values and that the power loss can be prevented using the MAX statistic. Although these findings are restricted to a small example of the linkage study family, the extension to other studies and different statistics is promising and can be further verified by more examples.

Table 4.2: Critical values for one-side MAX test in linkage studies and simulation results

				Simulated α Level		
ASF	Method	c_1	c_2	α_1	α_2	Overall
1. O'Brien Fleming	2pt	3.14012	2.17220	.0017	.0232	.0249
	Length	3.14542	2.12182	.0017	.0286	.0303
	W	3.14017	2.09015	.0017	.0310	.0327
2. Pocock	2pt	2.34855	2.45966	.0158	.0077	.0235
	Length	2.35191	2.51236	.0158	.0067	.0225
	W	2.34858	2.47566	.0158	.0077	.0235
3. Power Φ=1	2pt	2.43140	2.37106	.0133	.0114	.0247
	Length	2.43496	2.40052	.0133	.0103	.0236
	W	2.43143	2.36522	.0133	.0114	.0247
4. Power Φ=1.5	2pt	2.55993	2.28821	.0084	.0149	.0233
	Length	2.56379	2.29101	.0084	.0149	.0233
	W	2.55996	2.25712	.0084	.0181	.0265
5. Power Φ=2	2pt	2.68311	2.24085	.0076	.0183	.0260
	Length	2.68727	2.22521	.0067	.0186	.0253
	W	2.68315	2.19217	.0076	.0206	.0282
O'Brien-Fleming (1979)		2.8029	1.9817	.0047	.0336	.0382
Lan-DeMets Method (OF)		2.9626	1.9686	.0035	.0350	.0386
Pocock (1977)		2.1780	2.1780	.0245	.0160	.0405
Standard Normal		1.9600	1.9600	.0413	.0226	.0639

Table 4.3: Critical values used in the power study reflexing O'Brien-Fleming bound

	Simulation Results			
	MAX	T_0	$T_{0.5}$	$T_{0.355}$
C_1	3.1401	2.8029	2.8029	2.8029
C_2	2.1722	1.9817	1.9817	1.9817
α_1	.0017	.0027	.0026	.0031
α_2	.0232	.0023	.0205	.0235
Cumulative	.0249	.0261	.0231	.0266

Table 4.4: Powers of different statistics while type I error is maintained under the null

Parameters		**MAX**	T_0 (Means)	$T_{.5}$ (Proportions)	$T_{.355}$ (MERT)
$\lambda = .1$	$a = 0$	.6268	.5252	.6469	.6313
	$a = .2$	.4092	.3723	.4080	.4263
	$a = .3$	.3106	.2997	.2952	.3265
	$a = .5$	.1537	.1747	.1203	.1596
$\lambda = .2$	$a = 0$	.9933	.9729	.9947	.9923
	$a = .2$	.9219	.8865	.9211	.9284
	$a = .3$	.8171	.7447	.7542	.7832
	$a = .5$	.4757	.5248	.3493	.4709

Chapter 5

Optimal Two-Stage Sequential Design for Case-Control Genetic Association Studies

5.1 Introduction

It has been shown in the previous two chapters that group sequential designs using MAX are feasible and practical for genetic studies with the family-association design and the linkage design. The methods for obtaining the critical values can be extended to population-based association studies as long as the null correlations among the test statistics are provided. Since the methods were based on controlling upper bound of the type I error, they guarantee preservation of type I error in all situations. However, the other important aspect of a study in a statistical sense concerns the assessment of sample size and demonstration of statistical power, which have not been explored. Therefore, in this chapter, we assess the operating characteristics of the MAX statistic used in a two-stage group sequential design in population-based case-control studies.

5.2 Background

As described earlier in Chapter 3, for a diallelic marker with alleles a and A, three genotypes are denoted by aa, aA and AA. The Cochran-Armitage trend tests (CATTs) are often used to detect association between the diallelic marker and the disease. The robust statistic MAX, which takes the maximum of the three CATTs corresponding to the recessive, additive and dominant models, is desired.

Due to their relative simplicity along with the need of improving efficiency, two-stage optimal sequential designs have been studied in many research areas. Simon [1989] proposed a two-stage optimal design that minimized the expected sample size for a clinical trial which determined if the treatment had sufficient biological activity to warrant more extensive development. Shu *et al.* [2007] studied the optimal designs for sequential evaluation of a medical diagnostic test. Although two-stage designs have been studied in case-control genetic studies, they are not typical two-stage sequential studies. For example, Satagopan and Elston [2003] proposed an optimized two-stage design for case-control studies by genotyping all markers (hypotheses) in a portion of samples at the first stage, and then the most promising markers were selected and genotyped using the remaining samples at the second stage. Their approach has been extended further in different situations with different optimality criteria [Satagopan and Elston 2003, Satagopan *et al.* 2004, Muller *et al.* 2007]. Either an allelic test or a single CATT was used in these studies. For discussion on multi-stage designs for genetic case-control studies, see Elston *et al.* [2008]. We consider a typical sequential design for a single marker using the robust statistic MAX, and study how to allocate samples (or information) to achieve minimum average sample size or maximum power.

In the following sections, we investigate the operating characteristics of the two-stage sequential design under a variety of parameters. For example, we take into account not only the design specific parameters, such as the allocation fraction of samples for the first stage, an alpha spending function and alternative differences, but

Table 5.1: Population-based case-control study design.

	aa	Aa	AA	Total
Cases	r_0	r_1	r_2	r
Controls	s_0	s_1	s_2	s
Total	n_0	n_1	n_2	n

also genetic-related parameters including the allele frequency and underlying genetic model. We present the power and the expected sample size under each scenario and make recommendations for the optimal two-stage sequential design for case-control studies. Finally, we apply the design using parameter estimates from several published SNP studies and confirm the savings of sample sizes compared to the original designs.

5.3 Methods

5.3.1 Cochran-Armitage trend tests and MAX

Table 5.1 displays genotype counts for a single marker in cases and controls. Without loss of generality, we assume A is the risk allele and also a minor allele with frequency (MAF) < 0.5. The CATTs for association in Table 5.1 are given by Sasieni [1997]:

$$Z_\theta = \frac{n^{\frac{1}{2}} \sum_{i=0}^{2} x_i (s r_i - r s_i)}{\left\{ rs[n \sum_{i=0}^{2} x_i^2 n_i - (\sum_{i=0}^{2} x_i n_i)^2] \right\}^{1/2}},$$

where $x = (x_0, x_1, x_2) = (0, \theta, 1)$ is the set of scores for the genotypes (aa, Aa, AA). Z_θ is asymptotically normally distributed under the null. The optimal sets of scores for the recessive, additive (multiplicative), and dominant models are $x = (0, 1, 1)$, $(0, \frac{1}{2}, 1)$ and $(0, 0, 1)$, respectively. For a given θ, the CATT in (6.4) follows asymptotically $N(0, 1)$ under H_0.

Let p_i and q_i be the genotype frequencies in the cases and controls, respectively. Then $(r_0, r_1, r_2) \sim mboxMultinomial(r; p_0, p_1, p_2)$ and $(s_0, s_1, s_2) \sim \text{Multinomial}(s; q_0, q_1, q_2)$. The null hypothesis is $H_0 : p_i = q_i, i = 0, 1, 2$. Denote genotype relative risks (GRRs) as $\gamma_i = f_i/f_0$ for cases and $\delta_i = (1-f_i)/(1-f_0)$ for controls, $i = 1, 2$, where the penetrance f_i is the probability of disease given the genotype with i copies of the risk allele. Then $p_i = f_i g_i/K = \gamma_i g_i / \sum \gamma_i g_i$, and $q_i = (1-f_i)g_i/(1-K) = \delta_i g_i / \sum \delta_i g_i$, where g_i, $i = 0, 1, 2$, are the genotype frequencies for the three genotypes $G_0 = aa$, $G_1 = aA$ and $G_2 = AA$, and K is the disease prevalence. The relationship between γ_1 and γ_2 under the dominant, additive and recessive models follows $\gamma_1 = \gamma_2$, $\gamma_1 = (1+\gamma_2)/2$, and $\gamma_1 = 1$, respectively.

The MAX statistic is given by $\text{MAX} = \max(|Z_0|, |Z_{1/2}|, |Z_1|)$ or $\text{MAX} = \max(Z_0, Z_{1/2}, Z_1)$, depending on whether the risk allele is unknown or is the minor allele. Nevertheless, when the risk allele is unknown, the one-sided MAX can still be used at the $\alpha/2$ level with each allele being treated as the risk allele. The asymptotic null correlations among the three CATTs are used in applying sequential analysis using MAX. These correlations were given in Freidlin *et al.* [2002]:

$$\begin{aligned}
\text{Corr}_{H_0}(Z_1, Z_{1/2}) &= \frac{p_2(p_1+2p_0)}{\{p_2(1-p_2)\}^{1/2}\{(p_1+2p_2)p_0+(p_1+2p_0)p_2\}^{1/2}}, \\
\text{Corr}_{H_0}(Z_{1/2}, Z_0) &= \frac{p_0(p_1+2p_2)}{\{(p_0(1-p_0)\}^{1/2}\{(p_1+2p_2)p_0+(p_1+2p_0)p_2)\}^{1/2}}, \\
\text{Corr}_{H_0}(Z_1, Z_0) &= \frac{p_0 p_2}{\{p_0(1-p_0)\}^{1/2}\{p_2(1-p_2)\}^{1/2}}.
\end{aligned}$$

5.3.2 Applying MAX to sequential design

Critical values for a group sequential design are determined by the distribution of the test statistic and a prespecified alpha spending function. As discussed earlier, conventional alpha spending methods, such as the Pocock method [Pocock 1977] and the O'Brien-Fleming method [O'Brien and Fleming 1979], can be readily applied to obtain the critical values for a single test statistic with a known form of distribution.

However, the test statistic MAX does not have an explicit form of distribution. we demonstrated earlier that the two-point formula (3.5) is simple to use and control type I errors reasonably well. We consider MAX = $\max(Z_0, Z_{1/2}, Z_1)$ with a target $\alpha/2$ level. The two-point formula can be written as:

$$P(\text{MAX} > c) \leq \bar{\Phi}(c) + \sum_{j=1/2,1} \left\{ \int_{-\infty}^{c} \bar{\Phi} \left[\frac{c - \rho_j t}{(1-\rho_j^2)^{1/2}} \right] \phi(t) dt \right\},$$

where ρ_j is the asymptotic null correlation between Z_j and $Z_{j-1/2}$, ϕ and Φ are the density and distribution functions of $N(0,1)$, respectively, and $\bar{\Phi} = 1 - \Phi$.

Consider a two-stage sequential design with overall level α. Assume the sample size in stage i is N_i with levels α_i, $i = 1, 2$, and $\alpha_1 + \alpha_2 = \alpha$. The sample allocation is denoted by $\pi = N_1/N_2$ (the samples in stage 2 include all samples in stage 1). Given α and π, the levels α_1 and α_2 are determined by a prespecified alpha spending function (ASP). Three commonly used ASFs are considered here (also see Chapters 3 and 4): i) $\text{ASP}_1(t) = 2\{1 - \Phi(z_{\alpha/2} t^{-1/2})\}$, ii) $\text{ASP}_2(t) = \alpha \log\{1 + (e-1)t\}$, and iii) $\text{ASP}_3(t) = \alpha t$, where $t = \pi$ is the information fraction. The first and second functions are equivalent to the discretized O'Brien and Fleming and Pocock types of spending functions, respectively. The third function is an uniform spending function of information. The corresponding critical values c_1 and c_2 for the two stages using MAX can be inversely computed using the Newton-Raphson method.

5.3.3 Optimal two-stage design for case-control genetic association studies

In a two-stage sequential design, there is a trade-off between minimizing the sample size and maximizing the statistical power. Both are important when designing a group sequential study. Decisions are often made to balance the trade-off between the two factors. Suppose one interim analysis is conducted at Stage 1 and one final analysis is done at Stage 2. Define the stopping rule as that the study stops if MAX is statistically significant at Stage 1. Denote the allocation fraction $\pi = N_1/N_2$ as

before, where N_1 and N_2 are the cumulative sample sizes at the two stages. The expected (average) sample size (ASN) for the two-stage sequential design can be calculated as:

$$\text{ASN} = N_1 \text{ P(Study stops at Stage 1)} + N_2 \text{ P(Study continues), under } H_1 \text{or } H_0.$$

Accordingly, the type I error and power can be respectively written as:

$$\begin{aligned} \alpha &= P(\text{MAX}_1 > c_1 | H_0) + P(\text{MAX}_1 \leq c_1 \text{ and } \text{MAX}_2 > c_2 | H_0), \\ \text{Power} &= P(\text{MAX}_1 > c_1 | H_1) + P(\text{MAX}_1 \leq c_1 \text{ and } \text{MAX}_2 > c_2 | H_1), \end{aligned}$$

where MAX_1 and MAX_2 are the test statistics, and c_1 and c_2 are the critical values.

