ON MODEL SELECTION STRATEGIES TO IDENTIFY GENES UNDERLYING BINARY TRAITS USING GENOME-WIDE ASSOCIATION DATA

Zheyang Wu and
Department of Mathematical Sciences, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA, 01609, USA, URL: http://users.wpi.edu/zheyangwu/

Hongyu Zhao
Department of Epidemiology and Public Health, Yale University School of Medicine, 300 George Street, Suite 503, New Haven, CT 06520, USA, URL: http://bioinformatics.med.yale.edu

Zheyang Wu: zheyangwu@wpi.edu; Hongyu Zhao: hongyu.zhao@yale.edu

Abstract

For more fruitful discoveries of genetic variants associated with diseases in genome-wide association studies, it is important to know whether joint analysis of multiple markers is more powerful than the commonly used single-marker analysis, especially in the presence of gene-gene interactions. This article provides a statistical framework to rigorously address this question through analytical power calculations for common model search strategies to detect binary trait loci: marginal search, exhaustive search, forward search, and two-stage screening search. Our approach incorporates linkage disequilibrium, random genotypes, and correlations among score test statistics of logistic regressions. We derive analytical results under two power definitions: the power of finding all the associated markers and the power of finding at least one associated marker. We also consider two types of error controls: the discovery number control and the Bonferroni type I error rate control. After demonstrating the accuracy of our analytical results by simulations, we apply them to consider a broad genetic model space to investigate the relative performances of different model search strategies. Our analytical study provides rapid computation as well as insights into the statistical mechanism of capturing genetic signals under different genetic models including gene-gene interactions. Even though we focus on genetic association analysis, our results on the power of model selection procedures are clearly very general and applicable to other studies.

Keywords and phrases

model selection; statistical power; random predictor; genome-wide association studies; gene-gene interaction

1. Introduction

Marker-by-marker analyses in genome-wide association studies (GWAS) have unraveled many genetic variants associated with a variety of complex traits. However, the current progresses are still limited in two aspects. First, for some diseases, such as asthma and coronary heart disease, fewer novel loci have been found than those for other diseases [7]. Second, the discovered genes only account for a small proportion of genetic risk in most diseases [5]. As a result, developing more sophisticated methods to better identify genetic variants associated with diseases has become a main focus of GWAS data analysis after initial scanning through single marker analysis.
Because of the genetic complexity of common diseases, a joint consideration of multiple markers is intuitively more informative when multiple genes and their interactions are involved in disease etiology. However, joint methods often lead to a sharp increase in the computational burden and in the stringency of statistical significance control, which can weaken their statistical power due to the large amount of candidate models considered. An optimal marker selection strategy should achieve a delicate balance between computational efficiency, satisfactory statistical power, and low error rates. To recognize the optimal marker selection strategy for a certain GWA study, researchers look for techniques to quickly evaluate possible strategies, marginal versus joint, for a variety of interesting genetic models.

There are three fundamental marker search strategies. The marginal search chooses the best fitted single-marker models separately. The exhaustive search selects the best fitted multiple-marker models from all possible combinations of predictors. The forward search looks for the preferred models conditional on the best fitted marker(s). Other marker search strategies are mostly extended from these three methods. For example, in a marginal-exhaustive two-stage search strategy, one can first screen the marker candidates through marginal search, and then choose the best multiple-marker models within the previously selected marker set. In the literature, the power evaluations of these methods have been explored by either simulations or with real data analyses [6][3][10][1]. Because real data are too specific and cannot be used in experimental design, and simulations are time-consuming and less insightful about the statistical mechanism of how genetic signals are captured, it is highly desirable to have analytical results.

The merits of analytical power calculations have been discussed and the methods for quantitative trait have been developed in the literature[12]. Because GWAS mostly focus on disease outcomes, which are binary, there is a need to derive results for binary traits. Statistically, genetic models for quantitative traits and binary traits are quite different. The genetic model for a quantitative trait is mostly based on a linear regression model with a random error component, whereas the genetic model for a binary trait specifies the disease risk for each possible genotype. Searching quantitative trait loci is commonly performed through fitting linear regression models, and the F-statistic is usually used to measure the model goodness-of-fit. On the other hand, searching binary trait loci is commonly performed through fitting logistic regression models, and the log-likelihood ratio test (LRT) or a score test statistic is generally applied for model comparisons. The distributions of the F-statistic and the LRT or a score statistic are quite different. So there is a need to rigorously explore whether the marker search methods will behave differently for binary traits in comparison to quantitative traits.

It is also of interest to study how error control criteria affect the relative performance of different search strategies. One contribution of this article is to compare and contrast two types of controls that have been widely applied in practice: discovery number control and type I error rate control. With the former, one collects a pre-specified number of models after ordering all candidate models. With the later, one selects models whenever they have test statistics that exceed the critical value based on a genome-wide type I error rate. The analytic power calculations help us to demonstrate and explain the power contrasts for each type of controls. Furthermore, we take into account linkage disequilibrium (LD) between the observed markers and the causative but unobserved loci. Besides the three basic search methods: the marginal search, the exhaustive search, and the forward search, this article also considers a marginal-exhaustive two-stage search strategy, which is practically appealing because of its computational efficiency.
We have implemented our method in an R package `markerSearchPower`, which provides researchers a convenient tool to find proper sample size in experimental design, to decide suitable strategies in data analysis, and to increase chance for true findings. Our statistical technique can also be used to address model selection problems of binary responses with general random predictor settings in other application areas.

The rest of this article is organized as follows. Section 2 sets up the genetic model to be studied, defines marker search strategies and statistical power. In Section 3, we introduce the score test statistics, derive asymptotic distributions for score tests and the power calculation formulas for four model selection strategies. In Section 4, the accuracy of our analytical results is demonstrated by simulations, and power comparisons among search strategies are illustrated in a large space of genetic parameters. In Section 5, we discuss the advantages and limitations of our analytical approaches, compare the results for quantitative traits and binary traits, and summarize the distinction between the performance under discovery number control and that under Bonferroni control.

2. Genetic Model and Marker Search

2.1. Model Setup

We assume that the odds of a binary trait (or hereafter described as “disease”) are specified by two loci which may or may not be directly genotyped. A genotype data set contains \( n \) independent individuals indexed by \( i = 1, \ldots, n \), and \( L \) candidate markers indexed by \( j, k = 1, \ldots, L \). A marker genotype is defined to be the number of its disease allele \( A \). Denoted by \( G_{ji} \) the random genotype value of the \( j \)th marker in the \( i \)th individual is

\[
G_{ji} = \begin{cases} 
2 & \text{Genotype} = A_jA_j, \text{ with probability } p_j^2 \\
1 & \text{Genotype} = A_ja_j, \text{ with probability } 2p_j(1-p_j), \\
0 & \text{Genotype} = a_ja_j, \text{ with probability } (1-p_j)^2 
\end{cases}
\]

where \( p_j \) is the disease minor allele frequency (MAF). Considering biallelic markers, the most general way to specify the underlying genetic model is through a 3-by-3 table of disease odds. Without loss of generality, we assume the first two markers, indexed by \( j = 1 \) or \( 2 \), are the associated markers which determine the odds of disease

\[
O(g_1, g_2) = \frac{p(D|g_1, g_2)}{p(D_{\overline{g_1}}, g_2)},
\]

where \((g_1, g_2)\) is a specific combination of genotypes, \( D \) and \( \overline{D} \) denote disease and non-disease, respectively. For example, the following are three common genetic models measuring different types of gene-gene interactions studied in the genetics literature [6]:

\[
\begin{array}{ccc}
A_1A_1 & A_2A_2 & A_2a_2 \\
A_1A_2 & \alpha (1+\theta_1)^2(1+\theta_2)^2 & \alpha (1+\theta_1)^2 \\
A_1a_1 & \alpha (1+\theta_1)(1+\theta_2)^2 & \alpha (1+\theta_1)(1+\theta_2) \\
a_1a_1 & \alpha (1+\theta_2)^2 & \alpha (1+\theta_2) \\
\end{array}
\]

Stat Sin. Author manuscript; available in PMC 2013 August 15.
In the above models, $\alpha$ and $\theta$’s are parameters for baseline and genotypic effects, respectively.

Based on the odds of disease, the genotypic disease risks are $p(D|g_1, g_2) = \frac{\alpha (1+\theta)^4}{\alpha (1+\theta)^2 + \alpha}$ and the joint distribution of genotypes in diseased individuals is

$$p(g_1, g_2|D) = \frac{p(D|g_1, g_2) p(g_1, g_2)}{\sum_{g_1} \sum_{g_2} p(D|g_1, g_2) p(g_1, g_2)}.$$

Similarly, we can get the joint distribution of genotypes in controls $p(g_1, g_2|\bar{D})$. For any marker-pair involving one non-associated marker $j$, it is clear that $p(g_1, g_j|D) = p(g_1|D) p(g_j)$ with $p(g_1|D) = \sum_{g_2} p(g_1, g_2|D)$. For any non-associated marker-pair, we have $p(g_j, g_k|D) = p(g_j) p(g_k), \; k > j \geq 3$.

The odds of disease can be defined through a logistic model

$$\log (O(g_1, g_2)) = b_0 + b_1 g_1 + b_2 g_2 + b_3 g_1 g_2. \; (2.4)$$

With various values of $b_1$, $b_2$ and $b_3$, equation (2.4) defines a flexible genetic interaction model. For example, the genetic model in (2.2) can be rewritten as a logistic model (2.4) with $b_0 = \log (\alpha)$, $b_1 = b_2 = 0$, $b_3 = \log (1 + \theta)$.

