A Robust Method for Testing Association in Genome-Wide Association Studies

Zhongxue Chen[a]   Hon Keung Tony Ng[b]

[a]Biostatistics Epidemiology Research Design Core, Center for Clinical and Translational Sciences (CCTS), The University of Texas Health Science Center at Houston, Houston, Tex., and [b]Department of Statistical Science, Southern Methodist University, Dallas, Tex., USA

Abstract

In genetic association studies, due to the varying underlying genetic models, no single statistical test can be the most powerful test under all situations. Current studies show that if the underlying genetic models are known, trend-based tests, which outperform the classical Pearson χ^2 test, can be constructed. However, when the underlying genetic models are unknown, the χ^2 test is usually more robust than trend-based tests. In this paper, we propose a new association test based on a generalized genetic model, namely the generalized order-restricted relative risks model. Through a Monte Carlo simulation study, we show that the proposed association test is generally more powerful than the χ^2 test, and more robust than those trend-based tests. The proposed methodologies are also illustrated by some real SNP datasets.

Copyright © 2011 S. Karger AG, Basel

1. Introduction

In case-control genome-wide association (GWA) studies, hundreds of thousands of single nucleotide polymorphisms (SNPs) are tested to determine whether they are associated with the common disease of interest. If a SNP is in linkage disequilibrium with a disease locus, it will not be independent of the status of the disease. Although there may exist gene-gene or gene-environment interactions, the first and crucial step in GWA studies is to identify single SNPs that are associated with disease.

For a SNP with two alleles, A and a, which is assumed to be at risk, there are three genotypes: AA, Aa, and aa. Suppose that there are r cases and s controls in the study. In the r cases, there are r_0, r_1, and r_2 affected people with genotypes AA, Aa, and aa, respectively. There are s_0, s_1, and s_2 people with genotypes AA, Aa, and aa, respectively, in the s unaffected controls.

Testing whether there is an association between the genotype and the disease status is equivalent to testing the association in the 2×3 contingency table. Pearson's χ^2 test with 2 degrees of freedom (df) is one of the most commonly used statistical methods for testing the association in a contingency table. Note that for SNP data, it is reasonable to assume that the relative risk associated

Assoc. Prof. Hon Keung Tony Ng, PhD
Department of Statistical Science
Southern Methodist University
PO Box 750332, Dallas, TX 75275-0332 (USA)
Tel. +1 214 768 2465, E-Mail ngh @ mail.smu.edu

Table 1. SNP data in GWA studies

Genotype:	AA	Aa	aa	Total
Case	r_0	r_1	r_2	r
Control	s_0	s_1	s_2	s
Total	n_0	n_1	n_2	n

with Aa is between the risks associated with AA and aa. However, Pearson's χ^2 test does not utilize this feature of the order-restricted risks in the SNP data. On the contrary, the Cochran-Armitage trend test (CATT) was designed to incorporate the trend to increase the detecting power. The CATT with appropriate scores can be more powerful than Pearson's χ^2 test if the underlying genetic model is known [1–4]. In general, the scores used in the CATT for SNP data are $(0, x, 1)$, where x is a number between 0 and 1 and the optimal value of x depends on the true underlying genetic model. For instance, if the genetic models are recessive, additive/multiplicative (log additive), and dominant, the optimal values of x in the CATTs are 0, 0.5 and 1, respectively. The CATTs with optimal scores have been shown to be more powerful than Pearson's χ^2 test, provided that the underlying genetic model is known [4].

However, the genetic model is usually unknown in practice and the CATT with a non-optimal score may perform poorly. In other words, the CATT is not as robust as the χ^2 test, and it is sensitive to the departure of assumed genetic models. To increase the robustness of CATT, several trend-based methods have been proposed for the situations where the underlying genetic models are unknown [5–17]. For instance, the maxmin efficiency robust test (MERT) by Gastwirth [18, 19], and the maximum of the three optimal CATTs under recessive, additive, and dominant models (MAX3) have been studied [7]. Zheng and Ng [16] also proposed a two-phase procedure (GMS) with the selection of the genetic model based on the data in the first stage and then used the optimal score based on the chosen model for the CATT in the second stage. Although the above CATT-based methods have been shown to be robust compare to the CATT, there are some limitations for these methods when the analytic null distributions are either unavailable or too complicated. Consequently, Monte Carlo or numerical methods are required to compute the p values of these test procedures.