The goals of our optimal designs are to minimize the ASN or to achieve the maximum power while the type I error is controlled, and find the ranges of parameter values to achieve the minimum ASN or maximum power. Note that a parameter considered in the design is a function of other parameters in the design. Using the notation in Section 2.1, given values of the disease prevalence K and the MAF (and hence the genotype frequency g_i, $i = 0, 1, 2$ under HWE), the power of MAX for a given sample size is a function of the GRRs and the underlying genetic model. In a two-stage design, the ASN is determined by the power of detecting the difference at each stage, which is determined not only by the above parameters but also the critical values, which are further determined by the alpha spending function and the allocation fraction. Here we study the optimal designs such that either the minimum ASN or the maximum power are achieved. First, we present the optimal designs under a fixed target sample size and study the impact on ASN and power for different combinations of the specified values of the allele frequency, genetic model (dominant, additive, or recessive), the alternative genotype relative risks (γ_1, γ_2), and the alpha spending function (ASF). Then, from a different perspective, we provide recommended sample sizes for optimal designs under the specified values of those parameters when the power is fixed. The results are also compared with those for a single-stage design.

5.4 Simulation studies and examples

In this section, we will first present results from simulation studies with various assumptions of values for the parameters. Then we will further evaluate the findings in the simulation studies by applying to real examples with more realistic values for some of the parameters in the simulation.

The simulation studies were carried out based on the disease prevalence $K = 0.1$. Each simulation was replicated 10,000 times. The critical values for MAX were obtained given the values of the allocation fraction π from 0.1 to 0.5 with an increment of 0.05 and an ASF described earlier. When the target sample size is fixed, the simulation was generated, and the ASN and power were calculated. When the power is fixed, the simulation was repeated until the sample size that achieved the specified power was found. The ASN and power were then calculated based on the simulated data sets. For all simulations, Hardy-Weinberg Equilibrium was assumed and all tests were two-sided at an overall alpha level of 0.05.

5.4.1 Fix the sample size

Given the target sample size of 1,000 cases and 1,000 controls for a single-stage design, we simulated results to obtain the ASNs and the powers under the alternative hypotheses for each of the three ASFs. The results are reported in Tables 5.2 through 5.4 for three different ASFs. In each table, results are presented under different MAFs p=0.1, 0.3, or 0.5, a genetic model, dominant (DOM), additive (ADD), or recessive (REC), and GRRs (γ_1, γ_2). The ASNs and powers and the corresponding allocation fractions π are presented for two scenarios: 1) when the minimum ASN is achieved (Columns 5 through 7); and 2) when the maximum power is achieved (Columns 8 through 10). The last column presents the power of a single-stage study with 1,000 cases and 1,000 controls.

While the optimal designs that achieve the minimum ASNs suggest the allocation

fractions of between 0.3 and 0.5 for Stage 1 depending on the ASFs and powers, the allocation fractions for the designs where the maximum powers are achieved are lower, ranging from 0.1 to 0.3. The gain of the power is relatively small for having extra samples, especially when the study power is high. For example, in Table 5.3, under the additive model ($\gamma_1 = 1.25$ and $\gamma_2 = 1.5$) where p=0.3, the minimum ASN is 722 subjects per group when the allocation fraction is 0.5 with the corresponding power of 85%. The ASN rises to 958 subjects per group when the allocation fraction is 0.1, the power is maximized and increased to 89%, only 4% more than the power achieved earlier. The power for a single-stage study with a sample size of 1,000 per group is 90%.

For a given ASF, higher MAF results in lower ASN and higher power. For example, for ASF = ASF_2, under the additive model ($\gamma_1 = 1.25$ and $\gamma_2 = 1.5$), the minimum ASN when p=0.1 is 869 with a power of 50%, whereas the correponding minimum ASNs and powers are 722 and 85% when $p = 0.3$, and 711 and 87% when p=0.5.

Among the three genetic models, when the GRR for genotype AA (γ_2) is fixed, the test has the highest power under the dominant model and has the lowest power under the recessive model. Focusing on the same section of results where ASF = ASF_2, $p = 0.3$ and $\gamma_2 = 1.5$ in Table 5.3, the minimum ASN is 547 with a power of 99% under the dominant model, whereas it is 722 with 85% power under the additive model, and 783 with 75% power under the recessive model.

Finally, across different ASFs, the allocation fractions are consistently suggested to be around 0.5 to achieve the minimum ASN when the alpha spending is conservative at the first stage, e.g. ASF = ASF_1 or the O'Brien-Fleming type. When the allowance for type I error increases, the allocation fractions that achieve the minimum ASN go slightly lower (ranging from 0.25 to 0.5). The Pocock type of ASF (ASF_2) appears to result in the smallest ASNs with the powers similar to those in the corresponding setting for the other ASFs.

5.4.2 Fix the statistical power

For two-stage sequential designs, when the power is given, say 80%, the required sample size to achieve this given power under a certain alternative can be obtained along with the allocation fraction. Under similar simulation procedures, we present in Tables 5.5 through 5.7 the planned sample sizes (N) as well as the ASNs that achieve the power 80% under different scenarios. Columns 5 through 8 show results when the ASN is the minimum, whereas Columns 9 through 12 show the results when the ASN is the maximum in comparison.

When the allocation ratio is set around half ($\pi = 0.5$), the probability of stopping at Stage 1 is the highest and the ASN reaches the minimum. On the other hand, when the allocation ratio is set low (between 0.1 and 0.15), the required sample size is relatively smaller because while the probability of stopping at the first stage is small, the chance of mistakenly stopping (type I error) is small as well. The ASNs, however, are about 7% higher than those in the case where the allocation fraction is 0.5.

The results also confirm the conservativeness between the three ASFs. When the O'Brien-Fleming type of spending (ASF = ASF_1) is utilized, the required sample sizes are the smallest but the ASNs are maintained to be close to the targeted sample size because the probability of stopping at Stage 1 is smaller than in those using the other ASFs.

In summary, for a two-stage sequential study, although the target sample size is generally required to be larger than what it is for a single-stage study, the ASN from a sequential study is smaller than that in a single study. Figure 5.2 also shows a visual example of such benefit with ASF = ASF_2. The patterns are similar for the other ASFs (result are not shown) even though the extent of the benefit might vary.

5.4.3 Examples

In this section, we apply the two-stage sequential design using MAX to four existing genome-wide association studies (GWAS): the age-related macular degeneration (AMD) study [Klein *et al.* 2005], the prostate cancer study[Yeager *et al.* 2007], the breast cancer study [Hunter *et al.* 2007], and the hypertension study [WTCCC 2007]. Based on the observed genotype distributions in cases and controls for a total of 17 SNPs in these studies (shown in Table 5.8) cited in Li *et al.* [2008b], we studied the ASNs and powers if these studies had used a two-stage design. For GWAS, we used a significance level of 0.0001, which was used by Sladek *et al.* [2007] for initial single-marker scans. Based on results from the previous section, we fix the allocation fraction at 0.5. The MAF is estimated using the genotype counts of each SNP. Table 5.9 shows the ASNs and corresponding powers for each of the three ASFs from using half of the study sample size (cases+controls) for the first stage and allowing stopping if significance was found at the desired alpha level. As a comparison, the corresponding power for a single-stage study is also presented. As expected, the sample saving is minimal when alpha spending is conservative for Stage 1 (ASF_1). Using more aggressive alpha spending, the saving advantages become more obvious. For example, in the AMD study and when the Pocock type of alpha spending (ASF_2) was used, for both of the SNPs presented in the table, sample size saving would be 14% of that original planned with about 5% reduction in power. In another example, using a uniform alpha spending function (ASF_3), the results generally suggest that we can save over 10% of the samples if the study has been conducted using a two-stage design with about 5% sacrifice in power.

5.5 Discussion

In this chapter we studied the operating characteristics of a group sequential case-control genetic study of the association between a single candidate marker and a

disease using the robust statistic MAX. The results have shown advantages of having a two-stage design on savings of the average sample size while maintaining the power in a slightly reduced rate. Our results indicated that the typical allocation fraction of half often balances the trade-off between the sample sizes and powers. The choice of alpha spending function can impact the sample size. Overly conservative alpha spending, such as the O'Brien-Fleming type of spending, helps little in sample size saving in a two-stage design. On the other hand, when the underlying difference is large, e.g., when the genotype relative risk for Genotype AA (γ_2) is above 2.0, or a slightly more spending during Stage 1 is allocated, the optimal design can be achieved with the interim analysis being performed for fewer subset of the target samples, i.e. allocation fraction being around 0.3 or so and results in a more than 50% reduction in average sample size compared to a single stage study.

The idea of having a two-stage sequential design under strong alternatives that allows stopping for significance in order to reduce average sample size can be extended to the opposite direction of the stopping rule, i.e., the study stops for futility or no association. Our simulation results also infer savings on sample size for such a design although careful design of the allocation of type II error (β) needs to be considered. Furthermore, the study can be designed to allow stopping for both significance and non-significance, where the study may reach the highest efficiency. Our preliminary results of the impact factors on sample sizes can be a starting point for further research for optimal choices in the genetic study setting. Our software can also be used as a practical tool in calculating critical values, sample size and power to design a two-stage sequential study using case-control samples.

Note that the calculation of critical values for MAX is an approximation. Although the approximation method has been proven through simulation to work well for a two-stage study, it can be overly conservative if more stages are planned and therefore affects the choice of required sample sizes. Further investigation may be needed to compare different alpha spending functions in terms of their impact on the designs

of the study given the distribution of MAX under different alternatives with different values of the allele frequency and specification of the genetic model.

Table 5.2: Optimal allocation fractions π to achieve the minimum ASN or the maximum power given the target sample size for a single-stage design ($N = 1000$ per group): ASF = ASF_1. The alternatives are specified by the GRRs (γ_1, γ_2) under three genetic models with different MAFs p.

				Min ASN is Achieved			Max Power is achieved			Power for a single
p	Model	γ_1	γ_2	ASN	Power	π	Power	ASN	π	stage study
0.1	DOM	1.5	1.5	786	0.978	0.5	0.978	1000	0.15	0.975
		2	2	468	1.000	0.4	1.000	468	0.4	1.000
	ADD	1.25	1.5	964	0.598	0.5	0.608	1000	0.25	0.601
		1.5	2	745	0.991	0.5	0.991	916	0.35	0.989
	REC	1	1.5	998	0.146	0.5	0.156	1000	0.15	0.151
		1	2	988	0.430	0.5	0.454	1000	0.15	0.451
0.3	DOM	1.5	1.5	678	0.997	0.5	0.998	723	0.45	0.997
		2	2	426	1.000	0.4	1.000	426	0.4	1.000
	ADD	1.25	1.5	878	0.895	0.5	0.911	973	0.35	0.902
		1.5	2	561	1.000	0.5	1.000	561	0.5	1.000
	REC	1	1.5	918	0.827	0.5	0.831	985	0.35	0.823
		1	2	600	1.000	0.5	1.000	600	0.5	1.000
0.5	DOM	1.5	1.5	802	0.968	0.5	0.973	1000	0.2	0.972
		2	2	528	1.000	0.45	1.000	528	0.45	1.000
	ADD	1.25	1.5	863	0.919	0.5	0.919	968	0.35	0.915
		1.5	2	573	1.000	0.5	1.000	586	0.45	1.000
	REC	1	1.5	736	0.993	0.5	0.993	736	0.5	0.990
		1	2	439	1.000	0.4	1.000	439	0.4	1.000

Table 5.3: Optimal allocation fractions π to achieve the minimum ASN or the maximum power given the target sample size for a single-stage design ($N = 1000$ per group): ASF = ASF_2. The alternatives are specified by the GRRs (γ_1, γ_2) under three genetic models with different MAFs p.