The genetic model can also be defined through a disease prevalence $p(D)$ together with a 3-by-3 table of genetic relative risks

$$GRR(g_1, g_2) = \frac{p(D|g_1, g_2)}{p(D|0, 0)}.$$

In this setup, it is easy to get the corresponding odds of disease through

$$p(D|g_1, g_2) = GRR(g_1, g_2) p(D|0, 0) \frac{GRR(g_1, g_2) p(D)}{\sum_{g_1} \sum_{g_2} GRR(g_1, g_2) p(g_1, g_2)}.$$
non-causative but associated with causative markers index by, say \(-1\) and \(-2\). Let marker 1 be in LD with marker \(-1\) and marker 2 be in LD with marker \(-2\). The odds of disease at markers 1 and 2 is

\[
O = \frac{p(D|g_1, g_2)}{p(D|g_1, g_2)} = \frac{\sum_{g_{-1}, g_{-2}} p(D|g_{-1}, g_{-2}) p(g_1|g_{-1}) p(g_2|g_{-2}) p(g_{-1}, g_{-2})}{\sum_{g_{-1}, g_{-2}} p(D|g_{-1}, g_{-2}) p(g_1|g_{-1}) p(g_2|g_{-2}) p(g_{-1}, g_{-2})}. \tag{2.5}
\]

The method of calculating \(p(g_1|g_{-1})\) and \(p(g_2|g_{-2})\) is illustrated in the supplement, which follows the LD models defined in the literature [6].

### 2.2. Model Fitting and Selection Procedures

We define the above genetic models of disease susceptibility by specifying the disease odds from a perspective data-generating point of view. In case-control retrospective studies, the samples are collected from a random sample of \(n_1\) cases and \(n_0\) controls. For a given GWAS data set, marker search requires to fit the following one-marker or two-marker logistic regression models:

\[
\logit(p_i) = \beta_0 + \beta_1 g_{ji}, \tag{2.6}
\]

\[
\logit(p_{jk}(Y_i=1|g_{ji}, g_{ki})) = \beta_{0jk} + \beta_{1jk} g_{ji} + \beta_{2jk} g_{ki} + \beta_{3jk} g_{ji} g_{ki}, \tag{2.7}
\]

where \(\logit(p) = \log(p/(1-p))\), \(g_{ji}\) and \(g_{ki}\) are the observed genotypes of markers \(j\) and \(k\) in individual \(i\), and \(Y_i=1\) or 0 indicates disease or non-disease status. The marginal search method looks for the best fitted models (2.6) over all single markers. The exhaustive search method seeks the best fitted models (2.7) over all marker-pairs. The forward search method first selects the best fitted model (2.6), then picks the best two-marker models (2.7) while retaining the previously chosen marker. Extended from these approaches, a marginal-exhaustive two-stage search method applies marginal search as the first stage and the exhaustive search in the chosen set of markers as the second stage. After any search procedure, the markers contained in the selected models are treated as the putative disease-associated markers.

The significance of statistical associations, or equivalently the goodness of model fitting, relies on either log-likelihood ratio test (LRT) or score test. As shown in the simulations below, the two tests are similar for the purpose to select markers when sample size is moderately large. Because LRT has no closed form, we used score test statistic to calculate power.

As a model can contain none, one, or two associated markers, it is necessary to consider both stringent and relaxed criteria to decide whether any chosen model is what we are looking for. Accordingly, the following two definitions of power in genetics literature [6] [10] are considered in any selection procedure:

#### A. Power is defined as the probability of identifying the true genetic association model (in marginal search, it is defined as the probability of detecting both associated markers).

#### B. Power is defined as the probability of detecting at least one associated marker.

We study two criteria for significance level control. The first is a discovery number control, where one selects the top \(R\) most significant models. The power under this control is a generalization of detection probability (DP) [4] into the context of model selection. The
second is a type I error rate control at a genome-wide significance level $\alpha$, which applies the Bonferroni correction according to the number of models to be compared. Specifically, in marginal search, the correction is $a/L$ for comparing totally $L$ one-marker models, the null distribution is $\chi^2_L$. In exhaustive search, the correction is $\alpha/L$ for comparing totally $L$ one-marker models, the null distribution is $\chi^2_L$. In the forward search, the correction of the first step is $a/L$ with the null distribution $\chi^2_L$; the correction of the second step is $a/(L - 1)$ with the null distribution $\chi^2_{L-1}$. In the following we construct the score test statistics for logistic regression, derive the null and alternative asymptotic distributions of relevant test statistics in the procedure of marker search, and present the formulae of power calculation under significance control criteria.

3. Model Selection Power

3.1. Score Test Statistic

We adopt a score test [13] into the context of genome-wide association case-control studies. Let $T^i = (T^i_1, \ldots, T^i_m)$, $i = 1, \ldots, n_1$, be a random sample of $m$ covariates of logistic regression in cases, $T^i = (T^i_0, \ldots, T^i_0)$, $i = 1, \ldots, n_0$, be a random sample of $m$ covariates in controls. $T^i_1$ and $T^i_0$ have distributions $p (t_1, \ldots, t_m | D)$ and $p (t_1, \ldots, t_m | \bar{D})$, respectively. Let $T_{m \times n} = (T_1, \ldots, T_n)$, $T_{m \times n} = (T'_1, \ldots, T'_n)$, $T'_1$, $\ldots$, $T'_n$ represent the combined variables with total sample size $n = n_1 + n_0$. To test the null hypothesis that there is no association between the outcome and the covariates, the score statistic is

$$S = nU' \Gamma^{-1} U$$

where $U_{m \times 1} = \frac{n_1}{\sqrt{n}} (\overline{T}_1 - \overline{T}_0)$ and $\Gamma_{m \times m} = \frac{n_0}{\sqrt{n}} \left( \frac{1}{2} \sum_{i=1}^{n_1} T_i T_i' - \overline{T} \overline{T}' \right)$ with

$$\overline{T} = \frac{1}{n} \sum_{i=1}^{n_1} T_i / n_1, \overline{T}' = \frac{1}{n} \sum_{i=1}^{n_0} T_i / n_0,$$

and $\overline{T} = \sum_{i=1}^{n} T_i / n$ being the vectors of sample averages.

In GWAS, let $Z_j = (Z_{j1}, \ldots, Z_{jm})$ and $X_j = (X_{j1}, \ldots, X_{jm})$ be samples of genotype values for the $j$th marker in cases and controls, respectively. That is, the vector of random genotype of marker $j$ is $G_j = (G_{j1}, \ldots, G_{jm})$. In general, the elements of $T^i_1$ and $T^i_0$ are functions of random genotypes corresponding to the form of the logistic regression model. Particularly, for a single-marker model (2.6) of marker $j$, $j = 1, \ldots, L$, there is $m = 1$ covariate such that $T^i_1 = (Z_{j1})$, $T^i_0 = (X_{j1})$ and $T = G_j$. So the score test is

$$S_j = \frac{n}{2} \left( \frac{2r(1-r)^2}{\overline{Z}_j^2 + (1-r)^2 \overline{X}_j^2 - (1-r)\overline{Z}_j \overline{X}_j} \right)^2.$$  

where $r = n_1/n$, $\overline{Z}_j = \sum_{i=1}^{n_1} Z_{ji}/n_1$, $\overline{X}_j = \sum_{i=1}^{n_0} X_{ji}/n_0$, $\overline{Z}_j^2 = \sum_{i=1}^{n_1} Z_{ji}^2/n_1$ and $\overline{X}_j^2 = \sum_{i=1}^{n_0} X_{ji}^2/n_0$. For a two-marker model (2.7) of markers $j$ and $k$, there are $m = 3$ covariate such that $T^i_1 = (Z_{j1}, Z_{k1}, Z_{jk})$, $T^i_0 = (X_{j1}, X_{k1}, X_{jk})$, and $T_{3 \times 2} = (G_{j1}, G_{k1}, G_{jk})'$, where $'$ denotes the element-wise cross-product of two vectors. Thus the score test becomes

Stat Sin. Author manuscript; available in PMC 2013 August 15.
\[ S_{jk} = n U_{jk} (\Gamma_{jk})^{-1} U_{jk}, \quad (3.2) \]

where

\[ U_{jk} = \frac{n_1 n_0}{n} \left( \bar{Z}_j - \bar{X}_j, \bar{Z}_k - \bar{X}_k, \bar{Z}_j \bar{Z}_k - \bar{X}_j \bar{X}_k \right), \]

\[ \Gamma_{jk} = \frac{n_1 n_0}{n} \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{pmatrix}, \quad (3.3) \]

with \( \Gamma_{jk} \) being symmetric and

\[ \gamma_{11} = \frac{n_1 Z_j^2 + n_0 X_j^2}{n} \left( \frac{n_1 Z_j + n_0 X_j}{n} \right)^2, \]
\[ \gamma_{12} = \frac{n_1 Z_j Z_k + n_0 X_j X_k}{n} \left( \frac{n_1 Z_j + n_0 X_j}{n} \right) \left( \frac{n_1 Z_k + n_0 X_k}{n} \right), \]
\[ \gamma_{13} = \frac{n_1 Z_j^2 + n_0 X_j^2}{n} \left( \frac{n_1 Z_j + n_0 X_j}{n} \right) \left( \frac{n_1 Z_k + n_0 X_k}{n} \right), \]
\[ \gamma_{22} = \frac{n_1 Z_j^2 + n_0 X_j^2}{n} \left( \frac{n_1 Z_k + n_0 X_k}{n} \right)^2, \]
\[ \gamma_{23} = \frac{n_1 Z_j Z_k + n_0 X_j X_k}{n} \left( \frac{n_1 Z_j + n_0 X_j}{n} \right) \left( \frac{n_1 Z_k + n_0 X_k}{n} \right), \]
\[ \gamma_{33} = \frac{n_1 Z_j^2 + n_0 X_j^2}{n} \left( \frac{n_1 Z_k + n_0 X_k}{n} \right)^2. \]

Note that \( \bar{Z}_j \bar{Z}_k = \sum_{i=1}^{n_1} Z_{ji} Z_{ki} / n_1, \bar{X}_j \bar{X}_k = \sum_{i=1}^{n_0} X_{ji} X_{ki} / n_0 \), the other averages functions of cross-product terms are analogously defined.