This paper is organized as follows. First, in Section 2.1, we propose a generalized genetic model for the SNP data, namely a generalized *order-restricted relative risks*

(ORRR) model, in which we assume that the two relative risks are monotonically increasing or decreasing. This ORRR model covers a wide range of ideal models. For instance, the recessive, additive, multiplicative and dominant models are special cases of the ORRR model. Then, we propose a statistical test based on the ORRR model in Section 2.2. Moreover, a restricted likelihood ratio test under the ORRR model is also considered in Section 2.2. Since the new test uses the order-restricted property of the relative risks, it is expected to be more powerful than the χ^2 test under many situations in general. On the other hand, unlike the CATT, the proposed test does not assume a specific genetic model; it is not sensitive to the misspecification of the underlying genetic models and therefore is more robust than CATT. In Section 3, a Monte Carlo simulation study is used to study the performance of the proposed procedure. We show that our proposed method is more robust than the existing methods and has decent power properties. The proposed methodologies are illustrated using some real SNP data in Section 4. Conclusions are provided in Section 5.

2. Proposed Test Procedure

2.1. A Generalized Genetic Model and Existing Methods

Table 1 gives the data structure of a case-control GWA study. The relative risks of genotypes Aa and aa to AA are defined as:

$$\begin{cases} \lambda_1 = \Pr(case|Aa) / \Pr(case|AA) \\ \lambda_2 = \Pr(case|aa) / \Pr(case|AA) \end{cases} \quad (1)$$

For many genetic models, we can reasonably assume $\Pr(case \mid Aa)$, the disease risk associated with genotype Aa, is between $\Pr(case \mid AA)$ and $\Pr(case \mid aa)$. Specifically, if a is the at-risk allele, the relative risks satisfy: $\lambda_1 \geq 1$ and $\lambda_2 \geq \lambda_1$ with at least one of the inequalities being strictly greater. If A is the at-risk allele, the relative risks satisfy $1 \geq \lambda_1$ and $\lambda_1 \geq \lambda_2$ where at least one of the inequalities is strict. The monotonicity of the relative risks is also known as order-restricted relative risks. Here, a genetic model with order-restricted relative risks is called a generalized ORRR model. We can see that the aforementioned ideal models (assuming the at-risk allele is a), that is, recessive ($\lambda_1 = 1, \lambda_2 > \lambda_1$), additive ($\lambda_1 = (1 + \lambda_2)/2$), multiplicative ($\lambda_2 > \lambda_1^2$), and dominant ($\lambda_1 = \lambda_2 > 1$), are all special cases of the generalized ORRR model.

As mentioned in Section 1, in addition to the χ^2 test, existing statistical test procedures, including CATT, MAX3, GMS and MERT, for the null hypothesis that there is no association between disease and the genotype are also considered here. The CATT statistic can be written as [10]:

$$Z_x = \frac{n^{1/2} \sum_{i=0}^{2} x_i (sr_i - rs_i)}{\left\{ rs \left[n \sum_{i=0}^{2} x_i^2 n_i - \left(\sum_{i=0}^{2} x_i n_i \right)^2 \right] \right\}^{1/2}}, \quad \text{where } (x_0, x_1, x_2) = (0, x, 1).$$

The statistic for MAX3 is [7]:

$$MAX3 = MAX\{|Z_0|, |Z_{1/2}|, |Z_1|\}.$$

The statistic for GMS is [16, 17]:

$GMS =$
$Z_0 I(Z_{1/2} > 0) I(Z_{HWDTT} > c) + Z_{1/2} I(Z_{1/2} > 0) I(|Z_{HWDTT}| < c)$
$+ Z_1 I(Z_{1/2} > 0) I(Z_{HWDTT} < -c) - Z_1 I(Z_{1/2} \leq 0) I(Z_{HWDTT} > c)$
$+ Z_{1/2} I(Z_{1/2} \leq 0) I(|Z_{HWDTT}| \leq c) - Z_0 I(Z_{1/2} \leq 0) I(Z_{HWDTT} < -c),$

where I is the indicator function, and the Hardy-Weinberg disequilibrium trend test (HWDTT) statistic is given by [20]:

$$Z_{HWDTT} = \frac{(rs/n)^{1/2}(\hat{\Delta}_P - \hat{\Delta}_Q)}{\{1 - n_2/n - n_1/(2n)\}\{n_2/n + n_1/(2n)\}},$$

$\hat{\Delta}_P = r_2/r - (r_2/r + r_1/(2r))^2$, $\hat{\Delta}_Q = s_2/s - (s_2/s + s_1/(2s))^2$, and c is a constant and usually chosen as 1.645.
The statistic for MERT is [19]:

$MERT = (Z_0 + Z_1)/\{2(1 + \hat{\rho}_{01})\}^{1/2},$

where $\hat{\rho}_{01} = (n_0 n_2)^{1/2}/\{(n_0 + n_1)(n_1 + n_2)\}^{1/2}$.