				Min ASN is Achieved			Max Power is achieved			Power for a single
p	Model	γ_1	γ_2	ASN	Power	π	Power	ASN	π	stage study
0.1	DOM	1.5	1.5	635	0.962	0.45	0.973	943	0.10	0.977
		2	2	340	1.000	0.30	1.000	340	0.30	1.000
	ADD	1.25	1.5	869	0.501	0.50	0.581	985	0.10	0.605
		1.5	2	597	0.981	0.40	0.987	933	0.10	0.988
	REC	1	1.5	968	0.119	0.45	0.137	997	0.10	0.153
		1	2	914	0.351	0.50	0.425	995	0.10	0.452
0.3	DOM	1.5	1.5	547	0.994	0.45	0.998	727	0.20	0.998
		2	2	291	1.000	0.25	1.000	291	0.25	1.000
	ADD	1.25	1.5	722	0.852	0.50	0.891	958	0.10	0.903
		1.5	2	431	1.000	0.35	1.000	440	0.30	1.000
	REC	1	1.5	783	0.745	0.50	0.810	970	0.10	0.827
		1	2	469	0.999	0.35	1.000	755	0.15	1.000
0.5	DOM	1.5	1.5	654	0.945	0.50	0.966	948	0.10	0.970
		2	2	397	1.000	0.30	1.000	397	0.30	1.000
	ADD	1.25	1.5	711	0.870	0.50	0.907	955	0.10	0.918
		1.5	2	441	1.000	0.35	1.000	480	0.45	1.000
	REC	1	1.5	592	0.982	0.45	0.989	926	0.10	0.990
		1	2	306	1.000	0.25	1.000	306	0.25	1.000

Table 5.4: Optimal allocation fractions π to achieve the minimum ASN or the maximum power given the target sample size for a single-stage design ($N = 1000$ per group): ASF = ASF_3. The alternatives are specified by the GRRs (γ_1, γ_2) under three genetic models with different MAFs p.

				Min ASN is Achieved			Max Power is achieved			Power for a single
p	Model	γ_1	γ_2	ASN	Power	π	Power	ASN	π	stage study
0.1	DOM	1.5	1.5	652	0.964	0.45	0.977	960	0.10	0.979
		2	2	354	1.000	0.30	1.000	354	0.30	1.000
	ADD	1.25	1.5	882	0.526	0.50	0.584	979	0.15	0.602
		1.5	2	613	0.983	0.45	0.988	831	0.20	0.988
	REC	1	1.5	977	0.115	0.50	0.143	996	0.15	0.154
		1	2	928	0.355	0.50	0.422	997	0.10	0.450
0.3	DOM	1.5	1.5	566	0.996	0.45	0.997	928	0.10	0.998
		2	2	300	1.000	0.25	1.000	300	0.25	1.000
	ADD	1.25	1.5	743	0.863	0.50	0.898	969	0.10	0.905
		1.5	2	444	1.000	0.35	1.000	444	0.35	1.000
	REC	1	1.5	794	0.778	0.50	0.822	982	0.10	0.832
		1	2	487	1.000	0.40	1.000	799	0.15	1.000
0.5	DOM	1.5	1.5	661	0.953	0.50	0.967	959	0.10	0.971
		2	2	411	1.000	0.35	1.000	411	0.35	1.000
	ADD	1.25	1.5	729	0.876	0.50	0.914	936	0.15	0.918
		1.5	2	460	1.000	0.35	1.000	460	0.35	1.000
	REC	1	1.5	609	0.986	0.50	0.990	744	0.25	0.991
		1	2	321	1.000	0.25	1.000	321	0.25	1.000

Table 5.5: Sample size and optimal allocation fraction to achieve the power 80%: ASF = ASF_1.

				Min ASN is achieved				Max ASN is resulted			
p	Model	γ_1	γ_2	π	N	ASN	P(stopping at Stage 1)	π	N	ASN	P(stopping at Stage 1)
0.1	DOM	1.5	1.5	0.5	523	481	0.158	0.3	523	521	0.006
		2	2	0.5	165	152	0.154	0.1	164	164	0.000
	ADD	1.25	1.5	0.5	1572	1446	0.161	0.2	1547	1547	0.000
		1.5	2	0.5	451	417	0.153	0.25	450	449	0.000
	REC	1	1.5	0.5	7659	7074	0.153	0.25	7581	7573	0.002
		1	2	0.5	2186	2039	0.135	0.3	2190	2187	0.002
0.3	DOM	1.5	1.5	0.5	377	348	0.153	0.1	368	368	0.000
		2	2	0.5	133	122	0.155	0.1	132	132	0.000
	ADD	1.25	1.5	0.5	747	690	0.153	0.25	745	744	0.002
		1.5	2	0.5	237	217	0.166	0.2	240	240	0.000
	REC	1	1.5	0.5	947	872	0.159	0.25	947	947	0.001
		1	2	0.5	284	263	0.151	0.15	289	289	0.000
0.5	DOM	1.5	1.5	0.5	552	510	0.152	0.25	557	556	0.001
		2	2	0.5	210	194	0.150	0.15	209	209	0.000
	ADD	1.25	1.5	0.5	714	656	0.163	0.15	708	708	0.000
		1.5	2	0.5	250	230	0.154	0.2	253	253	0.000
	REC	1	1.5	0.5	445	410	0.157	0.1	440	440	0.000
		1	2	0.5	142	131	0.153	0.1	143	143	0.000

Table 5.6: Sample size and optimal allocation fraction to achieve the power 80%: ASF = ASF_2.

				Min ASN is achieved				Max ASN is resulted			
p	Model	γ_1	γ_2	π	N	ASN	P(stopping at Stage 1)	π	N	ASN	P(stopping at Stage 1)
0.1	DOM	1.5	1.5	0.4	596	469	0.354	0.1	541	527	0.027
		2	2	0.45	194	148	0.428	0.1	169	166	0.023
	ADD	1.25	1.5	0.5	1845	1385	0.499	0.1	1615	1574	0.028
		1.5	2	0.5	540	410	0.483	0.1	465	454	0.027
	REC	1	1.5	0.5	8830	6712	0.480	0.1	7903	7718	0.026
		1	2	0.45	2566	1967	0.425	0.1	2254	2229	0.013
0.3	DOM	1.5	1.5	0.5	440	330	0.500	0.15	401	377	0.071
		2	2	0.5	157	118	0.488	0.1	139	135	0.026
	ADD	1.25	1.5	0.45	869	659	0.439	0.1	781	756	0.035
		1.5	2	0.5	287	215	0.505	0.1	249	243	0.030
	REC	1	1.5	0.5	1123	844	0.497	0.1	986	959	0.030
		1	2	0.5	338	256	0.482	0.15	303	288	0.058
0.5	DOM	1.5	1.5	0.5	654	495	0.489	0.1	577	559	0.033
		2	2	0.4	235	185	0.354	0.1	216	211	0.023
	ADD	1.25	1.5	0.45	827	630	0.435	0.1	740	719	0.032
		1.5	2	0.5	296	221	0.508	0.1	258	251	0.030
	REC	1	1.5	0.45	518	393	0.437	0.1	459	447	0.030
		1	2	0.5	172	129	0.503	0.1	149	145	0.027

Table 5.7: Sample size and optimal allocation fraction to achieve the power 80%: ASF = ASF_3.

				Min ASN is achieved				Max ASN is resulted			
p	Model	γ_1	γ_2	π	N	ASN	P(stopping at Stage 1)	π	N	ASN	P(stopping at Stage 1)
0.1	DOM	1.5	1.5	0.5	586	460	0.430	0.1	525	516	0.020
		2	2	0.45	184	146	0.370	0.15	173	168	0.035
	ADD	1.25	1.5	0.45	1737	1382	0.371	0.1	1601	1571	0.021
		1.5	2	0.5	508	400	0.425	0.15	467	451	0.039
	REC	1	1.5	0.5	8512	6676	0.431	0.1	7698	7569	0.019
		1	2	0.45	2471	1986	0.357	0.1	2235	2219	0.008
0.3	DOM	1.5	1.5	0.5	422	328	0.444	0.1	381	374	0.020
		2	2	0.5	150	117	0.439	0.15	138	132	0.045
	ADD	1.25	1.5	0.5	850	661	0.445	0.1	752	736	0.024
		1.5	2	0.5	272	211	0.449	0.1	249	244	0.021
	REC	1	1.5	0.45	1045	832	0.371	0.1	962	946	0.019
		1	2	0.5	322	254	0.425	0.1	291	288	0.013
0.5	DOM	1.5	1.5	0.45	618	492	0.373	0.1	565	554	0.023
		2	2	0.45	235	186	0.379	0.1	212	208	0.020
	ADD	1.25	1.5	0.5	810	627	0.454	0.1	724	706	0.027
		1.5	2	0.45	279	220	0.387	0.1	255	250	0.019
	REC	1	1.5	0.5	496	387	0.438	0.1	450	440	0.023
		1	2	0.5	162	126	0.439	0.1	146	143	0.019

Figure 5.1: ASNs and sample sizes for one-stage and two-stage studies at 80% Power: γ_2=1.5: ASF = ASF_1.

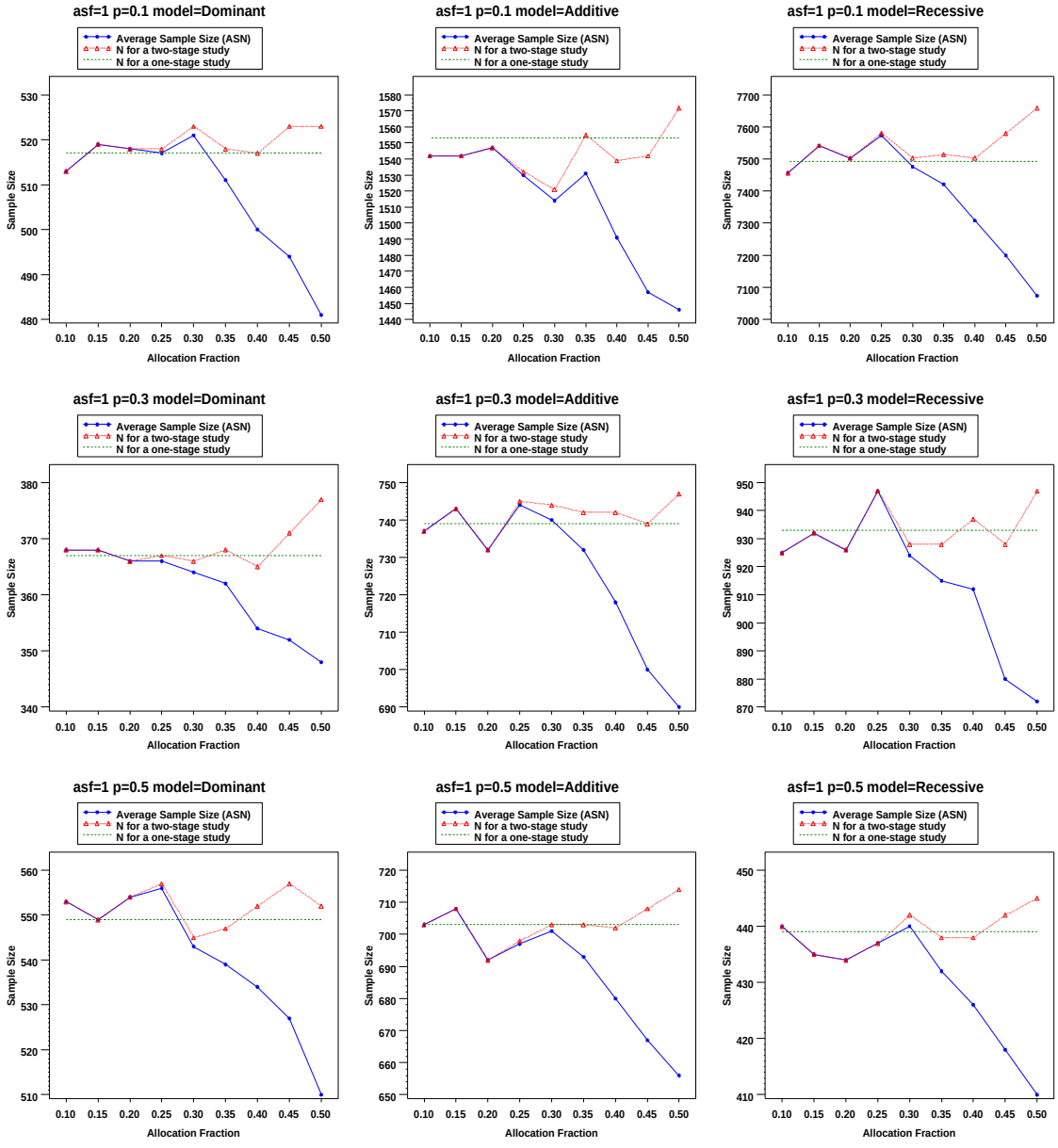


Figure 5.2: ASNs and sample sizes for single-stage and two-stage studies at 80% Power: γ_2=1.5, ASF = ASF_2.