### 3.2. Asymptotic Distributions

In GWAS, the marker genotypes of individuals are not controllable but randomly observed. It is crucial to consider the genotype predictors as random variables. We apply the following generalization of Delta method to derive the null and the alternative distributions of the score test statistics, which are functions of random predictors. Let \( W_i = (W_{1i}, \ldots, W_{mi}), i = 1, \ldots, n \), be \( n \) independent and identically distributed random vectors of dimension \( m \). The corresponding mean vector is \( \theta = (\theta_1, \ldots, \theta_m) \) with \( \theta_s = E(W_{si}) \), and the covariance matrix is \( \Sigma = \text{Cov}(W_i) \) with \( (\Sigma)_{st} = \text{Cov}(W_{si}, W_{ti}) \), \( s, t = 1, \ldots, m \). Let \( W = (W_1, \ldots, W_m) \) be the vector of the sample means, i.e., \( \bar{W} = \frac{1}{n} \sum_{i=1}^{n} W_{si} \). Consider a real valued function \( h(W) \) of \( W \), if \( \nabla \theta (\theta) \equiv \left( \frac{\partial h(\theta)}{\partial \theta_1}, \ldots, \frac{\partial h(\theta)}{\partial \theta_m} \right)' \neq 0 \),

\[ \sqrt{n} [h(\bar{W}) - h(\theta)] \xrightarrow{L} N \left( 0, \tau^2 \right), \quad (3.4) \]

where \( \tau^2 = [\nabla h(\theta)]' \Sigma [\nabla h(\theta)] \) and \( \xrightarrow{L} \) denotes convergence in law. If \( \nabla h(\theta) = 0 \),

Stat Sin. Author manuscript; available in PMC 2013 August 15.
\[ n \left[ h(\overline{W}) - h(\theta) \right] \overset{L}{\to} cX_d^2. \]  

where if \( A \equiv D^2h(\theta) \Sigma, \) with \( D^2h(\theta) = \frac{\partial^2}{\partial \theta^2} h(\theta) \) being the Hessian matrix of \( h(\theta), \) then

1. \( c = \frac{1}{d}, \quad d = \text{rank}(A), \) if \( A \) is idempotent,
2. \( c \approx \frac{\text{max}(\sigma_1^2)}{\text{max}(\sigma_2^2)}, \quad d \approx \frac{\text{max}(\sigma_1^2)}{\text{max}(\sigma_2^2)}, \) if \( A \) is not idempotent.

Furthermore, if \( \nabla h_1(\theta) \neq 0 \) and \( \nabla h_2(\theta) \neq 0, \)

\[ \text{Cov}(\sqrt{n}h_1(\overline{W}), \sqrt{n}h_2(\overline{W})) \overset{P}{\to} [\nabla h_1(\theta)]' \sum \nabla h_2(\theta). \]  

Clearly, the score tests in (3.1) and (3.2) are functions of the genotypic sample means. The distribution of \( \overline{W} \) can be derived from the genotypic distributions in cases or in controls determined by genetic models. For the score tests involved in each marker search method, the following sections specify the distribution of \( \overline{W} \) for the given markers involved in the model fittings. When the causative loci are not genotyped but markers 1 and 2 are in LD with them, by (2.5) we obtained the joint genotypic distribution \( p(g_A1, g_A2|D) \) in cases and \( p(g_A1, g_A2|\overline{D}) \) in controls. The distributions of relevant score tests then come after the mean vector and covariance matrix involving the associated markers 1 and 2:

\[ \theta_{A1:A2} = E(\overline{W}_{A1:A2}), \]
\[ \Sigma_{A1:A2} = Var(\overline{W}_{A1:A2}). \]

### 3.3. Marginal Search

#### 3.3.1. Asymptotic Distribution of Test Statistic—
The relevant tests and the corresponding distributions for marginal search are as follows. For the \( j \)-th single marker, \( j = 1, \ldots, L, \) let \( T_j \equiv \sqrt{\frac{n}{2}} h(\overline{W}_j) \) with \( \overline{W}_j = (\overline{Z}_j, \overline{X}_j, \overline{Z}_2, \overline{X}_2) \) being a vector of sample averages over cases and controls. Let \( W_j = (Z_j, X_j, Z_2, X_2) \) represent the random genotypic vector of any observation. For an associated marker \( j = 1 \) (similarly for \( j = 2 \)), \( Z_1 \) has the distribution \( p(g_1|D) = \Sigma g_2 p(g_1, g_2|D), \) and \( X_1 \) has the distribution \( p(g_1|\overline{D}) \) for the non-associated markers \( j = 3, \ldots, L, \) \( Z_j \) and \( X_j \) have the same distribution \( p(g_j). \) The means and variance matrices are gotten by \( \theta = E(\overline{W}_j) \) and \( \Sigma_j = Cov(\overline{W}_j), \) respectively. When \( n_1 = n_0, \) by (3.4),

\[ T_j \sim \sqrt{n/2} h(\theta_j) \overset{L}{\to} N(0, \tau_j^2), \]

where \( \tau_j^2 = |\nabla h(\theta_j)|' \sum_j |\nabla h(\theta_j)|. \)

To calculate the alternative distributions, note that \( T_1 \) and \( T_2 \) are correlated because the odds of disease is a function of both markers 1 and 2. The joint distribution of \( (T_1, T_2)' \) is asymptotically multivariate normal.
and is a part of the joint distribution in (3.8). Corresponding to the non-associated markers \( j = 3, \ldots, L \), the null distribution for marginal search is \( T_j \overset{L}{\sim} N(0, 1) \) by (3.4), or consistently, \( S_j \overset{L}{\sim} \chi^2_1 \) by (3.5). Further, by (3.6), the correlations between \( T_j, j = 3, \ldots, L \), and \( T_1 \) (or \( T_2 \)) are asymptotically

### 3.3.2. Power under Discovery Number Control

For the discovery number control, the power of detecting alternative model(s) in the top \( R \) most significantly fitted models is equal to the probability that an alternative model is better fitted than the \( R \)th (or in marginal search under power definition (A), the \((R - 1)\)th) best fitted null models. Specifically, when the number of discoveries is controlled by \( R \), the power of marginal search under definition (A) for detecting both associated markers 1 and 2 is

\[
P(S_1 \land S_2 \geq S_{(r)}) = \int \int P(S_{(r)} \leq t_1^2 \land t_2^2) \, dG(t_1, t_2),
\]

where \( S_1 \land S_2 = \min \{S_1, S_2\} \), \( r = L - 2 - R + 1 \), \( S_{(r)} \) is the \( r \)th smallest (or the \( R \)th largest) order statistics in the set \( \{S_j, j = 3, \ldots, L\} \), and \( G(t_1, t_2) \) is the cumulative distribution function (CDF) of \((T_1, T_2)^T\) in (3.7). Let \( G_1(\cdot) \) be the CDF of \( \chi^2_1 \). Then

\[
P(S_{(r)} \leq x) = G_1(x) \sum_{l=0}^{L-2-r} \binom{r+l-1}{l} (1-G_1(x))^l.
\]

To get the power of marginal search under definition (B) that either associated marker 1 or marker 2 is selected, we calculate the probability that either \( S_1 \) or \( S_2 \) is larger than the cutoff point: \( P(S_1 \lor S_2 \geq S_{(r)}) \), where \( S_1 \lor S_2 = \max \{S_1, S_2\} \).

### 3.3.3. Power under Bonferroni Control

Since the null distribution of a score statistic used in marginal search is \( \chi^2_1 \), the cutoff under the Bonferroni corrected type I error rate control is \( c = G_1^{-1}(1 - \alpha/L) \), where \( G_1^{-1} \) is the quantile function of \( \chi^2_1 \) and \( \alpha \) is the genome-wide significance level. Under power definition (A) or (B), the probability of finding both or either associated marker is \( P(S_1 \land S_2 \geq c) \) or \( P(S_1 \lor S_2 \geq c) \).

### 3.4. Exhaustive Search

#### 3.4.1. Asymptotic Distribution of Test Statistic

The relevant test distributions for exhaustive search are based on the score tests in (3.1) and (3.2). For the statistics involving associated markers 1 and 2, let \( T_{12} = \sqrt{n/2}h_{12}(\bar{W}_{12}) \), \( T_i = \sqrt{n/2}h_i(\bar{W}_{i}) \), \( i = 1, 2 \), where

\[
\bar{W}_{12} = \begin{pmatrix} \bar{Z}_1, \bar{X}_1, \bar{Z}_2, \bar{X}_2, Z_1Z_2, X_1X_2, Z_1^2, Z_2^2, X_1^2, X_2^2 \end{pmatrix}^{T} \nonumber \\
\frac{Z_1^2Z_2, X_1^2X_2, Z_1Z_2^2, X_1X_2^2, Z_1^2Z_2^2, X_1^2X_2^2}{Z_1^2Z_2^2, X_1^2X_2^2, Z_1Z_2^2, X_1X_2^2, Z_1^2Z_2^2, X_1^2X_2^2}.
\]

Stat Sin. Author manuscript; available in PMC 2013 August 15.
is a vector of sample averages over cases and controls. Let $W_{12}$ be the corresponding genotypic vector of one random observation, we have the mean vector $\hat{\Theta}_{12} = E(W_{12})$ and the variance matrix $\Sigma_{12} = Var(W_{12})$. Based on the asymptotic distribution results in (3.4) and (3.6), when $n_1 = n_0$,

$$(T_{12}, T_1, T_2) \sim \mu_{T_{12}, T_1, T_2} \xrightarrow{L} MVN(0, \tau_{T_{12}, T_1, T_2}),$$  (3.8)

where

$$\mu_{T_{12}, T_1, T_2} = \sqrt{2}(h_{12}(\hat{\Theta}_{12}), h_1(\hat{\Theta}_{12}), h_2(\hat{\Theta}_{12}))',$$

$$\tau_{T_{12}, T_1, T_2} = D'\Sigma_{12}D,$$

with $D = (\nabla h_{12}(\hat{\Theta}_{12}), \nabla h_1(\hat{\Theta}_{12}), \nabla h_2(\hat{\Theta}_{12}))$.