2.2. The Proposed Test

The proposed test is designed to detect the alternative hypothesis that the underlying genetic model belongs to the generalized ORRR model. Suppose the allele frequencies for AA, Aa, and aa are p_0, p_1, p_2 for case and q_0, q_1, and q_2 for control, respectively. Under the null hypothesis that there is no association between disease and the genotype, we have $p_0 = q_0$, $p_1 = q_1$, and $p_2 = q_2$.

Equation 1 can be expressed as

$$\begin{cases} \lambda_1 = \dfrac{p_1 \Pr(AA)}{p_0 \Pr(Aa)} \\ \lambda_2 = \dfrac{p_2 \Pr(AA)}{p_0 \Pr(aa)} \end{cases} \tag{2}$$

where $\Pr(AA) = \Pr(AA \mid \text{case})\Pr(\text{case}) + \Pr(AA \mid \text{control})\Pr(\text{control}) = kp_0 + (1 - k)q_0$ and k is the disease prevalence. Similarly we have $\Pr(Aa) = kp_1 + (1 - k)q_1$ and $\Pr(AA) = kp_2 + (1 - k)q_2$.
Then, equation 1 can be written as

$$\begin{cases} \lambda_1 = \dfrac{p_1(kp_0 + (1-k)q_0)}{p_0(kp_1 + (1-k)q_1)} \\ \lambda_2 = \dfrac{p_2(kp_0 + (1-k)q_0)}{p_0(kp_2 + (1-k)q_2)} \end{cases} \tag{3}$$

Assuming the at-risk allele is a, the alternative hypothesis we want to test is based on the generalized ORRR genetic model for which the relative risks satisfy $\lambda_1 \geq 1$ and $\lambda_2 \geq \lambda_1$, with at least one of the inequalities being strict.
From equation 3, we can write $\lambda_1 \geq 1$ and $\lambda_2 \geq \lambda_1$ as

$$\begin{cases} p_1 q_0 \geq p_0 q_1 \\ p_2 q_1 \geq p_1 q_2 \end{cases} \tag{4}$$

Let

$$\begin{cases} T_1 = p_1 q_0 - p_0 q_1, \\ T_2 = p_2 q_1 - p_1 q_2, \end{cases}$$

then detecting the alternative hypothesis is equivalent to detecting that both T_1 and T_2 are non-negative and at least one of them is strictly greater than 0. Therefore, we propose a statistical test procedure based on T_1 and T_2.

Given the observed data presented in table 1, the sample estimates of T_1 and T_2 can be obtained as

$$\begin{cases} \hat{T}_1 = \dfrac{r_1 s_0 - r_0 s_1}{rs} \\ \hat{T}_2 = \dfrac{r_2 s_1 - r_1 s_2}{rs} \end{cases} \tag{5}$$

respectively. It can be shown that under the null hypothesis of no association (i.e. $p_i = q_i$, $i = 0, 1, 2$), the expected values of $\hat{T}_1$ and $\hat{T}_2$ are zeros (i.e. $E(\hat{T}_1) = E(\hat{T}_2) = 0$) and the variance-covariance matrix of $\hat{T}_1$ and $\hat{T}_2$ is

$$\sum = \begin{bmatrix} Var(\hat{T}_1) & Cov(\hat{T}_1, \hat{T}_2) \\ Cov(\hat{T}_1, \hat{T}_2) & Var(\hat{T}_2) \end{bmatrix},$$

where

$Var(\hat{T}_1) = p_0 p_1(2p_2 + n(p_0 + p_1))/rs,$
$Var(\hat{T}_2) = p_2 p_1(2p_0 + n(p_2 + p_1))/rs,$
$Cov(\hat{T}_1, \hat{T}_2) = p_0 p_1 p_2(2 - n)/rs.$

Since the variance-covariance matrix is unknown, it can be estimated by replacing the p_is by their consistent estimators, $\hat{p}_i = n_i/n$, $i = 0, 1, 2$. The estimated variance-covariance matrix can be expressed as

$$\hat{\Sigma} = \frac{n_1}{rsn^3}\begin{bmatrix} n_0(2n_2 + n(n_0 + n_1)) & n_0 n_2(2 - n) \\ n_0 n_2(2 - n) & n_2(2n_0 + n(n_2 + n_1)) \end{bmatrix}.$$