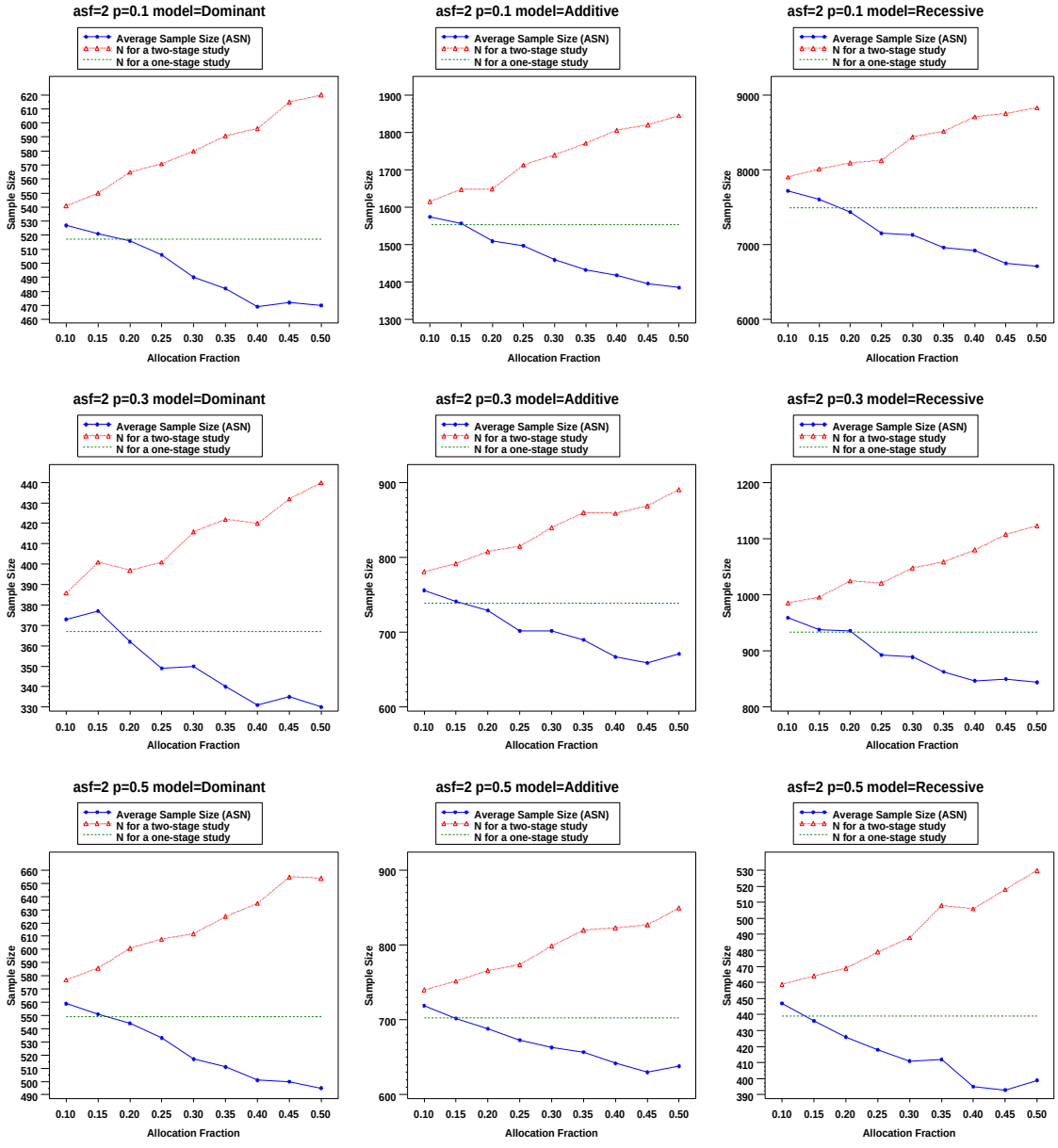


Figure 5.3: ASNs and sample sizes for single-stage and two-stage studies at 80% Power: γ_2=1.5, ASF = ASF_3.

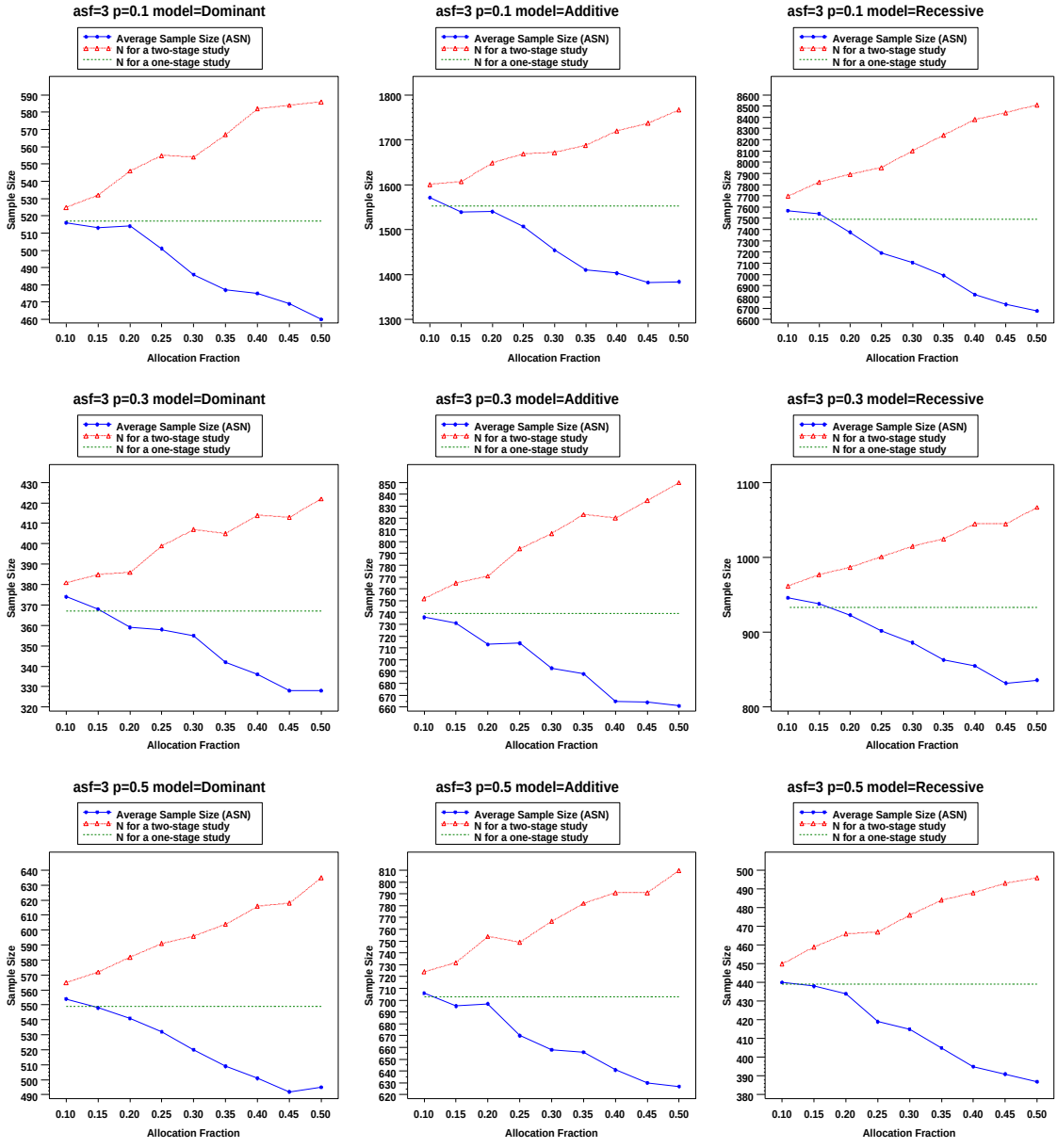


Table 5.8: Description of genotype distributions in 17 SNPs from 4 GWAS studies.

	Cases			Controls				
SNP ID	AA	AB	BB	AA	AB	BB	Total	MAF
AMD								
rs38390	50	35	11	6	25	19	146	0.59
rs1329428	2	24	68	5	29	14	142	0.24
Prostate Cancer								
rs1447295	25	283	864	10	218	929	2329	0.12
rs698267	223	598	351	301	579	277	2329	0.48
rs7837688	27	283	861	11	206	939	2327	0.12
Breast Cancer								
rs10510126	10	180	955	14	272	854	2285	0.11
rs12505080	50	477	608	99	408	628	2270	0.26
rs17157903	18	316	777	26	220	862	2219	0.14
rs1219648	250	543	352	170	538	433	2286	0.42
rs7696175	187	605	353	249	496	396	2286	0.43
rs2420946	242	546	357	165	537	440	2287	0.41
Hypertension								
rs2820037	40	587	1325	72	684	2180	4888	0.15
rs6997709	118	716	1116	237	1201	1500	4888	0.27
rs7961152	416	963	570	492	1448	992	4881	0.43
rs11110912	67	647	1237	83	804	2049	4887	0.18
rs1937506	113	742	1097	244	1205	1484	4885	0.27
rs2398162	111	624	1205	194	1121	1608	4863	0.24

Table 5.9: ASNs and powers for a two-stage design for the 17 SNPs

	ASF_1			ASF_2			ASF_3		
SNP ID	ASN	Power two-Stage	Power one-Stage	ASN	Power two-Stage	Power one-Stage	ASN	Power two-Stage	Power one-Stage
AMD									
rs38390	146	0.902	0.902	126	0.857	0.906	128	0.871	0.902
rs1329428	142	0.859	0.856	123	0.801	0.856	124	0.819	0.854
Prostate Cancer									
rs1447295	2327	0.496	0.496	2219	0.419	0.501	2224	0.450	0.506
rs698267	2325	0.688	0.688	2138	0.592	0.671	2153	0.614	0.674
rs7837688	2319	0.737	0.737	2108	0.667	0.735	2111	0.689	0.739
Breast Cancer									
rs10510126	2268	0.831	0.831	1986	0.768	0.824	2007	0.790	0.831
rs12505080	2269	0.515	0.515	2174	0.426	0.519	2179	0.463	0.528
rs17157903	2217	0.542	0.542	2103	0.471	0.552	2107	0.490	0.547
rs1219648	2279	0.778	0.778	2032	0.708	0.772	2043	0.735	0.785
rs7696175	2286	0.227	0.227	2249	0.176	0.237	2252	0.188	0.233
rs2420946	2280	0.768	0.768	2033	0.697	0.765	2048	0.727	0.775
Hypertension									
rs2820037	4864	0.766	0.766	4371	0.710	0.776	4401	0.716	0.762
rs6997709	4879	0.671	0.671	4498	0.591	0.670	4528	0.623	0.674
rs7961152	4869	0.676	0.676	4466	0.609	0.684	4517	0.627	0.680
rs11110912	4868	0.722	0.722	4421	0.650	0.727	4439	0.681	0.730
rs1937506	4875	0.646	0.646	4529	0.569	0.651	4550	0.587	0.648
rs2398162	4838	0.800	0.800	4284	0.728	0.793	4329	0.755	0.800

Chapter 6

Two-Stage Adaptive Design for Case-Control Genetic Association Studies

6.1 Introduction

Adaptive designs have been increasingly considered during the conduct of clinical trials in recent years. While the "classical" design is the most preferable in planning a study, it requires knowledge or good estimates of parameters for determining the study characteristics such as sample size, power, test statistic and study endpoint. However, these study parameters are often unknown, mis-estimations affect the success of the study in terms of its targeted statistical power or ability to draw a plausible conclusion. A two-stage adaptive design is often used in the situation to "rescue" the study from potential loss from various perspectives of the study such as time, cost and statistical power. It allows an assessment of the study parameter(s) during the interim of the study and then making an appropriate adjustment to the study characteristics to obtain the maximum of the information in the study.

The elements to be modified in adaptive designs in a statistical sense typically include sample size, study endpoint/test hypothesis, randomization plan if in a clinical

trial, and test statistics. Statistical approaches that combine the p-values obtained from different stages proposed by researchers from decades ago and are revisited and applied for modern adaptive designs. Bauer and Köhne [1994] compared a few methods and demonstrated that using Fisher's combination test for independent test statistics worked well in the situation of adaptive design. Hommel *et al.* [2005] extended this approach and proposed a combination test for the scenario where the test statistics are correlated.

Not as popular as adaptive designs for sample size modifications, adaptive designs that allow early stopping and modification of test statistics have been studied by a few researchers. For example, Lang *et al.* [2000] proposed using an adaptive statistic based on estimation through an isotonic regression of the interim data to test equality of several normal means against a monotone alternative. Kieser *et al.* [2000] proposed to adapt an "optimal" statistic through bootstrapping approaches of several statistics. Both studies demonstrated power advantages over the classical designs. For genetic studies, while the two-stage non-sequential designs have been widely considered by researchers for the purpose of cost and time saving in genome wide association studies, the idea of adaptive design has been considered in a few researchers. For example, Song *et al.* [2007] proposed an adaptive two-stage procedure that screens single-nucleotide polymorphisms (SNPs) using the Hardy-Weinberg disequilibrium trend test (HWDTT) in a first stage, and then tests a reduced number of SNPs that pass the screening step in a second stage using the Cochran-Armitage trend test (CATT). Zheng and Ng [2008] presented a two-phase test combining the HWDTT test and consequently selected genetic model specific CATT test based on the results of the HWDTT test. None of these studies were performed following the formal traditional group sequential design.