Because $S_{12}$ has 3 degrees of freedom, the convergence in (3.8) is relatively slower than that in (3.7). In the following we describe an approximation for the mean of $T_{12}$, in the case that sample size is small (e.g. $n_1 < 1000$) and genetic effect is weak (e.g. $\Theta < 0.2$ in model (2.2)). Note that if we consider an observed (and thus fixed) data design matrix $t = (g_1, g_2, g_1 * g_2)$ of a logistic regression in the form (2.7), where $g_j = (z_{j1}, ..., z_{jn_1}, x_{j1}, ..., x_{jn_0})'$ $j = 1$ or 2, it has been shown that [13]

$$S_{12} \sim \chi^2_{3, \delta_n(t)}$$

where

$$\delta_n(t) = n^t b \Gamma_{12}(t) b,$$

with $\Gamma_{12}(t)$ given in the form (3.3) and $b = (b_1, b_2, b_3)'$ being the vector of coefficients in (2.4). Now we define $\delta_n = \delta_n(E(T))$ in our set-up for random genotype $T = (G_1, G_2, G_1 * G_2)$ with $G_j = (Z_{j1}, ..., Z_{jn_1}, X_{j1}, ..., X_{jn_0})'$ $j = 1$ or 2. We can use a weighted Chi-square with one degree of freedom to approximate $S_{12}$, i.e. $S_{12} \Rightarrow n_k \chi^2_{1, \delta_n}$, solving the equations assuming equal mean and variance [9], i.e. $a_n (1 + \lambda_n) = 3 + \delta_n$ and $a_n^2 (2 + 4 \lambda_n) = 6 + 4 \delta_n$ we get $\lambda_n = \frac{\delta_n}{1 + \delta_n}$, $\lambda_n = 2t_n - 1 + \sqrt{4t_n^2 - 2t_n}$, and $a_n = \frac{\delta_n}{1 + \delta_n}$. So we can apply the following approximation for weak genetic effects:

$$E(T_{12}) \approx \sqrt{\delta_n \lambda_n}.$$

By (3.5), the score test $S_{jk}$ for model (2.7) of two non-associated markers $j$ and $k$, $3 \leq j < k \leq L$, has an asymptotic distribution

$$S_{jk} \xrightarrow{L} \chi^2_3.$$  (3.9)
Let $S_{kj}$ denote the score test for the extra terms in model (2.7) over model (2.6). The following is a useful decomposition

$$S_k = S_j + S_{kj}. \quad (3.10)$$

So the correlation between the score tests $S_{jk1}$ and $S_{jk2}$, sharing the same marker $j$, can be captured by $S_j$, while $S_{kj1}$ and $S_{kj2}$ can be treated independent. Furthermore, by (3.5), for $k \geq 3$,

$$S_{kj} \to \chi^2_L. \quad (3.11)$$

### 3.4.2. Power under Discovery Number Control

With the test distributions derived above, we can calculate the probability of identifying the whole associated genetic model through exhaustive search under power definition (A). Define sets $A_1 \equiv \{S_{ij}, i = 1, 2, j = 3, \ldots, L\}$ and $A_2 \equiv \{S_{jk}, 3 \leq j < k \leq L\}$. Let $S_{A_1[R]}$ denote the $R$th largest score tests in a set $A$. When controlling the false discovery number by $R$, the probability of detecting the associated marker-pair is

$$P\left(S_{12} \geq S_{A_1 \cup A_2[R]} \right) = \int \int \int P\left(\tilde{r}_{12}^2 \geq S_{A_1 \cup A_2[R]} \right) dG(t_{12}, t_1, t_2).$$

where $G(t_{12}, t_1, t_2)$ is the CDF of (3.8), $A_1' = [\tilde{r}_i^2 + S_{ji}, i = 1, 2, j = 3, \ldots, L]$ is from the decomposition (3.10), and

$$P\left(\tilde{r}_{12}^2 \geq S_{A_1 \cup A_2[R]} \right) = \sum_{r=0}^{R-1} \sum_{(r_1, r_2, r_3) \in S_r} P_1 P_2 P_3,$$

where

$$S_r = \{r_1, r_2, r_3\}; \sum r_i = r, 0 \leq r_1, r_2, r_3 \leq (L-2), 0 \leq r \leq N,$$

$$P_1 = \binom{L-2}{r_1} \left( 1 - G_1\left(\tilde{r}_{12}^2-r_1^2\right) \right)^{r_1} G_1\left(\tilde{r}_{12}^2-r_1^2\right)^{L-2-r_1},$$

$$P_2 = \binom{L-2}{r_2} \left( 1 - G_1\left(\tilde{r}_{12}^2-r_2^2\right) \right)^{r_2} G_1\left(\tilde{r}_{12}^2-r_2^2\right)^{L-2-r_2},$$

$$P_3 = \binom{N}{r_3} \left( 1 - G_3\left(\tilde{r}_{12}^2\right) \right)^{r_3} G_3\left(\tilde{r}_{12}^2\right)^{N-r_3},$$

$N = \binom{L-2}{2}$ is the number of variables in $S_2$, $G_1(\cdot)$ is the CDF of distribution (3.11), and $G_3(\cdot)$ is the CDF of distribution (3.9). With the same argument given in the literature [12], the test statistics within the sets $A^* \equiv \{S_{ji}, j = 3, \ldots, L\}$ and $S_2 \equiv \{S_{jk}, 2 \leq j < k \leq L\}$ can be treated as asymptotically independent as $L \to \infty$. 

Stat Sin. Author manuscript; available in PMC 2013 August 15.
To simplify the heavy computation needed for the above formula, we can use the following approximations of the integrands. Simulations (results not presented) illustrated that these approximations are pretty accurate in the context of Monte Carlo integration. Let \( m = 2(L - 2) + N \) be the total number of the elements in \( A_1' \cup A_2 \). \( Q \) denotes the quantile function of mixed distribution of these elements. For a given \((t_{12}, t_1, t_2)\), \( P \left( \frac{t_{12}^2}{Q} \geq S_{A_1' \cup A_2} \right) \) can be approximately replaced with

\[
I \left( \frac{t_{12}^2}{Q} \geq \frac{m - R + 0.5}{N} \right) \approx I \left[ \frac{t_{12}^2}{Q} > \text{Rth largest value in } A_3 \right],
\]

where \( I \{ E \} \) denotes the indicator function of event \( E \), and the set

\[
A_3 \{ Q_1 (r) + t_{12}^2, Q_1 (r) + t_{12}^2, Q_3 (r), r = 1, \ldots, R \},
\]

with \( Q_1 (r) = G_1^{-1} \left( 1 - \frac{r}{N} \right) \) and \( Q_3 (r) = G_3^{-1} \left( \frac{r}{N} - \frac{0.5}{N} \right) \).

According to power definition (B), the probability for exhaustive search to detect either associated marker is

\[
P \left( \max (|S_{12} \cup A_1| > S_{A_1}) \right) = 1 - \int \int \int P_{t_{12}, t_1, t_2} dG(t_{12}, t_1, t_2)
\]

where

\[
P_{t_{12}, t_1, t_2} = P \left( \max \left( \frac{t_{12}^2}{Q} \cup A_1' \right) \leq S_{A_1} \right)
\]

\[
= \int P \left( \frac{t_{12}^2}{Q} \cup A_1' \leq x \right) g_{G_{1}^{-1}(x)} (d) \, dx
\]

\[
= \int G_1 (x - t_{12}^2) G_1 (x - t_{12}^2)^{L - 2} g_{G_{1}^{-1}(x)} (d) \, dx,
\]

\( g_{3, (N-R+1)} (\cdot) \) is the PDF of the \((N - R + 1)\)th order statistics distribution with the following density function:

\[
g_{3, (N-R+1)} (x) = \frac{N!}{(N-R)! (R-1)!} G_3 (x)^{N-R} [1 - G_3 (x)]^{R-1} g_3 (x),
\]

\( G_3 (c) \) and \( g_3 (\cdot) \) are the CDF and PDF of the distribution of (3.9), respectively.