Because $\hat{\Sigma}$ is a positive definite square matrix, eigen-decomposition of $\hat{\Sigma}$ gives $\hat{\Sigma} = PDP'$, where D is a diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix},$$

d_1 and d_2 are the eigenvalues of $\hat{\Sigma}$, and the columns of P are the corresponding eigenvectors of $\hat{\Sigma}$ which satisfies $PP' = I$.
Let

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \hat{\Sigma}^{-\frac{1}{2}}\begin{bmatrix} \hat{T}_1 \\ \hat{T}_2 \end{bmatrix},$$

under the null hypothesis, Z_1 and Z_2 are asymptotically independent standard normally distributed random variables, where

$$\hat{\Sigma}^{-\frac{1}{2}} = P\begin{bmatrix} d_1^{-\frac{1}{2}} & 0 \\ 0 & d_2^{-\frac{1}{2}} \end{bmatrix}P'.$$

It can be shown that all elements in $\hat{\Sigma}^{-1/2}$ are non-negative. Therefore, under the alternative hypothesis, we would expect $E(Z_1)$ and $E(Z_2)$ to be non-negative where at least one of them is

strictly positive. Note that if the at-risk allele is A instead of a in table 1 (i.e. column 1 and column 3 are being switched), statistics Z_1 and Z_2 are still valid with $E(Z_1)$ and $E(Z_2)$ being non-positive and with at least one of them being strictly negative under the alternative hypothesis.

By taking the order-restricted property of the SNPs data into account, we consider the statistics

$$\begin{cases} C_1 = -2\log\big(\Phi(Z_1)\Phi(Z_2)\big) \\ C_2 = -2\log\big(\Phi(-Z_1)\Phi(-Z_2)\big) \end{cases}, \qquad (6)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The asymptotic distributions of C_1 and C_2 are as follows.

Theorem 1
Under the null hypothesis, C_1 and C_2 are asymptotically χ^2 distributed with 4 df.

The proposed test statistic is the maximum of C_1 and C_2 denoted as

$$W = \max\{C_1, C_2\}. \qquad (7)$$

It should be noted that C_1 and C_2 are not independent. However, the following asymptotic property can lead us to an approximation of the p value associated with statistic W.

Theorem 2
Under the null hypothesis of no association, the survival function of W is asymptotically bounded by

$$2\gamma - \gamma^2 \le \Pr(W > w) \le 2\gamma, \qquad (8)$$

where $\gamma = 1 - \chi_4^2(w)$ and $\chi_4^2(\cdot)$ is the cumulative distribution function of the χ^2 distribution with 4 df.

Theorem 2 suggests that we can estimate the p value of the test procedure based on W by 2γ, and with small γ, the approximation is very accurate.

Under the ORRR genetic models, some other statistical tests can also be applied. For instance, the restricted likelihood ratio test (RLRT) has been proposed to detect the association for the $2 \times k$ contingency table [21, 22]. For SNP data with $k = 3$, the RLRT statistic is:

$$RLRT = 2\left[\sum_{i=0}^{2} r_i \log\big(\hat{\pi}_i / \hat{p}\big) + \sum_{i=0}^{2} s_i \log\big((1 - \hat{\pi}_i)/(1 - \hat{p})\big)\right],$$

where $\hat{p} = r/n$, and $\hat{\pi}_i = r_i/n_i$, $i = 0, 1, 2$, are order-restricted MLEs satisfying $\hat{\pi}_0 \le \hat{\pi}_1 \le \hat{\pi}_2$ or $\hat{\pi}_0 \ge \hat{\pi}_1 \ge \hat{\pi}_2$.

Usually $\hat{\pi}_i$ is estimated using the pool adjacent violators algorithm (PAVA) [23], and the above statistic has a weighted χ^2 distribution ($\bar{\chi}^2$) under the null hypothesis [24]. Its p value is $\Pr(RLRT > c) = w_1 \Pr(\chi_1^2 > c)\, w_2 \Pr(\chi_2^2 > c)$. For the SNP data, $k = 3$ and the weights can be estimated using [22] $w_1 = 0.5$, $w_2 = 0.5 - \cos^{-1}(-\sqrt{r_1 r_2/[(r - r_1)(r - r_2)]})/(2\pi)$. For association studies, the order (increase or decrease) is usually unknown before we observe the data. According to Barlow et al. [24], we can first compute the p values for both increasing and decreasing alternatives and then compute the overall p value as two times the smaller one.