In this chapter, as the uncertainty of the underlying genetic model continues to be a challenge in the statistical analysis, we explore adaptive group sequential designs in population-based case-control association studies. We propose two model

selection procedures to select "optimal" statistics based on interim data and apply these statistic to the rest part of the data. One method is based on the maximum CATT test statistic, the other is based on the HWDTT statistic. We compare the power results for both adaptive and non-adaptive group sequential designs under various specified genetic parameters and alternative hypotheses.

6.2 Methods

6.2.1 Combination tests

This section considers the practically relevant situation of a two-stage study where a single adaptive interim analysis is planned. In the literature, various approaches have been proposed for significance tests.

Let p_1 be the observed error probability or p-value for the test in the subsample investigated before the interim analysis (Stage 1) and p_2 be that for the subsample recruited after adaptation (Stage 2). If p_1 and p_2 are uniformly distributed on [0,1] under H_0, using the Fisher's combination test [Hedges and Olkin 1985], the null hypothesis can be rejected at the end of the study if

$$p_1 p_2 \leq exp[-\frac{1}{2}\chi_4^2(1-\alpha)] \tag{6.1}$$

where $\chi_4^2(1-\alpha)$ is the $(1-\alpha)$-quantile of the central χ^2 distribution with 4 degrees of freedom (d.f.).

For test statistics that are asymptotically normally distributed with known variance and equal sample, the uniform most powerful test method, referred to as the inverse normal method may be used [Bauer and Kohne 1994]. The rejection region is written as:

$$\frac{1}{\sqrt{2}}\Phi^{-1}(1-p_1) + \frac{1}{\sqrt{2}}\Phi^{-1}(1-p_2) \leq z_{1-\alpha}, \tag{6.2}$$

where Φ^{-1} is the inverse of the standard normal distribution function.

By taking into account the sample size imbalance, Lehmacher and Wassmer [1999] incorporated weights in the inverse normal method and proposed the rejection region being as:

$$\sqrt{\frac{n_1}{n_1+n_2}}\Phi^{-1}(1-p_1)+\sqrt{\frac{n_1}{n_1+n_2}}\Phi^{-1}(1-p_2)\leq z_{1-\alpha}, \tag{6.3}$$

where n_1 and n_2 are the samples sizes for Stages 1 and 2, respectively.

Bauer and Kohne [1994] demonstrated that using Fisher's combination tests in adaptive designs provided simplicity and generalizability in adaptive designs and surprisingly suffered small loss of power compared to the other methods. The procedure of using such a test is described as following: When one allows early stopping in a two-stage study, the test decisions can be phrased as: After the first stage, early rejection occurs for $p_1 \leq \alpha_1$ and early acceptance for $p_1 \geq \alpha_0$. For $p \in [\alpha_1, \alpha_0]$, the study is continued. To control the type I error rate of α over the two stages, the early stopping boundaries α_0 and α_1 are to be chosen such that they fulfill the equation

$$\alpha_1+\int_{\alpha_1}^{\alpha_0}\int_0^{c_\alpha/p_1} dp_2dp_1=\alpha_1+c_\alpha(\ln\alpha_0-\ln\alpha_1)=\alpha$$

If $p_1 \leq \alpha_1$ one could already stop at the interim analysis with the rejection of H_0. The condition $p_2 \leq 1$ guarantees that the combination test must reject H_0. regarding the choice of α_1 and α_0 and the critical value c_α for the two p-values obtained from a two-stage study.

6.2.2 Adaptive procedures

As described in Chapter 5, CATTs are well suited for testing association between a marker and a disease in population based case-control studies (Table 5.1). It is equivalent to the score test in a logistic regression model for case-control data with a single covariate (genotype) coded with score $(0, x, 1)$ for genotypes (g_0, g_1, g_2), where

$0 \leq x \leq 1$. Recall the statistic is:

$$Z_\theta = \frac{n^{\frac{1}{2}} \sum_{i=0}^{2} x_i (sr_i - rs_i)}{\left\{ rs[n \sum_{i=0}^{2} x_i^2 n_i - (\sum_{i=0}^{2} x_i n_i)^2] \right\}^{1/2}},$$

where $x = (x_0, x_1, x_2) = (0, x, 1)$ and $x = 1, 1/2, 0$ for the dominant, additive, and recessive models, respectively. For complex diseases, however, the genetic models of the true disease loci are unknown. In this situation, robust tests, such as MAX, are preferable. As the previous chapters presented the execution of conducting statistical testing using the MAX statistic in a group sequential study, the chapter further explores the application of an adaptive design in such a study. We consider two different adaptive test statistics, both motivated by the problem that the underlying genetic model is unknown. The construction of the adaptive statistics targets a higher likelihood of identification or selection of the genetic model based on the data obtained in the first stage, hence a more powerful test is advised in the second stage.

The first test statistic is based on the observed maximum value of the test statistics for the first stage data. The score set or genetic model that results in the maximum test statistic is selected (referred to as the MAX approach hereafter). In another words, the adaptive test statistic $T_{adp1} = T(\theta)$, is constructed by the trend score set $(0, \theta, 1)$ is such that $\theta = arg_{\theta \in (0,1/2,1)}\{T(\theta) = max(T(0), T(1/2), T(1)\}$.

The second adaptive statistic is motivated by the genetic model selection (GMS) approach proposed by Zheng and Ng [2008]. The selection procedure is based on using the HWDTT proposed by Song and Elston [2006]. The HWDTT tests the difference of a Hardy-Weinberg disequilibrium (HWD) coefficient measure [Weir 1996], denoted as $\Delta = Pr(AA) - [Pr(AA) + Pr(Aa)/2]^2$ between the cases (Δ_p) and controls (Δ_q) being zero under the null hypothesis of no association. When an association exists, the direction of the difference ($\Delta_p - \Delta_q$) is associated with the underlying genetic model [Wittke-Thompson *et al.* 2005]. Specifically, the HWDTT test statistic is written as

$$Z_{HWDTT} = \frac{(rs/n)^{1/2}(\hat{\Delta}_p - \hat{\Delta}_q)}{[1 - n_2/n - n_1/(2n)][n_2/n + n_1/(2n)]},$$

where $\hat{\Delta}_p = \hat{p}_2 - (\hat{p}_2 + \hat{p}_1/2)^2$ and $\hat{\Delta}_q = \hat{q}_2 - (\hat{q}_2 + \hat{q}_1/2)^2$ are the estimates of Δ_p and Δ_q, respectively, with $\hat{p}_i = r_i/r$ and $\hat{q}_i = s_i/s$.

The adaptive statistic $T_{adp2} = T_{GMS}$, where $T_{GMS} = T(0)$ if $Z_{HWDTT} > c$, $T(1)$ if $Z_{HWDTT} < -c$, and $T(1/2)$ otherwise. The c is proposed to be $\Phi^{-1}(0.95) = 1.645$. This approach is referred as the HWDTT approach hereafter.

6.3 Simulation studies

Similar to the setups described in Chapter 5, the simulation studies were carried out with a targeted sample size of 1000 (500 cases and 500 controls) and the interim time point being set at when half of the samples were obtained. The samples were generated following the multinomial distribution among cases ($Mul(r; p_0, p_1, p_2)$) and controls ($Mul(s; q_0, q_1, q_2)$) that were specified by the disease prevalence (K) of 0.1; the minor allele frequency (MAF) of 0.1, 0.3, or 0.5; the underlying genetic model (dominant, additive, or recessive); and the genotype relative risk (GRR) for genotype AA ($\gamma_2 = f_2/f_0$) of 1 (null distribution), 1.2, 1.5, or 2.0. f_2 and f_0 are the penentrance factors for genotypes AA and aa, respectively. Both Hardy-Weinberg equilibrium (HWE) and its violation were considered in data generation. Ten thousand (10,000) replicates were used.

6.3.1 Probabilities of selecting the correct model

Table 6.1 shows the probabilities of selecting the three genetic models under various alternatives by using the MAX approach and the HWDTT approach described earlier. The sample size in this study is 500 (250 cases and 250 controls). Overall, both approaches show advantages and disadvantages over each other. The MAX approach tends to aggressively choose between the dominant model (DOM) and the recessive model (REC) by the data. On the other hand, the HWDTT approach conservatively puts majority of the weight on the additive model (ADD) and transfers the weight to the correct model (DOM or REC) as the data shows increasing evidence. Under

the null hypothesis of no association, the MAX approach only has a 20% of chance to choose the additive model and a 40% chance to choose either the dominant or the recessive models. The HWDTT approach puts 90% bet on the additive model, and 5% on each of the other two models. When the true underlying model is dominant and recessive, the MAX approach has more probability to choose the correct model than the HWDTT approach. However, it also more likey chooses the opposite model. The HWDTT approach results in little chance of selecting the recessive when the true underlying model is dominant and vice versa. When the true model is additive, the MAX approach has similar probabilities choosing one of the three models, whereas the HWDTT approachs maintains its 90% chance of having a correct selection. As expected, for both approachs, the accuracy of the selection improves as the allele frequency (MAF) or the genotype relative risk increase. As a short summary for $\gamma_2 = 2.0$, the MAX approach has roughly 48% to 62% accuracy for choosing the additive model, and 60% to 87% for choosing the dominant or recessive models. The HWDTT approach, on the other hand, has 90% accuracy of selecting the additive model, about 30% for rare alleles (MAF=0.1) and 66% for common alleles (MAF=0.3 or 0.5) of selecting the correct dominant or recessive models.

6.3.2 Adaptive vs. non-adaptive designs

Table 6.2 presents the results of type I errors and powers for both the non-adaptive (Columns 4 to 7) and the adaptive two-stage group sequential designs (Columns 8 and 9). For reference, the power for the single stage design using the MAX statistic is also provided.

The test statistics used for the first stage included the model-specific tests for the dominant, additive and recessive genetic models, as well as the MAX statistic. For the model-specific test statistic, the p-values were obtained following the standard normal distribution. The p-value for the MAX statistic was obtained following integration of the tri-variate standard normal distribution executed by using the "pmvnorm"

function in the R software. The test statistic used for the second stage data include the same model-specific or the MAX test used for the first stage (non-adaptive), and the two adaptive statistics based on the MAX and the HWDTT selection approaches (adaptive). The adaptive test statistics were combined with the MAX statistics at the first stage in the assessment for powers and their p-values were obtained following standard normal distribution. For simplicity, we only allowed stopping for efficacy or significance. Therefore the parameters described in the Bauer and Kohne approach were set as: $\alpha_0 = 1$, $\alpha_1 = 0.0087$, and $c_\alpha = 0.0087$. It was considered statistically significant if p_1 or $p_1 p_2$ was less than 0.0087.

The simulation results confirmed the control of type I errors when using the combination test approaches for both the non-adaptive statistics and adaptive statistics. The robustness of the MAX test was also demonstrated within the non-adaptive tests, i.e., while the MAX test achieved less power than the locally optimal test as if the genetic model was correctly selected, it prevented significant power loss from that when the genetic model was misspecified. In the table, "ADP_1" represents the design using the adaptive statistic based on the MAX approach, and "ADP_2" represents the design using the adaptive statistic based on the HWDTT approach. Compared with the non-adaptive design using MAX as the test statistic for both stages, ADP_1 and ADP_2 show similar power achievement. The ADP_2 appears to provide slightly higher power (2-3%) than the MAX test when the allele frequency is rare (MAF=0.1) and especially the underlying genetic model is additive. On the other hand, the non-adaptive desing use MAX, or the adaptive design ADP_1 show slight advantage over the ADP_2 design for common alleles (MAF=0.3 or 0.5).

Table 6.3 presents the powers when there is a departure from HWE. The data were generated according to a Wright coefficient of inbreeding (F) of 0.05. As demonstrated by C. C. Li (1991), this will increase the frequencies of homozygotes by $Fp(1-p)$ while the frequencies of heterozygotes will be decreased uniformly by a factor of $(1 - F)$, where p is the minor allele frequency (MAF). The impact of departure from HWE has

little impact on type I error and are consistent with the previous findings. That is, when the underlying model is additive, the adaptive design using statistics selected by the HWDTT approach (ADP_2) is more favorable than the non-adaptive MAX test or the adaptive test by the MAX approach (ADP_1). When the underlying model is dominant and recessive, ADP_2 shows similar robustness as the other designs.

6.4 Summary

In this chapter, we studied the power properties for two proposed two-stage adaptive designs. Both designs target selections of test statistic that have better likelihood to represent the true underlying genetic model. The results suggested that both of the proposed genetic model selection approaches are limited in terms of degree of accuracy. The adaptation of test statistic based on the MAX selection approach provided very close results to the MAX statistic used at both stages. The HWDTT approach provides an adaptive statistic that results in better efficiency and slightly higher power.