If \( R \) is neither too small nor too large, i.e., \( \frac{r}{N} \rightarrow c, 0 < c < 1 \), as \( N \rightarrow \infty \), we can use quantiles to replace the order statistics in order to simplify the calculation ([2] Chapter 4.6), i.e.

\[
S_{A_1} \rightarrow G_3^{-1} \left( \frac{N - R + 0.5}{N} \right) = Q.
\]

So for given \((t_{12}, t_1, t_2)\), we can approximate the integrand

\[
P_{t_{12}, t_1, t_2} \text{ with}
\]

\[
\]
3.4.3. Power under Bonferroni Control—Traditional type I error control for exhaustive search does not consider the models with one associated and one non-associated markers, so the null distribution is from non-associated two-marker models [6] [10]. Let $G^{-1}_3$ be the quantile functions of $\chi^2$ in (3.9). The cutoff is $c = G^{-1}_3 \left( 1 - \alpha / \binom{L}{2} \right)$. The probability of finding the whole associated genetic model under power definition (A) is

$$P(S_{12} \geq c) = \int P \left( t_{12}^2 \geq c \right) dG(t_{12}).$$

With a similar argument for the power under discovery number control, the probability of finding either associated marker under power definition (B) is

$$P(\max(|S_{12}| \cup A_1) \geq c) = 1 - \int \int \int P_{t_{12}, t_{11}, t_{22}}(c) dG(t_{12}, t_{11}, t_{22}),$$

where

$$P_{t_{12}, t_{11}, t_{22}}(c) = P \left( \max \left( \left\{ t_{12}^2 \right\} \cup A_1 \right) \leq c \right) = I \left[ t_{12}^2 \leq c \right] \left[ G_1 \left( c - t_{11}^2 \right) G_1 \left( c - t_{22}^2 \right) \right]^{L-2},$$

with $G_1(\cdot)$ being the CDF of distribution (3.11).

3.5. Forward Search

3.5.1. Asymptotic Distribution of Test Statistic—For forward search, first we derive the distributions of test statistics, which will be used to calculate the power of this search procedure. For the score tests involving the associated markers 1 and 2, let $T_{hj} = \sqrt{n} S_{hj}$, where $S_{hj}$ follows (3.10), $i = 1, 2, j = 3, \ldots, L$. We can rewrite

$$T_{hj} = \sqrt{n} h_{hj} (W_{12j}),$$

where

$$W_{12j} = \begin{pmatrix}
\overline{Z}_1, \overline{X}_1, \overline{Z}_2, \overline{X}_2, \overline{Z}_1 \overline{X}_1, \overline{Z}_2, \overline{X}_1, \overline{X}_2, \overline{Z}_1 Z_2, \overline{Z}_1 Z_2 X_1 X_2, \overline{Z}_1 Z_2 X_1, \overline{Z}_1 Z_2 X_2, \\
\overline{Z}_1 Z_2, \overline{X}_1 Z_2, \overline{Z}_2 Z_2, \overline{X}_2 Z_2, \overline{Z}_1 Z_2, \overline{Z}_1 Z_2 X_1 X_2, \overline{Z}_1 Z_2 X_1, \overline{Z}_1 Z_2 X_2, \overline{Z}_1 Z_2, \overline{Z}_1 Z_2 X_1 X_2, \\
\overline{Z}_1 Z_2, \overline{X}_1 Z_2, \overline{Z}_2 Z_2, \overline{X}_2 Z_2, \overline{Z}_1 Z_2, \overline{Z}_1 Z_2 X_1 X_2, \overline{Z}_1 Z_2 X_1, \overline{Z}_1 Z_2 X_2, \overline{Z}_1 Z_2, \overline{Z}_1 Z_2 X_1 X_2,
\end{pmatrix}$$

is a vector of sample averages over cases and controls. Let $W_{12j}$ be the corresponding random genotypic vector of any observation. The mean vector is $\theta_{12} = E(W_{12j})$ and the
variance matrix is $\Sigma_{12j} = \text{Var}(W_{12j})$. Following (3.4) and (3.6), we have the asymptotic joint distribution

$$
\begin{pmatrix}
T_1, T_2, T_{1|j}, T_{2|j}
\end{pmatrix} - \mu_{T_1, T_2, T_{1|j}, T_{2|j}} \xrightarrow{L} \text{MVN}\left(0, \tau_{T_1, T_2, T_{1|j}, T_{2|j}}\right),
$$

where

$$
\mu_{T_1, T_2, T_{1|j}, T_{2|j}} = \sqrt{\frac{12}{n}}\left(h_1(\theta_{12j}), h_2(\theta_{12j}), h_{1|j}(\theta_{12j}), h_{2|j}(\theta_{12j})\right),
\tau_{T_1, T_2, T_{1|j}, T_{2|j}} = D \Sigma_{12j} D,
$$

with $D = (\nabla h_1(\theta_{12j}), \nabla h_2(\theta_{12j}), \nabla h_{1|j}(\theta_{12j}), \nabla h_{2|j}(\theta_{12j}))$. Through calculation [11], we have $\nabla h_{1|j}(\theta_{12j}) = \nabla h_j(\theta_{12j}), i = 1, 2$. By (3.4) and (3.6) it is clear that

$$
\text{Var}\left(T_i\right) = (\nabla h_i(\theta_{12j}))' \Sigma_{12j} \nabla h_i(\theta_{12j})
= (\nabla h_i(\theta_{12j}))' \Sigma_{12j} \nabla h_{1|j}(\theta_{12j}) = \text{Cov}\left(T_i, T_{1|j}\right).
$$

So $T_i$ and $T_{1|j}$ have correlation coefficient converging to 1. This explains why forward selection has similar power as marginal search for detecting either associated marker: if a genetic effect cannot standout in a marginal scan, it does not likely show strong signal in the following step either. Furthermore, $T_i$ and $T_{1|j}$ are asymptotically independent, i.e.

$$
\text{Cov}\left(T_j, T_{1|j}\right) \to 0, \quad i=1,2, j=3,\ldots,L.
$$

When comparing a model involving two incorrect markers $j$ and $k (3 \leq j < k \leq L)$ in (2.7) with a model for marker $j$ in (2.6), by (3.5), the corresponding score test statistic $S_{k|j}$ has the asymptotic chi-square distribution:

$$
S_{k|j} \xrightarrow{L} \chi_2^2.
$$

3.5.2. Power under Discovery Number Control—In the forward search procedure, we first apply marginal search to find the most significant marker among models in (2.6). Based on the selected marker, we then fit models in (2.7) in the second step to find the markers that have strong joint association. When controlling for $R$ total discoveries, under power definition (A) for finding the whole associated model, we need to calculate the probability that the forward search chooses marker 1 or 2 in the first step, and then picks the genetic model in the second step. Define $i^* \equiv \arg \max_{i=1,2} \left\{ S_i \right\}, A_i^* \equiv \left\{ S_{i,3}^*, \ldots, S_{i,L}^* \right\}$, as $L \to \infty$. The power can be written as
where \( G(t_{12}, t_1, t_2) \) is the CDF of \((T_{12}, T_1, T_2)\)' given in (3.8), \( S_{(L-2)} = \max_{j \geq 3} \{ S_j \} \), 
\( A_j = r_j^2 + \sum_{j=3}^{r} S_j \), \( j = 3, \ldots, L \) by the decomposition (3.10), and 
\[ P \left( r_{12}^2 > S_{(u-2)} \right) = \left( G_1 \left( r_{12}^2 \right) \right)^{L-2}, \]
\[ P \left( r_{12}^2 \geq S_{(u-1)} \right) = G_2(u)^{L-2} \sum_{l=0}^{R-1} \left( \begin{array}{c} R-l-1 \cr l \end{array} \right) \left[ 1-G_2(u) \right]^l, \]
where \( u = r_{12}^2 - r_{12}^2 \), \( r = L - 2 - R + 1 \). \( G_1(\cdot) \) is the CDF of \( \chi^2_1 \) for the distribution of \( S_j \), and \( G_2(\cdot) \) is the CDF of \( \chi^2_2 \) for the distribution of \( S_j^* \). \( r^* \) is fixed for an observed value \((t_1, t_2)'\) of the random vector \((T_1, T_2)'\), so it is easy to implement the power calculation with Monte Carlo integration.

Note that \( S_{(L-2)} \) and \( S_{(u-1)} \) are asymptotically independent. This is because \( \text{corr}(S_j, S_j^*) < 1 \) for each \( j \geq 3 \). So as \( L \to \infty \),
\[ P \left( \hat{j} \neq k^*; S_{j} = S_{(u-2)}, S_{k^*} = S_{(u-1)} \right) \to 1. \]

But when \( \hat{j} \neq k^* \), \( S_j^* \) and \( S_k^* | j^* \) are always independent.

When \( R \) and \( L \) are large, we can simplify the formula of \( P \left( r_{12}^2 \geq S_{(u-1)} \right) \) by approximating the \( R \)th largest variable in set \( \{ S_j^*, j = 3, \ldots, L \} \) with \( G_2^{-1} \left( 1 - \frac{1}{R} \right) \), where \( G_2^{-1} \) is the quantile function of \( S_j^* \). So we can approximately replace \( P \left( r_{12}^2 \geq S_{(u-1)} \right) \) with
\[ I \left\{ u \geq G_2^{-1} \left( 1 - \frac{R-n}{R} \right) \right\} \] to calculate the integration.