3. Monte Carlo Simulation Study

A Monte Carlo simulation study is used to study the performance and the power properties of the proposed procedure as well as some existing procedures in the literature. We assume that the rows of case and control in table 1 follow multinomial distributions with probabilities $p = (p_0, p_1, p_2)$ and $q = (q_0, q_1, q_2)$, respectively.

Let

$$\frac{p_0}{q_0} = u, \quad \frac{p_1}{q_1} = \lambda_1 \frac{p_0}{q_0} = \lambda_1 u, \quad \frac{p_2}{q_2} = \lambda_2 \frac{p_0}{q_0} = \lambda_2 u,$$

where λ_1 and λ_2 are the relative risks. Since $p_0 + p_1 + p_2 = 1$, we have

$$u = \frac{1}{q_0 + \lambda_1 q_1 + \lambda_2 q_2}, \quad p_0 = \frac{q_0}{q_0 + \lambda_1 q_1 + \lambda_2 q_2},$$

$$p_1 = \frac{\lambda_1 q_1}{q_0 + \lambda_1 q_1 + \lambda_2 q_2} \quad \text{and} \quad p_2 = \frac{\lambda_2 q_2}{q_0 + \lambda_1 q_1 + \lambda_2 q_2}.$$

Therefore, for given q_is, and λ_1, λ_2, the values of the corresponding p_is can be obtained from the above formulas.

For the controls, we first assume Hardy-Weinberg equilibrium (HWE) holds and the minor allele frequencies (MAF) are 0.3 and 0.5. The numbers of cases (r) and controls (s) both equal 2,500 in our simulations. We use different λ_1 and λ_2 to compare the performance of our proposed method with those of GMS, MERT, MAX3, Pearson's χ^2 test, CATT with $x = 0.5$, and the RLRT. In our simulation study, we use significance level $\alpha = 10^{-5}$ to reflect the real situation of GWA studies where the total number of SNPs are large and the point-wise significance levels are usually very small. For each setting, we used 1,000,000 realizations to estimate the type I error rates (sizes) and power values of those test procedures. To estimate the p values of MAX3, GMS, and MERT, we used R package 'Rassoc' with option 'asy' which uses the asymptotic null distribution [17].

Note that under the null hypothesis that $\lambda_1 = \lambda_2 = 1$, the estimated rejection rates are the estimated type I error rates. These estimated type I error rates of different test procedures are presented in table 2. Figures 1 and 2 plot the estimated rejection rates of different test procedures when HWE holds. We also considered the situations where HWE does not hold for controls. Specifically, we assume the probabilities for genotypes (AA, Aa, aa) are (0.1, 0.3, 0.6) or (0.6, 0.3, 0.1) in controls. Figures 3 and 4 plot the estimated rejection rates for these two settings, respectively.

From the simulation results, for all of the methods except for RLRT, the methods control type I error rates
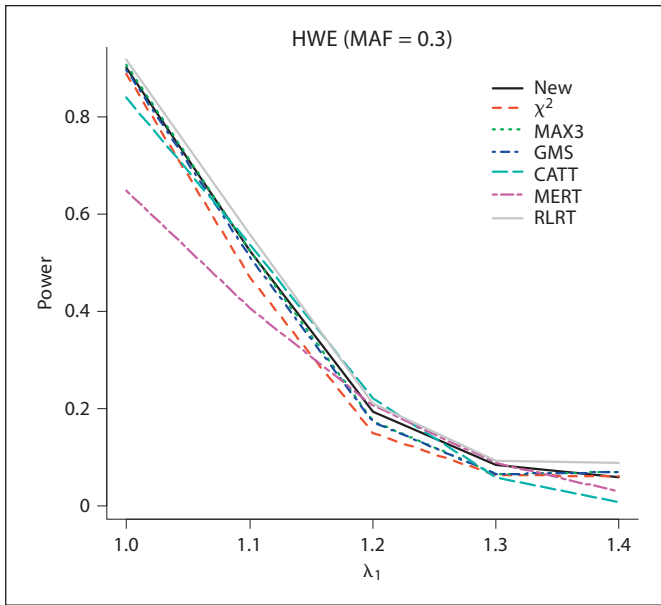
Fig. 1. The estimated power values of the different test procedures when HWE (MAF = 0.3) holds for controls with $\lambda_1 = 1.0, 1.1, 1.2, 1.3, 1.4$, and $\lambda_2 = 1.4$.