While the idea of using adaptive designs based on moderate accurate genetic model selection is appealing, the approaches of selecting the correct genetic model may be further explored. For example, for the HWDTT approach, one might want to choose a cutoff that is more data driven and hence provides more power. For the MAX approach, there might be a way to adjust the bias of choosing the opposite model.

Table 6.1: Distribution (in percentage) of model selection by the maximum and HWDTT approaches

			Selection by MAX			Selection by HWDTT		
MAF	True Model	GRR (γ_2)	DOM	ADD	REC	DOM	ADD	REC
0.1	NULL	1	34.5	20.4	45.1	4.9	90.4	4.7
	DOM	1.2	39.9	25.5	34.7	6.8	90.4	2.8
		1.5	54.9	36.2	8.9	13.1	86.0	0.9
		2	66.4	33.4	0.3	31.9	68.0	0.1
	ADD	1.2	35.0	22.4	42.6	4.8	90.5	4.7
		1.5	38.2	32.8	29.1	5.5	90.5	4.0
		2	41.5	47.9	10.6	6.9	89.7	3.4
	REC	1.2	33.9	20.1	46.0	3.4	88.6	8.0
		1.5	30.1	19.7	50.2	2.0	83.3	14.8
		2	22.1	19.2	58.6	0.7	70.4	28.9
0.3	NULL	1	37.5	19.8	42.7	5.0	89.5	5.5
	DOM	1.2	47.9	24.5	27.5	11.9	86.6	1.6
		1.5	68.6	26.8	4.7	31.3	68.5	0.2
		2	82.8	17.2	0.1	66.5	33.5	0.0
	ADD	1.2	37.6	24.9	37.6	5.0	90.2	4.8
		1.5	36.4	40.3	23.3	6.1	90.1	3.9
		2	33.6	59.0	7.4	9.1	88.5	2.4
	REC	1.2	30.9	20.2	48.9	1.5	86.0	12.5
		1.5	13.8	20.4	65.8	0.3	68.4	31.3
		2	1.7	14.3	84.1	0.0	32.0	68.0
0.5	NULL	1	40.1	20.2	39.6	5.2	89.8	5.0
	DOM	1.2	50.6	21.6	27.8	13.5	85.2	1.3
		1.5	71.1	21.3	7.6	33.6	66.2	0.2
		2	85.6	13.8	0.6	65.0	35.0	0.0
	ADD	1.2	37.7	24.9	37.5	4.9	90.6	4.5
		1.5	33.3	40.5	26.2	6.4	90.0	3.7
		2	25.2	62.8	12.0	8.5	88.9	2.6
	REC	1.2	29.1	21.7	49.2	1.3	84.7	14.0
		1.5	5.7	22.1	72.2	0.2	64.2	35.7
		2	0.1	12.8	87.1	0.0	33.8	66.2

Table 6.2: Power comparisons among non-adaptive and adaptive designs under HWE (F=0)

			Two-Stage Non-Adaptive				Two-Stage Adaptive		Single Stage
MAF	Model	γ_2	DOM	ADD	REC	MAX	ADP_1	ADP_2	MAX
0.1	NULL	1	0.0492	0.0486	0.0406	0.0390	0.0391	0.0430	0.0498
	DOM	1.2	0.1959	0.1875	0.0396	0.1433	0.1432	0.1648	0.1962
		1.5	0.7716	0.7456	0.0657	0.6767	0.6778	0.7133	0.7827
		2	0.9996	0.9987	0.1239	0.9979	0.9951	0.9984	0.9996
	ADD	1.2	0.0911	0.0924	0.0472	0.0693	0.0721	0.0824	0.0973
		1.5	0.3030	0.3159	0.0718	0.2458	0.2376	0.2770	0.3351
		2	0.8159	0.8298	0.1540	0.7544	0.7449	0.7859	0.8471
	REC	1.2	0.0527	0.0523	0.0472	0.0428	0.0478	0.0491	0.0608
		1.5	0.0496	0.0575	0.0822	0.0613	0.0657	0.0630	0.0935
		2	0.0584	0.0846	0.2093	0.1485	0.1493	0.1325	0.2257
0.3	NULL	1	0.0496	0.0498	0.0493	0.0478	0.0484	0.0476	0.0522
	DOM	1.2	0.2844	0.2449	0.0682	0.2297	0.2283	0.2366	0.3034
		1.5	0.9036	0.8311	0.1427	0.8420	0.8418	0.8384	0.9131
		2	0.9999	0.9986	0.3103	0.9994	0.9991	0.9993	0.9999
	ADD	1.2	0.1309	0.1449	0.0887	0.1290	0.1241	0.1350	0.1653
		1.5	0.5165	0.5846	0.2741	0.5201	0.5060	0.5495	0.6305
		2	0.9607	0.9797	0.6897	0.9649	0.9579	0.9721	0.9835
	REC	1.2	0.0536	0.0848	0.1217	0.0968	0.0970	0.0922	0.1245
		1.5	0.0923	0.2763	0.5023	0.3999	0.4006	0.3579	0.5144
		2	0.2252	0.7585	0.9653	0.9239	0.9288	0.8847	0.9653
0.5	NULL	1	0.0495	0.0516	0.0505	0.0512	0.0503	0.0504	0.0515
	DOM	1.2	0.2080	0.1525	0.0635	0.1642	0.1647	0.1614	0.2186
		1.5	0.7551	0.5392	0.1235	0.6572	0.6546	0.6137	0.7664
		2	0.9926	0.9290	0.2300	0.9826	0.9784	0.9670	0.9937
	ADD	1.2	0.1220	0.1582	0.1211	0.1438	0.1385	0.1492	0.1833
		1.5	0.4615	0.6135	0.4178	0.5575	0.5370	0.5883	0.6596
		2	0.9129	0.9767	0.8431	0.9596	0.9520	0.9691	0.9819
	REC	1.2	0.0707	0.1652	0.2315	0.1802	0.1779	0.1769	0.2374
		1.5	0.1656	0.6886	0.8513	0.7656	0.7697	0.7368	0.8618
		2	0.4633	0.9956	0.9999	0.9993	0.9989	0.9988	0.9999

Table 6.3: Power comparisons among non-adaptive and adaptive designs without HWE (F=0.05)

			Two-Stage Non-Adaptive				Two-Stage Adaptive		Single Stage
MAF	Model	γ_2	DOM	ADD	REC	MAX	ADP_1	ADP_2	MAX
0.1	NULL	1	0.0540	0.0554	0.0437	0.0472	0.0488	0.0500	0.0494
	DOM	1.2	0.1948	0.1815	0.0505	0.1483	0.1457	0.1617	0.1914
		1.5	0.7649	0.7341	0.0854	0.6715	0.6772	0.6985	0.7806
		2	0.9989	0.9980	0.1792	0.9965	0.9946	0.9969	0.9989
	ADD	1.2	0.0899	0.0907	0.0498	0.0732	0.0722	0.0812	0.0969
		1.5	0.3143	0.3343	0.0967	0.2687	0.2577	0.2920	0.3553
		2	0.8216	0.8399	0.2325	0.7722	0.7646	0.7976	0.8575
	REC	1.2	0.0514	0.0545	0.0560	0.0562	0.0531	0.0543	0.0613
		1.5	0.0559	0.0691	0.1085	0.0856	0.0890	0.0879	0.1176
		2	0.0707	0.1311	0.3031	0.2239	0.2244	0.2125	0.3115
0.3	NULL	1	0.0519	0.0541	0.0484	0.0509	0.0518	0.0511	0.0541
	DOM	1.2	0.2757	0.2354	0.0741	0.2228	0.2192	0.2277	0.3000
		1.5	0.8995	0.8251	0.1610	0.8393	0.8410	0.8367	0.9130
		2	1.0000	0.9988	0.3603	0.9998	0.9995	0.9997	1.0000
	ADD	1.2	0.1321	0.1543	0.0922	0.1354	0.1266	0.1391	0.1731
		1.5	0.5344	0.6091	0.3141	0.5452	0.5265	0.5746	0.6545
		2	0.9659	0.9820	0.7412	0.9684	0.9635	0.9746	0.9870
	REC	1.2	0.0581	0.0895	0.1341	0.1052	0.1058	0.0989	0.1371
		1.5	0.1074	0.3211	0.5510	0.4508	0.4534	0.4128	0.5660
		2	0.2919	0.8328	0.9768	0.9504	0.9506	0.9203	0.9801
0.5	NULL	1	0.0491	0.0503	0.0499	0.0511	0.0502	0.0535	0.0488
	DOM	1.2	0.2188	0.1668	0.0740	0.1805	0.1765	0.1777	0.2255
		1.5	0.7623	0.5571	0.1274	0.6683	0.6689	0.6273	0.7721
		2	0.9960	0.9428	0.2657	0.9873	0.9852	0.9771	0.9965
	ADD	1.2	0.1328	0.1676	0.1232	0.1478	0.1447	0.1547	0.1948
		1.5	0.4837	0.6341	0.4362	0.5759	0.5552	0.6012	0.6762
		2	0.9317	0.9804	0.8669	0.9682	0.9595	0.9750	0.9857
	REC	1.2	0.0680	0.1742	0.2383	0.1823	0.1805	0.1826	0.2520
		1.5	0.1812	0.7054	0.8509	0.7746	0.7791	0.7517	0.8600
		2	0.5117	0.9960	0.9998	0.9993	0.9988	0.9988	0.9997

Chapter 7

Discussion and Future Research in Related Areas

7.1 Discussion

Group sequential designs have been shown to achieve efficiencies in time and cost in research studies. In recent years, genetic studies which examine association and linkage between diseases and genetic markers have become popular as the genetic testing technologies are broadly available. As a result, the identification of the susceptible genetic markers can directly help researchers to target the development of treatment and prevention of the disease from the genetic aspect. One unique aspect of the statistical tests employed in genetic studies is that they can be constructed to be optimal according to the underlying genetic model. However, for many complex diseases, the genetic model is often unknown. To resolve this problem, researchers have proposed robust statistics, among which the MAX statistic has become favorable.

This dissertation explores several aspects of applying group sequential designs in genetic studies with a robust or an adaptive statistic. It first presents approaches of obtaining critical values for the robust statistic MAX, which does not follow asymptotic normal distribution and requires multiple integrations in its distribution function. These methods are alternative applications of Efron [1997] in calculating the

upper bound of simultaneous hypothesis tests with a normal distribution. The attractiveness of these methods is that they reduce multi-dimensional calculations into one or two-dimension ones. They are based on approximations and lead to conservative result. The precision relies on the number of statistics used in the simultaneous tests and the correlations among them. Our results showed that the type I errors were adequately controlled under these approaches in various settings of genetic studies. Looking forward, they shall make multi-stage (3 or more) group sequential robust studies immediately feasible without complex calculation procedures and excessive computing time.

While applying group sequential robust designs seems practical and feasible with the approaches presented in the dissertation, there are several design parameters that can affect the study efficiency. In a series of simulations covering a wide spectrum of genetic effect mechanisms, we proposed optimal designs in terms of the choices of design parameters such as the allocation fraction and alpha spending function and studies the impact of other parameters on the designs. For example, we demonstrated that earlier looks and more aggressive spendings of alpha for interim monitoring can promote about 50% savings under strong alternatives. These results can be a good reference for study designs in the future.

Finally, as an innovative step, this dissertation also presents a special, more complex type of group sequential design that allows adaptation or modification of the test statistic inferring data from the first stage. The adaptive designs have shown to be appealing because of their efficiencies and likelihood of providing higher power than non-adaptive robust designs. The adaptation is based on genetic model selection according to the observed data. We used the MAX and the HWDTT approach, and preliminarily found that the later approach provided slightly higher power than the non-adaptive robust design.

In summary, the group sequential robust design combines two pieces: the group sequential design and the robust test, which have been individually studied in the

literature. This dissertation provides the statistical methods needed to execute such designs in practice for different types of genetic studies, from planning a study with stopping rules, targeted sample size and power, to allowing modification of the study plan with an adaptive approach. Our results show that this design can be readily put to use and will gain benefits just as those in traditional group sequential designs. On the other hand, being preliminary, they identify areas that can be potentially improved through future research. Below are the discussions of some of the areas.

7.2 Limitations and research needs

7.2.1 Extension of the two-stage robust design

In this dissertation, although we focused on two-stage group sequential design, our approach can be extended to more than two stages. The degree of accuracy of the approximation, however, may decrease when there are too many interim looks and when the test statistics are less correlated. Therefore an improved calculation method is needed to provide more precise results.

The major focus of this dissertation is on a single candidate-gene association study, which is more suitable for study in its later stage after a few suspected alleles are identified after massive screenings. The application of group sequential designs needs to be modified when applied in a genome-wide association study (GWAS), where hundreds of thousands of single-nucleotide polymorphisms (SNPs) are genotyped. Both the sampling and genotyping costs might be taken into account [Satagopan and Elston 2003, Satagopan *et al.* 2004, Müller *et al.* 2007, Elston *et al.* 2007]. The stopping rule needs to be modified and considered together with other restrictions in practice. One may consider "the no-early-stopping" adaptive designs that hopefully selects with high likelihood of the underlying genetic model and conduct a more powerful test in the end.