Under power definition (B) for finding either associated marker, the power of the forward search is the sum of \( P_A^i \): the probability of detecting marker 1 or 2 in the 1st step, and \( P_B^i \): the probability that step 1 fails but step 2 picks up at least one associated marker. When controlling for \( R \) total discovered models, it is straightforward that
\[
P_A = P\left(S_{\ell} > S_{(\ell-2)}\right) = \int \int \left(G_1 \left(t_1^2 \lor t_2^2\right)^{L-2}\right) dG(t_1, t_2),
\]

where \(G(t_1, t_2)\) is the CDF of the joint distribution of \((T_1, T_2)'\) given in (3.7). Define \(j^* \equiv \arg \max_{k \geq 3} \{S_k\}, \quad A_{j^*} \equiv \{S_{k|j^*}, k \geq 3, k \neq j^*\}.\) The second probability is

\[
P_{\tilde{h}} = P\left(S_1 \lor S_2 < j^* \cap (S_{1|j^*} \lor S_2|j^*) \geq S_{A_{j^*}|j^*}\right).
\]

For any \(k \geq 3, S_{k|j^*}\) and \(S_2\) are independent, so \(S_{k|j^*}\) and \(S_j^*\) are independent. By the results in (3.12) and (3.13), the distribution of \(S_{k|j^*}\) does not depend on \(j^*\). So, \(S_{k|j^*}\) has the same distribution of \(S_{j, f}, f = 3, \ldots, L\). Then

\[
P_{\tilde{h}} = \int P_{t_1, t_2} P_{t_1|t_2} dG\left(t_1, t_2, t_{1|j^*}, t_{2|j^*}\right).
\]

where \(G(t_1, t_2, t_{1|j}, t_{2|j})\) is the CDF of \((T_1, T_2, T_{1|j}, T_{2|j})'\) given in (3.12), and

\[
P_{t_1, t_2} = P\left(S_{(L-2)} \lor \left(t_1^2 \lor t_2^2\right)^{L-2}\right) = 1 - \left(G_1 \left(t_1^2 \lor t_2^2\right)^{L-2}\right),
\]

\[
P_{t_1, t_2|t_1} = P\left(t_{1|j^*}^2 \lor t_{2|j^*}^2 \geq S_{A_{j^*}|j^*}\right) = G_2\left(t_{1|j^*}^2 \lor t_{2|j^*}^2\right)^{L-3-r} \sum_{l=0}^{L-1-r} \left[1 - G_2\left(t_{1|j^*}^2 \lor t_{2|j^*}^2\right)^l\right],
\]

with \(r = L - 3 - R + 1, G_2(\cdot)\) is the CDF of \(S_{k|j^*}, k \geq 4,\) given in (3.13). We can approximate \(S_{A_{j^*}, [R]}\) through the quantile function \(G_2^{-1}(1 - \frac{\alpha}{L-2})\) to simplify the calculation of integration.

### 3.5.3. Power under Bonferroni Control

When we utilize the Bonferroni control in forward search, the first step selects the most significant single marker only if the test is larger than the cutoff \(c_1 = G_1^{-1}(1 - \alpha/L),\) where \(G_1^{-1}\) is the quantile function of \(\chi^2_1.\) In the second step, the null distribution is always \(\chi^2_1\) no matter which marker is first selected. Let the cutoff be \(c_2 = G_2^{-1}(1 - \alpha/(L-1)),\) where \(G_2^{-1}\) be the quantile function of \(\chi^2_1.\) For finding the true genetic model under power definition (A), the analytical power calculation is given by

\[
P\left(S_{\ell} > S_{(L-2)} \cap S_{\ell} > c_1 \cap S_{12} - S_{\ell} > c_2\right) = \int \int \left(G_1 \left(t_1^2 \lor t_2^2\right)^{L-2} \left\{t_1^2 \lor t_2^2\right\} > c_1 \left\{t_1^2 - t_2^2\right\} > c_2\right) dG(t_1, t_2),
\]

Similarly as the calculation under discovery number control, the power of forward search for finding either associated marker is \(P_A + P_B,\) where
As shown above, $T_i$ and $T_{ij}$ have large correlation coefficient converging to 1. Furthermore, 
thus $P_B \cong 0$.

3.6. Marginal-Exhaustive Two-Stage Search—We also study the power of a marginal-exhaustive two-stage method [6]. In the first stage for screening, it selects a set of single markers $I_1 \subseteq \{1, 2, \ldots, L\}$ with a liberal type I error-rate cutoff $c_1 = G^{-1}_1(1 - \alpha/2)$, where $G^{-1}_1$ is the quantile function of $\chi^2_1$. Then in the second stage, it applies the exhaustive search to the selected markers set. In the context of score test statistics, we adopt the approach by Marchini et al. [6] to define the statistic in the second stage. Specifically, the statistic is 
$S_{lm} = S_{lm} - 2c_2$, where $S_{lm}$ be the score test for the marker-pair $(l, m) \subseteq I_1$. We would select markers $l$ and $m$ if $S_{lm} \geq c_2 \equiv G^{-1}_3(1 - \alpha/2)$, where $G^{-1}_3$ is the quantile function of $\chi^2_3$. To find the associated genetic model, the first stage has to find both associated markers marginally. So under definition (A), the power of the two-stage search is 

$$P(S_1 \wedge S_2 \geq c_1 \wedge S'_{12} > c_2) = \iint \left( \left( r_1^2 \wedge r_2^2 \right) > c_1 \wedge \left( r_{12}^2 - 2c_1 \right) > c_2 \right) dG(t_{12}, t_1, t_2).$$

4. Results

4.1. Comparison between Analytical and Simulation Results

In order to demonstrate the accuracy of our analytical power calculation, we compared the power values from calculations with those from simulations. For the feasibility of simulation, we considered $L = 300$ candidate markers with MAF $p_j = 0.3, j = 1, \ldots, L, n_1 = 1000$ cases, $n_0 = 1000$ controls, base line effect $\alpha = 0.007$ and genotypic effect $\theta = 0.3$ for
genetic model in (2.2). These setups lead to a disease prevalence close to 0.01. Table 1 shows the power of marker search procedures under discovery number control. Table 2 shows the power under the Bonferroni corrected type I error rate control with the similar setups as Table 1 except that sample size \(n_1 = n_0 = 5000\) and genotypic effect \(\theta = 0.2\) in (2.2). These parameters were chosen to get the power values that are not too small nor too large, but are in a spectrum of values of practical interests. We simulated 1000 data sets and ran search procedures for each, the empirical power is the proportion of successful detections. In simulations, we used both score test and log likelihood ratio test for model comparisons. With various parameter setups, more comparisons between the simulated and the calculated power values can be found in the supplement. The consistent closeness between the analytical and simulation results demonstrates that our power calculation methods perform well, and that score test and LRT have similar performance for model selection.

4.2. Power Comparisons of Marker Search Methods

We applied the analytical power calculations to compare different marker search methods in a hypothetical GWAS that contains \(n_1 = 1000\) cases, \(n_0 = 1000\) controls, and \(L = 300,000\) candidate markers with MAF \(p_j = 0.3, j = 1, \ldots, L\). Assume the true genetic model is a logistic model of form (2.4) with the baseline effect \(a = \log (0.007)\). Let the main effects \(b_1 = b_2\) and the interaction effect \(b_3\) varies from \(-1\) to \(1\) by a step size of \(0.1\). Figures 1 and 2 show the 3-D plots of statistical power over a set of main and interaction effects under the discovery number control \(R = 20\) and the Bonferroni corrected type I error rate \(\alpha = 0.05\), respectively. As demonstrated by the “trenches”, marginal search and forward search will unavoidably fail when a disease susceptibility is controlled by interactions that show no marginally detectable signal. Exhaustive search, on the other hand, can avoid this problem and detect the full signal in 2-dimension as long as the effect size is large enough.

In order to contrast one marker search method with another, we subtracted the power values of one method from those of another method. The differences of the power values are plotted in heat maps. The red areas in Figures 3 and 4 represent negative values, indicating the first method in a pair of comparison has lower power; the green areas represent the opposite.

Figure 3 illustrates that the marginal search is better than the exhaustive search in the green areas, where \(b_3\) is small, and \(b_1\) and \(b_2\) are moderate. Relative to marginal search, forward search benefits from modeling interactions in finding certain true epistatic models, as shown in the red areas in the upper middle panel. However, to find either associated marker, forward search is uniformly less powerful than marginal search through the whole genetic model space, as shown in the whole lower middle panel that is white or green. Comparing with the forward search, the exhaustive search almost always performs similarly or better in finding the whole genetic model. With regard to the influence of the power types, power definition (B) favors marginal selection more than power definition (A), which is evidenced by the enlarged green areas in the second row.

Type I error control with the Bonferroni correction leads to notably different patterns of power comparisons. By examining the left and right columns of Figure 3 together with those of Figure 4, we can see exhaustive search increases its advantage so as to uniformly beat both marginal search and forward search in finding the true genetic model over the whole genetic model space. The Bonferroni control also raises the performance of forward search over marginal search in finding either associated marker, which is demonstrated by comparing the middle column of Figure 3 with that of Figure 4. For both control criteria, when we relax the control level of \(R\) or \(\alpha\), there is a trend that marginal search becomes relatively more powerful than exhaustive and forward searches. This is shown in the
supplement which contains more maps of comparisons under different genetic parameter set-ups.

### 4.3. Power Comparisons When Marginal Effects Are Fixed

As illustrated above, interaction effect is crucial for the statistical power of marker selection. However, because there is usually a lack of knowledge of interaction effects from real studies, it is meaningful to compare the search methods when the marginal association, possibly revealed from an interaction effect, is fixed. Assume the three genetic models in (2.1) – (2.3) have the same marginal association, which is represented by the heterozygote odds ratio $\lambda$ at each causative marker. When the values of $\lambda$ and the population disease prevalence $p(D)$ are fixed, we can calculate $\alpha$ and $\theta$ (letting $\theta_1 = \theta_2$ in model (2.1)).

Furthermore, it is interesting to study the influence of LD when the true disease-causing loci are not observed but are linked with genotyped markers. We adopted the squared correlation coefficient $r^2$ to measure LD [8] while deriving the analytical power calculation to find those linked markers. The assumptions for the fixed marginal effect and the constraints of LD follow Marchini et al. [6]. Technical details are given in the supplement.

Assume $\lambda = 1.5$, $p(D) = 0.01$, $n_1 = n_0 = 2000$, the MAF $p_j = 0.05$, 0.1, 0.2, and 0.5, the LD strength $r^2 = 0.5$, 0.7, and 1, the LD constraint $p(A_i|A_{-i}) = 1$ and $p(A_i|A_{-i}) = q$, $i = 1, 2$, where $A_{-i}$ is the disease disease-causing allele at the unobserved locus indexed by $-i$, $A_i$ is the disease allele at the genotyped locus of marker $i$, which is in LD with the causative locus $-i$. In general, power definition, genetic model, allele frequency, and sample size are influential to the relative performance of search strategies. Under the discovery number control $R = 5$, Figure 5 shows the power comparisons for finding the joint association. Here, the marginal search is the best for detecting model (2.1). For detecting models (2.2) and (2.3), the exhaustive search is the best while the marginal selection becomes the worst. The forward search is similar (for detecting model (2.1)) or better than marginal search (for detecting models (2.2) and (2.3)). This is not surprising because model (2.1) is additive in the log scale of odds, whereas models (2.2) and (2.3) are interactive, which accommodate those strategies facilitated by interaction effects. Figure 6 shows the power comparisons in finding either associated marker. Under this power definition, the marginal search interestingly outperforms the forward search for the interaction models (2.2) and (2.3). Furthermore, with sample size increasing, the exhaustive search increases its power faster than the marginal search and the forward search for the interaction models (2.2) and (2.3) (see the supplement for the figures of a smaller sample size $n_1 = n_0 = 1000$). This indicates that the exhaustive search has a more stringent requirement for sample size, but it will provide greater potential to detect small interactive effect when we do have enough observations.

Under the Bonferroni corrected type I error rate control, Figure 7 (or 8) shows the power comparisons for finding the joint association model (or either associated marker). The genetic parameter set-up is the same as that in Figures 5 and 6. To find the joint association model, Figure 7 shows that the exhaustive search is uniformly the best for all models, over all allele frequencies. Comparing with the forward search, the marginal search is a more favored method for model (2.1), but not for models (2.2) and (2.3). As shown in Figure 8, marginal search and forward search are very similar for finding either associated marker.

We calculated the statistical power of a marginal-exhaustive two-stage method for finding the joint association. The first stage screens single markers at a liberal type I error control level $\alpha_1$. Then the second stage carries out the exhaustive search within the selected set of markers, at a Bonferroni corrected type I error level $\alpha/(\alpha_1 L)/2$. This method is appealing.
for its potential of reducing the computational burden in the exhaustive search while remaining the high power. Figure 9 shows the comparisons among the power of the marginal search in finding either or both associated markers, the power of the exhaustive search in finding the joint association, and the power of the two-stage method in finding the joint association. The genetic parameters are the same as those for Figures 5 – 8. The two-stage method performs similarly as or even slightly better than exhaustive search. However, this result is valid only under the moderate marginal association that guarantees a high probability of picking the associated markers in the screening stage. As shown in Figure 2, it is possible, at least in theory, that the marginal association from particular interactions may totally vanish. In this case, the two-stage method will certainly not be able to surpass the exhaustive search. Note that we only used minutes of computational time to analytically calculate power for generating Figure 9, which reproduces the similar comparison patterns as that shown by heavy simulations (Figure 2 in Marchini et al. [6]).

5. Discussion

In this article the analytical power calculation derived for marker search strategies offers valuable tools for GWAS data analysis. In practice, power study itself is important for designing experiments and choosing proper analysis techniques based on prior knowledge. However, the underlying genetic model, on which any power study has to be based, is often not known. Hence, it is important to evaluate power efficiently so that researchers could study a wide range of possibilities. Our analytical power calculation can significantly reduce computational burden of simulations, which researchers had to rely on before. Our R package will enable researchers to investigate the statistical power from their proposed sample sizes and compare the performance of different multi-marker analysis strategies to analyze GWAS data. It can be used to explore a large genetic model space with flexible assumptions of joint genetic effects with or without interactions.

It is generally believed that complex diseases are jointly influenced by multiple markers with potential interactions. Therefore, it is more intuitive to view the underlying genetic model as one multivariate model, instead of a group of single-marker models. Our assumptions for the underlying genetic models are more in line with this view by considering joint effects and interactions instead of the oversimplified single-marker models in the literature [4]. The genetic model studied here can be extended to more complex models with higher order interactions. Moreover, for the multivariate joint marker models, it is important to understand the mechanisms of how the joint genetic signal can be picked through various statistical model fitting and comparing procedures. Our study in the distributions and the correlation structures among test statistics provides insights into these questions.

Comparing the results obtained here for the binary trait with those for the quantitative trait [12], the patterns of the relative performances of the marginal search, exhaustive search, and forward search have both common and distinguishable characters for the two types of traits. For the common features, the strength of interaction effect is a key factor to the performance. Both marginal and forward searches fail when certain interactions show no marginal signal. In general, strong interaction effect favors the exhaustive and the forward searches, which are carried in a higher dimensional model space, to find the joint genetic association.

For the distinction, the exhaustive search is relatively more powerful for binary traits than for quantitative traits, especially in finding the joint association. The forward search is relatively less powerful for binary traits than for quantitative traits, especially in finding at least one associated marker. The underlying reason revealed by the analytical study is the
correlation between the test statistics of single-marker model fittings in (2.6) and the test statistics of the extra terms in (2.7) over (2.6). The correlation is stronger among score tests for binary traits than that among F test for quantitative traits. Thus compared with quantitative traits, in the exhaustive search for binary traits, finding one associated marker can better increase the chance to find the other associated marker. Nevertheless, in the forward search, if the selected marker in the first step is non-associated, it is much harder in the second step to discover the associated markers for binary traits than for quantitative traits.

Another difference is the symmetry of the influence of the genetic effects on the power. If the total genetic effect can be represented by \( b_1g_1 + b_2g_2 + b_3g_1g_2 \), the search power is the same for quantitative traits when the main effects \( b_1 = b_2 = 0 \) and the interaction effect \( b_3 = \pm c \). This is because the underlying quantitative trait genetic model is linear [12], and the regression models have the same goodness-of-fit for signals with opposite directions but the same magnitude. However, this is no longer true for binary traits because the total genetic effects with the same magnitude but opposite directions lead to different disease risks as shown in (2.4). This is why the heat maps in Figures 3–4 are not symmetric.

Statistical significance control is crucial to the statistical power of detecting genetic signals. We studied two different types of controls: the total discovery number \( R \)-control, which is related to false discovery number or false discovery proportion control, and the type I error rate \( \alpha \)-control with the Bonferroni adjustment. Interestingly these two types of controls grant different marker search strategies with distinguishable relative merit. First, the \( \alpha \)-control is more stringent than the \( R \)-control, and the \( R \)-control leads to higher statistical power. So even if we do not have many “significant” results in the sense of the Bonferroni control, we can still include the top ranked genetic variations into the validation stage, as long as resources permit. Second, because GWAS are mostly used to screen for candidate genes, the type I error rate control is not the number one aim. \( R \)-control is more commonly adopted in phased designs by researchers who want to control the number of markers for the follow-up validation study. Third, the Bonferroni corrected \( \alpha \)-control provides the exhaustive search more power than the marginal search, especially for finding the joint associations. So we expect more joint associations to be found when applying the 2-dimensional exhaustive scan in real GWAS data analysis. Lastly, for finding either associated marker, \( R \)-control makes forward selection less favored than marginal search, whereas the \( \alpha \)-control makes the two methods very similar.

The widely applied Bonferroni control procedure provides an intuitively simple rule for model selection. Nevertheless, it might not provide accurate type I error rate control for the whole model selection process. For example, in the exhaustive search, if the alternative is defined as the joint association [6][3][10][1], the models with partially associated markers should contribute to the null distribution. However, the Bonferroni correction procedure only applies \( \chi^2 \) as the null distribution, which ignores those “wrong” models. As for the forward selection, applying type I error rate control in each step does not necessarily lead to the overall family-wise type I error rate control.

Through the analytical study, one can see that the non-associated markers have little influence on the power of marker search when the numbers of cases and controls are large. This is because the non-associated markers correspond to the null distributions. As shown in the Methods section, these null distributions are asymptotically chi-square with degrees of freedom 1, 2, or 3, no matter what the genetic parameters of the non-associated markers are. It does not affect the results much whether we assume the homogenous or random MAFs for these non-associated markers. The factors of the associated markers and the LD are more
important. The R package \textit{markerSearchPower} allows flexible assumptions for these genetic parameters.

The association signals from GWAS do not imply causality because they only report the statistical association by evidencing the discrepancy of the genotype distributions between cases and controls. This study focuses on the ability of the logistic model fitting to find the difference of distributions between cases and control under the assumption of homogenous population.

Acknowledgments

We are grateful to Yale University Biomedical High Performance Computing Center and WPI Computing and Communications Center for computation support.

This work is supported in part by NIH grant GM 59507 and NSF grant DMS 0714817.