The current proposed design may be extended to non-genetic studies. For example, for survival studies, the assumption of proportional hazard is often violated. It is common to conduct a trial that monitors the survivals periodically. Different weight sets used in weighted log-rank tests may be optimal depending on the underlying data. Freidlin *et al.* [1999] demonstrated the robustness of applying MERT in this situation. Tarone [1981] proposed the maximum of the logrank statistic and modified Wilcoxon statistic. These statistics have shown robustness against power loss. One may apply group sequential design and use these statistics, following the approaches proposed in this dissertation. Because of the direct theories and derivations behind these approaches, the success of their application is promising.

7.2.2 Adaptive design and genetic model selection

The selection of the genetic model based on the observed data is challenging. The two approaches discussed in this dissertation show little advantage over the simple robust design using MAX in population-based case-control studies. Researchers have shown that using the maximum test statistic of the observed data to select the genetic model is biased and the adaptive procedure is shy a fraction of power compared to the non-adaptive robust procedure. The method of choosing the genetic model based on the HWDTT test has shown some advantages. Both methods may be improved through future research. First, it might help to explore how to correct the bias in the selection by maximum approach. Secondly, for the selection procedure by HWDTT, since we used a nominal cutoff of $\Phi^{-1}(0.95)$, it might be helpful to explore a more data driven cutoff that can result in higher probabilities of correct selection. Thirdly, it is also possible to combine the two selection methods and define a decision rule that selects the genetic model with higher accuracy. Last but not the least, genetic model selection methods are yet to be studied for family-based studies where the cases and controls are correlated.

In the adaptive design proposed in this dissertation, the statistical test was based

on traditional meta-analysis-like combination tests. These tests treat data from the two stages separately and consider the p-values are independent. There have been researches improving this combination test method, such as the weighted Z-score methods. Further investigation of performing a test using cumulative data in a true group sequential setting that consider correlations of the data as well as the test statistics is needed [Müller and Schäfer 2001].

Furthermore, adaptive studies that allow adaptations not only of test statistics, but also modification of targeted sample sizes, will be useful. Finally, extension to GWAS study with hundreds of genetic markers may be explored.

7.2.3 Repeated confidence intervals

In the proposed group sequential robust or adaptive designs, identification of the genetic markers that are significantly associated with the underlying disease is often of primary interest. Other statistical measures, such as the parameter estimates and their confidence intervals, are often desired. It is well known that estimation after a sequentially-run experiment can be biased if the sequential nature of the experiment is ignored. There has been a considerable amount of research on computing confidence intervals, point estimates and p-values in the group sequential design. The estimations are quite straightforward for the classical single statistic, non-adaptive group sequential design [Jennison and Turnbull 1989]. The principal used in controlling the overall type I error can be applied in parameter estimation and construction of their confidence intervals. In genetic studies using MAX, the rule will need to be modified to more accurately represent its distribution. There have been progressions of approaches proposed in the literature to deal with parameter estimation and confidence interval construction. Lawrence and Hung [2001] used a generalization of the adaptive statistic to produce a consistent point estimate and the confidence interval with asymptotically correct coverage for adaptive two-stage designs. Their approach did not encompass the group sequential setting, in which some α might be spent to

allow for early stopping. Lehmacher and Wassmer [1999] extended the repeated confidence interval approach to adaptive designs based on the inverse normal method. Their method permits data-driven sample size adaptations in a group sequential setting but does not accommodate other types of data-dependent changes. Mehta *et al.* [2007] extended the repeated confidence interval procedure by Jennison and Turnbull [1989] to the adaptive sequential setting and allowed a broad range of data-dependent changes to an ongoing group sequential trial, such as sample size changes, alterations in the spending function and alterations in the number of spacing of interim looks. Therefore, the extensions of these methods to adaptive designs in genetic studies need to be explored.

References

Armitage P (1955) Tests for linear trends in proportions and frequencies. *Biometrics*. **11**: 375386.

Bauer P, Brannath W, Posch M (2001) Flexible two-stage designs: An overview. *Method Inform Med*. **40**: 122-126.

Bauer P, Khne K (1994) Evaluation of experiments with adaptive interim analyses. *Biometrics*. **50**: 1029-1041. Correction in *Biometrics*. **52**: 380.

Betensky RA (1998) Construction of a continuous stopping boundary from an alpha spending function. *Biometrics*. **54**: 1061-1071.

Blackwelder WC, Elston RC (1985) A comparison of sibpair linkage tests for disease susceptibility loci. *Genet Epidemiol*. **2**: 8597.

Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. *Lancet*. **361**: 598-604.

Cochran WG (1954) Some methods for strengthening the common chi-squared tests. *Biometrics*. **10**: 417-451.

Cox DR, Hinkley DV (1974) *Theoretical statistics*. Chapman & Hall, London.

Efron B (1997) The length heuristic for a simultaneous hypothesis tests. *Biometrika*. **84**: 143-147.

Elston RC, Lin DY, Zheng G (2007) Multi-stage sampling for genetic studies. *Annu Rev Genomics Hum Genet*. **8**: 327-342.

Evangelou E, Trikalinos TA, Salanti G, Ioannidis JPA (2006) Family-based versus unrelated case-control designs for genetic associations. *PLoS Genet*. **2**: e123.

Ewens WJ, Spielman RS (1995) The transmission/disequilibrium test: history, subdivision, and mixture. *Am J Hum Genet.* **57**: 455-464.

Fisher L (1998) Self-designing clinical trials. *Stat Med.* **17**: 1551-1562

Freidlin B, Podgor MJ, Gastwirth JL (1999) Efficiency robust tests for survival or ordered categorical data. *Biometrics.* **55**: 883-886.

Freidlin B, Zheng G, Li Z, Gastwirth JL (2002) Trend tests for case-control studies of genetic markers: Power, sample size and robustness. *Hum Hered.* **53**: 146-152.

Gastwirth JL (1966). On robust procedure. *J Am Statist Assoc.* **61**: 929-948.

Gastwirth JL (1985) The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *J Am Statist Assoc.* **80**: 380-384.

Gastwirth JL, Freidlin B (2000) On power and efficiency robust linkage tests for affected sibs. *Ann Hum Genet.* **64**: 443-453.

Hattersley AT, McCarthy M (2005) What makes a good genetic association. *Lancet.* **366**: 1315-1323.

Hedges LV, Olkin I (1985) *Statistical Methods for Meta-Analysis.* Academic press, New York.

Hommel G, Lindig V, Faldum A (2005) Two-stage adaptive designs with correlated test statistics. *J Biopharm Stat.* **15**: 613-623.

Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager N, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet.* **39**: 870874.

Jennison C, Turnbull BW (1989) Interim analyses: the repeated confidence interval approach (with Discussion). *J Roy Stat Soc B.* **51**: 305-361.

Jennison C, Turnbull BW (2000) *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC. Boca Raton: FL.

Kieser M, Bauer P, Lehmacher W (1999) Inference on multiple endpoints in clinical trials with adaptive interim analyses. *Biometrical J.* **41**: 261-277.

Kim K, DeMets DL (1987) Design and analysis of group sequential tests based on Type I error spending functions. *Biometrika*. **74**: 149-54.

Klein RJ, Zeiss C, Chew EY, Tsai J-Y, Sackler RS, Haynes C, Henning AK, San-Giovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL et al. (2005) Complement factor H polymorphism in aged-related macular degeneration. *Science*. **308**: 385-389.

Knapp M, Seuchter SA, Baur MP (1994) Linkage analysis in nuclear families. 1: Optimality criteria for affected sib-pair tests. *Hum Hered*. **44**: 3743.

Konig IR, Schafer H, Muller H, Ziegler A (2001) Optimized group sequential study designs for tests of genetic linkage and association in complex diseases. *Am J Hum Genet*. **69**: 590-600.

Konig IR, Schafer H, Ziegler A, Muller H (2003) Reducing sample sizes in genome scans: Group sequential study designs with futility stops. *Genet Epidemiol*. **25**: 339-349.

Konig IR, Ziegler A (2003) Group Sequential study designs in genetic epidemiological case-control studies. *Hum Hered*. **56**: 63-72.

Kropf S, Hommel G, Schmidt U, Brickwedel J, Jepsen MS (2000) Multiple comparisons of treatments with stable multivariate tests in a two-stage adaptive design, including a test for noninferiority. *Biometrical J.* **42**: 951-965.

Lan KKG, DeMets DL (1983) Discrete sequential boundaries for clinical trials. *Biometrika*. **70**: 659-663.

Lang T, Auterith A, Bauer P (2000) Trend tests with adaptive scoring. *Biometrical J.* **42**: 1007-1020.

Lawrence J, Hung HMJ (2003) Estimation confidence intervals after adjusting the maximum information. *Biometrical J.* **45**: 143-152.

Lehmacher W, Wassmer G (1999) Adaptive sample size calculations in group sequential trials. *Biometrics*. **45**: 1286-1290.

Li CC (1969) Population subdivision with respect to multiple alleles. *Ann Hum Genet*. **33**: 23-29.

Li CC (1991) Genetics of subdivided populations and its relationships with certain measures of association. *Genet Epidemiol*. **8**: 1-11.

Li Z, Gastwirth JL (2003) On the power of affected relative pair designs for linkages studies. *Ann Hum Genet*. **68**: 65-68.

Li Z, Gastwirth JL, Gail M (2005) Power and related statistical properties of conditional likelihood score tests for association studies in nuclear families with parental genotypes. *Ann Hum Genet*. **69**: 296-314.

Li Q, Yu K, Li Z, Zheng G (2008a) MAX-rank: a simple and robust genome-wide scan for case-control association studies. *Hum Genet*. **123**: 617-623.

Li Q, Zheng G, Li Z, Yu K (2008b) Efficient approximation of p-value of the maximum of correlated tests, with applications to genome-wide association studies. *Ann Hum Genet*. **72**: 397-406.

Mehta CR, Bauer P, Posch M, Brannath W (2007) Repeated confidence intervals for adaptive group sequential trials. *Stat Med*. **26**: 5422-5433.

Müller HH, Pahl R, Schäfer H (2001) Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches. *Biometrics*. **57**: 886-891.

Müller HH, Pahl R, Schäfer H (2007) Including sampling and phenotyping costs into the optimization of two stage designs for genome wide association studies. *Genet Epidemiol*. **31**: 844-852.

O'Brien PC, Fleming TR (1979) A multiple testing procedure for clinical trials. *Biometrics*. **35**: 549-556.

ONeill R (2000) New approaches in data management and biostatistics. *International Journal of Pharmaceutical Medicine*. **14**: 241.

Pocock SJ (1977) Group sequential methods in the design and analysis of clinical trials. *Biometrika*. **64**: 191-199.

Risch N (1990a) Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet Epidemiol* **7**: 316.

Risch N (1990b) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet*. **46**: 229241.

Sasieni PD (1997) From genotypes to genes: Doubling the sample size. *Biometrics*. **53**: 1253-1261.

Satagopan JM, Elston RC (2003) Optimal two-stage genotyping in population-based association studies. *Genet Epidemiol*. **25**: 149-157.

Satagopan JM, Venkatraman ES, Begg CB (2004) Two-stage designs for gene-disease association studies with sample size constraints. *Biometrics*. **60**: 589-597.

Schaid DJ (1998). Transmission disequilibrium, family controls , and great expectations. *Am J Hum Genet*. **63**: 935-941.

Schaid DJ, Nick TG (1990) Sib-pair linkage tests for disease susceptibility loci: common tests vs. the asymptotically most powerful test. *Genet Epidemiol.* **7**: 359-370.

Schaid DJ, Sommer SS (1993) Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am J Hum Genet.* **53**: 1114-1126.

Selwyn MR, Fish SM (2004) Choice of alpha spending function and time points in clinical trials with one or two interim analyses. *Pharmaceut. Statist.* **3**: 193-203.

Shu Y, Liu A, Li Z (2007) Sequential evaluation of a medical diagnostic test with binary outcomes. *Stat in Med.* **26**: 4416-4427.

Simon R (1989) Optimal two-stage designs for phase II clinical trials. *Control Clin Trials.* **10**: 1-10.

Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S *et al.* (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature.* **445**: 881885.

Song K, Elston RC (2003) The Hardy-Weinberg disequilibrium (HWD) measure and test statistics for a disease-susceptibility locus with multiple alleles allowing for an inbreeding coefficient (F). *Genetica.* **119**: 269-293.

Song K, Lu Q, Lin X, Waterworth D, Elston RC (2007) Genome-wide association studies using an adaptive two-stage analysis for a case-control design. *BMC Proc.* **1(Suppl 1)**: S147.