References

Figure 1.
3-D plots of statistical power under the discovery number control over genetic effect space. Three model selection methods: marginal search in the left column, exhaustive search in the middle column and forward search in the right column. Two definitions of power: (A) detecting the joint association (or both associated markers in marginal search) in row 1, and (B) detecting either associated marker in row 2. The genetic models follows a logistic model with the main effect $b_1 = b_2$ and the epistatic effect $b_3$ both varying from $-1$ to $1$. The MAF $p_j = 0.3, j = 1, \ldots, L$. The total discovery number $R$ is set to be 10.
Figure 2.
3-D plots of statistical power under the Bonferroni control over genetic effect space. Three model selection methods: marginal search in the left column, exhaustive search in the middle column and forward search in the right column. Two definitions of power: (A) detecting the joint association (or both associated markers in marginal search) in row 1, and (B) detecting either associated marker in row 2. The genetic models follows a logistic model with the main effect $b_1 = b_2$ and the epistatic effect $b_3$ both varying from $-1$ to $1$. The MAF $p_j = 0.3$, $j = 1, \ldots, L$. The genome-wide significance level $\alpha$ is set to be $0.05$. 
Figure 3.
Power differences with varying main and interaction effects with the discovery number $R = 10$. Row 1 illustrates power definition (A), row 2 illustrates power definition (B). Left column: marginal search vs. exhaustive search; middle column: marginal search vs. forward search; right column: forward search vs. exhaustive search. Green areas indicate positive values of difference, red areas indicate negative values. The main effect $b_1 = b_2$ and the epistatic effect $b_3$ both vary from $-1$ to $1$, the allele frequency $p_j = 0.3$, $j = 1, \ldots, L$. 
Figure 4.
Power differences with varying main and interaction effects with the Bonferroni type I error rate $\alpha = 0.05$. Power definition (A) is applied in row 1 and (B) is used in row 2. Left column: marginal search vs. exhaustive search; middle column: marginal search vs. forward search; right column: forward search vs. exhaustive search. Green areas indicate positive values of difference, red areas indicate negative values. The main effects $b_1 = b_2$ varies from $-1$ to $1$, the epistatic effect $b_3$ varies from $-1$ to $1$, the allele frequency $p_j = 0.3$, $j = 1, \ldots, L$. 

Stat Sin. Author manuscript; available in PMC 2013 August 15.
Figure 5.
Power of finding the joint association with the discovery number $R = 5$. Green lines, marginal search; blue dashed lines, exhaustive search; red dot-dashed lines, forward search. The marginal odds ratio at both loci is 1.5, disease prevalence is 0.01, case and control numbers are both 2000. Columns of panels show genetic Models 1, 2, 3, respectively; rows show LD strength $r^2 = 0.5, 0.7$ and 1. The minor allele frequencies are 0.05, 0.1, 0.2 and 0.5 on the x-axis of each penal.
Figure 6.
Power of finding either associated marker with the discovery number $R = 5$. Green lines, marginal search; blue dashed lines, exhaustive search; red dot-dashed lines, forward search. The marginal odds ratio at both loci is 1.5, disease prevalence is 0.01, case and control numbers are both 2000. Columns of panels show genetic Models 1, 2, 3, respectively; rows show LD strength $r^2 = 0.5, 0.7$ and 1. The minor allele frequencies are 0.05, 0.1, 0.2 and 0.5 on the x-axis of each panel.
Figure 7.
Power of finding the joint association with the Bonferroni type I error rate $\alpha = 0.05$. Green lines, marginal search; blue dashed lines, exhaustive search; red dot-dashed lines, forward search. The marginal odds ratio at both loci is 1.5, disease prevalence is 0.01, case and control numbers are both 2000. Columns of panels show genetic Models 1, 2, 3, respectively; rows show LD strength $r^2 = 0.5, 0.7$ and 1. The minor allele frequencies are 0.05, 0.1, 0.2 and 0.5 on the x-axis of each penal.
Figure 8.
Power of finding either associated marker with the Bonferroni type I error rate $\alpha = 0.05$. Green lines, marginal search; blue dashed lines, exhaustive search; red dot-dashed lines, forward search. The marginal odds ratio at both loci is 1.5, disease prevalence is 0.01, case and control numbers are both 2000. Columns of panels show genetic Models 1, 2, 3, respectively; rows show LD strength $r^2 = 0.5, 0.7$ and 1. The minor allele frequencies are 0.05, 0.1, 0.2 and 0.5 on the x-axis of each penal.
Figure 9.
Power of four marker search methods controlled by the Bonferroni type I error rate $\alpha = 0.05$. Red dashed lines, marginal search for either association; green lines, marginal search for joint association; dark blue long-dashed lines, exhaustive search for joint association; light blue dot-dashed lines, two-stage search for joint association. The marginal odds ratio at both loci is 1.5, disease prevalence is 0.01, case and control numbers are both 2000. Columns of panels show genetic Models 1, 2, 3, respectively; rows show LD strength $r^2 = 0.5$, 0.7 and 1. The minor allele frequencies are 0.05, 0.1, 0.2 and 0.5 on the x-axis of each penal.
Under the control of the discovery number $R$, the comparisons of the power values between simulations (based on score test and log-likelihood ratio test) and analytical calculations (based on score test). Power definitions (A) and (B) are considered. $\alpha = 0.007, \theta = 0.3$.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Source</th>
<th>$R = 1^*$</th>
<th>$R = 5$</th>
<th>$R = 10$</th>
<th>$R = 15$</th>
<th>$R = 20$</th>
<th>$R = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Power definition (A)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal search</td>
<td>Score Simu.</td>
<td>0.14</td>
<td>0.38</td>
<td>0.51</td>
<td>0.58</td>
<td>0.64</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.14</td>
<td>0.38</td>
<td>0.51</td>
<td>0.58</td>
<td>0.64</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.15</td>
<td>0.38</td>
<td>0.50</td>
<td>0.58</td>
<td>0.65</td>
<td>0.72</td>
</tr>
<tr>
<td>Exhaustive search</td>
<td>Score Simu.</td>
<td>0.32</td>
<td>0.50</td>
<td>0.57</td>
<td>0.61</td>
<td>0.65</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.33</td>
<td>0.52</td>
<td>0.61</td>
<td>0.64</td>
<td>0.69</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.32</td>
<td>0.50</td>
<td>0.59</td>
<td>0.63</td>
<td>0.65</td>
<td>0.69</td>
</tr>
<tr>
<td>Forward search</td>
<td>Score Simu.</td>
<td>0.28</td>
<td>0.42</td>
<td>0.48</td>
<td>0.50</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.29</td>
<td>0.43</td>
<td>0.48</td>
<td>0.50</td>
<td>0.53</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.28</td>
<td>0.44</td>
<td>0.48</td>
<td>0.51</td>
<td>0.52</td>
<td>0.54</td>
</tr>
<tr>
<td><strong>Power definition (B)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal search</td>
<td>Score Simu.</td>
<td>0.55</td>
<td>0.83</td>
<td>0.90</td>
<td>0.92</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.55</td>
<td>0.83</td>
<td>0.90</td>
<td>0.92</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.55</td>
<td>0.83</td>
<td>0.91</td>
<td>0.94</td>
<td>0.95</td>
<td>0.98</td>
</tr>
<tr>
<td>Exhaustive search</td>
<td>Score Simu.</td>
<td>0.57</td>
<td>0.80</td>
<td>0.87</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.59</td>
<td>0.81</td>
<td>0.87</td>
<td>0.81</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.62</td>
<td>0.77</td>
<td>0.84</td>
<td>0.88</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td>Forward search</td>
<td>Score Simu.</td>
<td>0.64</td>
<td>0.76</td>
<td>0.83</td>
<td>0.87</td>
<td>0.91</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.64</td>
<td>0.76</td>
<td>0.83</td>
<td>0.87</td>
<td>0.91</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.64</td>
<td>0.74</td>
<td>0.82</td>
<td>0.88</td>
<td>0.90</td>
<td>0.94</td>
</tr>
</tbody>
</table>

(*$R = 2$ in marginal search under power definition (A)).
Under the Bonferroni corrected type I error with family-wise significance level $\alpha$, the comparisons of the power values between simulations (based on score test and log-likelihood ratio test) and analytical calculations (based on score test). Power definitions (A) and (B) are considered. $n_1 = n_2 = 5000$, $\alpha = 0.007$, $\theta = 0.2$.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Source</th>
<th>$\alpha = 0.01$</th>
<th>$\alpha = 0.05$</th>
<th>$\alpha = 0.10$</th>
<th>$\alpha = 0.15$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Power definition (A)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal search</td>
<td>Score Simu.</td>
<td>0.20</td>
<td>0.37</td>
<td>0.46</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.20</td>
<td>0.37</td>
<td>0.46</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.19</td>
<td>0.34</td>
<td>0.42</td>
<td>0.48</td>
</tr>
<tr>
<td>Exhaustive search</td>
<td>Score Simu.</td>
<td>0.86</td>
<td>0.92</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.87</td>
<td>0.92</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.86</td>
<td>0.92</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>Forward search</td>
<td>Score Simu.</td>
<td>0.55</td>
<td>0.73</td>
<td>0.79</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.56</td>
<td>0.73</td>
<td>0.80</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.51</td>
<td>0.70</td>
<td>0.76</td>
<td>0.80</td>
</tr>
<tr>
<td><strong>Power definition (B)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marginal search</td>
<td>Score Simu.</td>
<td>0.68</td>
<td>0.83</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.68</td>
<td>0.83</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.66</td>
<td>0.82</td>
<td>0.87</td>
<td>0.89</td>
</tr>
<tr>
<td>Exhaustive search</td>
<td>Score Simu.</td>
<td>0.87</td>
<td>0.93</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.87</td>
<td>0.92</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.87</td>
<td>0.93</td>
<td>0.95</td>
<td>0.96</td>
</tr>
<tr>
<td>Forward search</td>
<td>Score Simu.</td>
<td>0.69</td>
<td>0.82</td>
<td>0.88</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>LRT Simu.</td>
<td>0.69</td>
<td>0.82</td>
<td>0.88</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>Score Calcu.</td>
<td>0.67</td>
<td>0.82</td>
<td>0.86</td>
<td>0.89</td>
</tr>
</tbody>
</table>