Spielman R, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet.* **52**: 506-516.

Tarone RE, Gart JJ (1980) On the robustness of combined tests for trends in proportions. *J Am Stat Assoc.* **75**: 110116.

Tarone RE (1981) On the distribution of the maximum of the logrank statistic and the modified Wilcoxon statistic. *Biometrics.* **37**: 79-85.

Wang S, Tsiatis AA (1987) Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics.* **43**: 193-199.

Wang W, Barratee B, Clayton D, Todd J (2005) Genome-wide association studies: Theoretical and practical concerns. *Nat Rev Genet.* **6**: 109-118.

Weir BS (1996) *Genetic Data Analysis II: Methods for Discrete Population Genetic Data.* Sunderland, MA: Sinauer Associations Inc.

Whittemore AS, Tu IP (1998) Simple, robust linkage test for affected sibs. *Am J Hum Genet.* **62**: 12281242.

Whittemore AS, Tu IP (2000). Detection of disease genes by use of family data. I. Likelihood-based theory. *Am J Hum Genet.* **66**: 1328-1340.

Wittke-Thompson JK, Pluzhnikov A, Cox NJ (2005) Rational inferences about departure from Hardy-Weinberg equilibrium. *Am J Hum Genet.* **76**: 967-986.

The Wellcome Trust Case Control Consortium (WTCCC) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature.* **447**: 661-683.

Yan LK, Zheng G, Li Z (2008) Two-stage group sequential robust tests in family-based association studies: controlling type I error. *Ann Hum Genet.* **72**: 557-565.

Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, Minichiello MJ, Fearnhead P, Yu K, Chatterjee N et al. (2007) Genome-wide association study

of prostate cancer identifies a second risk locus at 8q24. *Nat Genet.* **39**: 645-649.

Zheng G, Freidlin B, Gastwirth JL (2002) Robust TDT-type candidate-gene association test. *Ann Hum Genet.* **66**: 145-155.

Zheng G, Freidlin B, Li Z, Gastwirth JL (2002) Choice of scores in trend tests for case-control studies of candidate-gene associations. *Biometrical J.* **45**: 335-348.

Zheng G, Chen Z (2005) Comparison of maximum statistics for hypothesis testing when a nuisance parameter is only present under the alternative. *Biometrics.* **61**: 254-258.

Zheng G, Ng H (2008) Genetic model selection in two-phase analysis for casecontrol association studies. *Biostatistics.* **9**: 391-399.

Appendix: R Programs:
Obtaining Critical Values Using Two-Point Formula

```
#########################################################
# Author:  Lihan Yan
# Date:    10/13/2006
# Purpose: Using Newton-Raphson Method to find c1,c2
#  in Efron(1997) two point formula given L and alpha
#########################################################
library(mvtnorm)

#function generating correlation matrix given rho1, rho2, rho3

#########################################################
#Function Name:  makecorr
#Parameter:      rho1, rho2, rho3
#Purpose:        Generate correlation matrix
#########################################################
makecorr <- function(rho1,rho2,rho3) {
 ##set up the correlation matrix
 corr1 = matrix(NA, nrow=J, ncol=J)
 corr2 = corr1
 corrmat = matrix(NA, nrow=J*2, ncol=J*2)

 ### construct correlation
 ####corr1 is the base
 corr1[1,2]=rho1
 corr1[1,3]=rho2
 corr1[2,3]=rho3

 for(j in 1:J) {
        for(k in j:J) {
           if(j==k) (corr1[j,k]=1)
           else (corr1[k,j]=corr1[j,k])
  }
 }

 ## corr2 modified by a fraction due to information change assume 1/2
 corr2 = sqrt (1/2) * corr1

 ## corrmat combines corr1 and corr2
 for(j in 1:J) {
```

```
        for(k in 1:J) {
        corrmat[j, k]= corr1[j,k]
        corrmat[j+3, k+3]=corr1[j,k]
        corrmat[j,k+3]=corr2[j,k]
        corrmat[j+3,k]=corr2[j,k]
   }
 }
 return(corrmat)
}

############################################################
# Function:  getminl
# Purpose: output the smallest L through
# permutation
############################################################
#library(mvtnorm)
##initialization

#number of tests at each stage
#J <- 3
# sample size or anything that affect information change
getminl <- function(rho1,rho2,rho3){

 corrmat <- makecorr(rho1,rho2,rho3)
 minL <- 1000  #arbitary value to initiate lower bound
 tempL <- rep(NA, J*2-1)
 L <- rep(NA, J*2-1)

 for(r1 in 1:J){
 for(s1 in 1:J){
 for(t1 in 1:J){
 for(r2 in 1:J){
 for(s2 in 1:J){
 for(t2 in 1:J){
   if (r1 != s1 && r1 != t1 && s1 != t1 && r2 != s2 && r2 != t2 && s2 != t2){
    tempL[1] = acos(corrmat[r1,s1])
    tempL[2] = acos(corrmat[s1,t1])
    tempL[3] = acos(corrmat[t1,r2+3])
    tempL[4] = acos(corrmat[r2+3,s2+3])
    tempL[5] = acos(corrmat[s2+3,t2+3])

    if (sum(tempL) < minL) {
      minL=sum(tempL)
```

```
      L=tempL
      o <- c(r1,s1,t1, r2+3,s2+3,t2+3)  #order of the optimal
    }
   # print(cbind(r1,s1,t1,r2,s2,t2,t(tempL[1:5])))
   } #end if
  }}}}}} #close 6 for loops

  return(cbind(t(L), t(o)))
}

#### Example: call FUNCTION getminl
#rho1<-sqrt(2/3)
#rho2<-sqrt((1+sqrt(2/3))/2)
#rho3<-rho2
#L<-getl(rho1,rho2,rho3)


##initialization

#number of tests at each stage
J <- 3

############################################################
# Function Name:  integrand and der
# Purpose:    function inside the phibar function and its
#             derivatives in prep for fc and dfc functions
############################################################
# integrand part in f step
integrand<-function(x,cj,rhoj) {
        z<-(cj-rhoj*x)/sqrt((1-rhoj^2))
        out<-(1-pnorm(z))*dnorm(x)
        return(out)
}

#integrand part in f derivative step
integrandf<-function(x,cj,rhoj) {
        z<-(cj-rhoj*x)/sqrt((1-rhoj^2))
        out <- dnorm(z)*dnorm(x)*1/(sqrt(1-rhoj^2))
        #out<-(1-pnorm(z))*dnorm(x)
        return(out)
}
```

```
#der <- function(cj,rhoj) {
    #z<-(cj-rhoj*cj)/sqrt((1-rhoj^2))
    #out<- ((1-pnorm(z))*dnorm(cj))+(-pnorm(cj))*dnorm(z)*1/(sqrt(1-rhoj^2))
#    return(out)
#}

#stage 1 max(t11,t12,t13) and c1

f2ptc1 <-function(crit1,alpha, r1, r2) {
   sumpart1<-integrate(integrand,lower=-Inf,upper=crit1,cj=crit1,rhoj=r1)$value
   sumpart2<-integrate(integrand,lower=-Inf,upper=crit1,cj=crit1,rhoj=r2)$value

   twoptc1 <- 1-pnorm(crit1) + sum(sumpart1, sumpart2) - alpha
   return(twoptc1)
}

df2ptc1 <-function(crit1, r1, r2) {

   #dsumpart1<- der(cj=crit1, rhoj=r1)
   #dsumpart2<- der(cj=crit1, rhoj=r2)
   dsumpart1 <- integrate(integrandf, lower=-Inf, upper=crit1, cj=crit1, rhoj=r1)$
   dsumpart2 <- integrate(integrandf, lower=-Inf, upper=crit1, cj=crit1, rhoj=r2)$
   leftover1 <- (1-pnorm((crit1-r1*crit1)/sqrt(1-r1^2)))*dnorm(crit1)
   leftover2 <- (1-pnorm((crit1-r2*crit1)/sqrt(1-r2^2)))*dnorm(crit1)

   dtwoptc1<-(-dnorm(crit1))-sum(dsumpart1, dsumpart2)+leftover1+leftover2
   return(dtwoptc1)
}

#stage 2 max(t21,t22,t23) and c2 given c1
f2ptc2 <-function(crit1,crit2, alpha, r3,r4,r5) {
   sumpart1<-integrate(integrand,lower=-Inf,upper=crit1,cj=crit2,rhoj=r3)$value
   sumpart2<-integrate(integrand,lower=-Inf,upper=crit2,cj=crit2,rhoj=r4)$value
   sumpart3<-integrate(integrand,lower=-Inf,upper=crit2,cj=crit2,rhoj=r5)$value

   twoptc2 <- sumpart1+sumpart2+sumpart3 - alpha
   return(twoptc2)
}

df2ptc2 <-function(crit1, crit2, r3,r4,r5) {

   dsumpart1 <- integrate(integrandf, lower=-Inf, upper=crit1, cj=crit2, rhoj=r3)$
   dsumpart2 <- integrate(integrandf, lower=-Inf, upper=crit2, cj=crit2, rhoj=r4)$
```

```
   dsumpart3 <- integrate(integrandf, lower=-Inf, upper=crit2, cj=crit2, rhoj=r5)$
   leftover1 <- (1-pnorm((crit2-r4*crit2)/sqrt(1-r4^2)))*dnorm(crit2)
   leftover2 <- (1-pnorm((crit2-r4*crit2))/sqrt(1-r5^2))*dnorm(crit2)

   dtwoptc2<- -dsumpart1-dsumpart2-dsumpart3+leftover1+leftover2
   return(dtwoptc2)
}


#############################################################
# Function Name:  find2ptc1, find2ptc2
# Parameters:   alpha = significance level
#############################################################
find2ptc1 <- function(alpha, r1, r2){
 crit <- rep(NA, 30)
 crit[1] <- qnorm(1-alpha)
 i <- 0
 eps1 <- 100

 while (eps1>1e-5) {
 #while (i<=20) {
  i <- i + 1
  crit[i+1]<-crit[i]-f2ptc1(crit[i], alpha, r1, r2)/df2ptc1(crit[i], r1, r2)
  eps1 <- abs(crit[i+1]-crit[i])
 # print(eps)
 }

 bound <- crit[i+1]
 #print(cbind('bound <- ',bound))
 return(cbind(bound, alpha))
}    #end find2ptc1

###iterative steps find c2
find2ptc2 <- function(crit1, alpha, r3, r4, r5){
   crit <- rep(NA, 30)
   crit[1] <- qnorm(1-alpha)
   i <- 0
   eps <- 100

   while (eps>1e-5) {
     i <- i + 1
     crit[i+1]<-crit[i]
       -f2ptc2(crit1,crit[i],alpha,r3,r4,r5)/df2ptc2(crit1,crit[i],r3,r4,r5)
```

```
     eps <- abs(crit[i+1]-crit[i])
    # print(eps)
   }

   bound <- crit[i+1]
   #print(cbind('bound <- ',bound))
   return(cbind(bound, alpha))
}    #end find2ptc2



################################################
# Two-stage procedure to find c1,c2
# Parameters:  corrmat, alpha1, alpha2
# Called function:  getl, makecorr, find2ptc
# Output:
################################################

findc1c2 <- function(rho1, rho2, rho3, alpha1,alpha2) {
   #set sequential correlations
   corr1 <- makecorr(rho1, rho2, rho3)
   corrmat <- matrix(NA, nrow=6, ncol=6)

   o<-getminl(rho1,rho2,rho3)[6:11]
   #print(o)

   for(i in 1:(J*2)){
      for(j in 1:(J*2)){
        corrmat[i,j]=corr1[o[i],o[j]]
       } } #close corrmat

   rho<-rep(NA,5)
   #print(corrmat)
   for (i in 1:5) {
     rho[i] = corrmat[i,i+1]
   }

   c1<-find2ptc1(alpha1, rho[1],rho[2])[1]
   c2<-find2ptc2(c1,alpha2, rho[3],rho[4],rho[5])[1]

   print(cbind(alpha1, alpha2, c1, c2))
   return(cbind(alpha1, alpha2, c1, c2))
}
```

```
# end function findc1c2


#sink('c:/OUTFILE.out')

####EXAMPLE:  INPUT DATA
#rho3<-0.086279596
#rho1<-0.960768923
#rho2<-0.359210604
#print("allele freq p=0.1")
#findc1c2(rho1,rho2,rho3, 0.001525323, 0.023474677)
#findc1c2(rho1,rho2,rho3, 0.01550286,0.00949714)
#findc1c2(rho1,rho2,rho3, 0.0125, 0.0125)
#findc1c2(rho1,rho2,rho3, 0.008838835,0.016161165)
#findc1c2(rho1,rho2,rho3, 0.00625,0.01875)
#findc1c2(rho1,rho2,rho3, 0.010, 0.015)
```