Fig. 2. The estimated power values of the different test procedures when HWE (MAF = 0.5) holds for controls with $\lambda_1 = 1.0, 1.1, 1.2, 1.3, 1.4$ and $\lambda_2 = 1.4$.

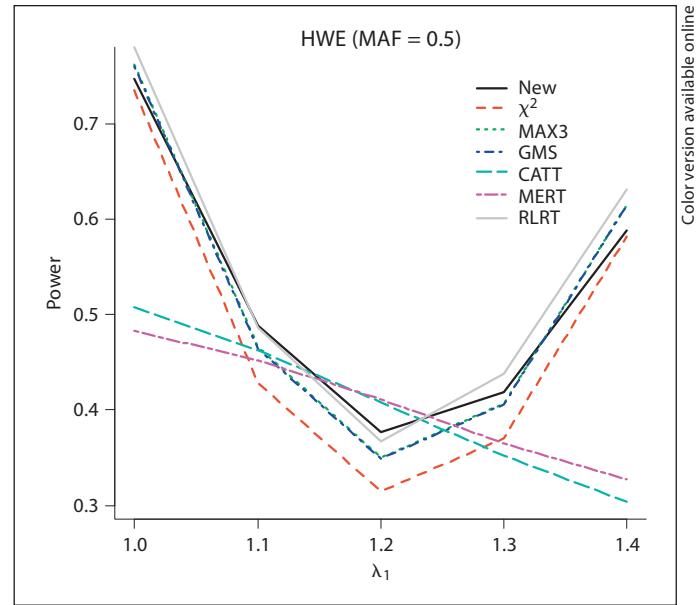**Table 2.** Estimated type I error rates for the test procedures discussed under different settings

Setting	W	χ^2	MAX3	GMS	CATT	MERT	RLRT
HWE (MAF = 0.3)	1.2×10^{-5}	1.3×10^{-5}	1.4×10^{-5}	1.4×10^{-5}	1.3×10^{-5}	1.5×10^{-5}	2.3×10^{-5}
HWE (MAF = 0.5)	0.7×10^{-5}	0.8×10^{-5}	0.8×10^{-5}	0.8×10^{-5}	0.9×10^{-5}	0.8×10^{-5}	0.9×10^{-5}
$q = (0.1, 0.3, 0.6)$	1.2×10^{-5}	1.1×10^{-5}	1.0×10^{-5}	1.0×10^{-5}	1.1×10^{-5}	1.4×10^{-5}	1.6×10^{-5}
$q = (0.6, 0.3, 0.1)$	0.6×10^{-5}	1.1×10^{-5}	0.6×10^{-5}	0.6×10^{-5}	1.1×10^{-5}	0.8×10^{-5}	0.5×10^{-5}

quite well. In general, RLRT and our proposed method have similar powers. However, our simulation study shows that sometimes RLRT had inflated type I error rates. For example, when HWE holds and MAF = 0.3, the estimated size using RLRT was 2.3×10^{-5}, which is statistically significantly different than the nominal level 1×10^{-5} at significance level 0.05. The situation can be even worse when MAF is smaller. If we assume HWE holds and MAF = 0.2 and 0.1, the estimated sizes from RLRT were 2.5×10^{-5}, and 2.8×10^{-5}, respectively. Also we can see that MAX3 and GMS have similar performances, while CATT with score $x = 0.5$ and MERT have power values close to each other.

For additive models ($\lambda_1 = 1.2, \lambda_2 = 1.4$), MERT and CATT are usually more powerful than other methods, as expected. However, when the true genetic model is dom-inant ($\lambda_1 = \lambda_2 = 1.4$) or recessive ($\lambda_1 = 1, \lambda_2 = 1.4$), MERT and CATT perform much worse than other methods. This indicates that MERT and CATT are sensitive to the underlying genetic models and therefore they are not robust. In contrast, MAX3 and GMS perform much better than MERT and CATT for dominant and recessive models; while they both have low power values for additive models. Under models other than recessive, additive, and dominant (i.e. $\lambda_1 = 1.1, \lambda_2 = 1.4$ and $\lambda_1 = 1.3, \lambda_2 = 1.4$), the proposed test and RLRT have among the three largest power values. Furthermore, if our proposed method is not the most powerful test for a given situation, it always has power value close to the largest one (usually the second or the third largest in power values). This indicates that our proposed method is robust in the sense that it has comparable power under different situations considered
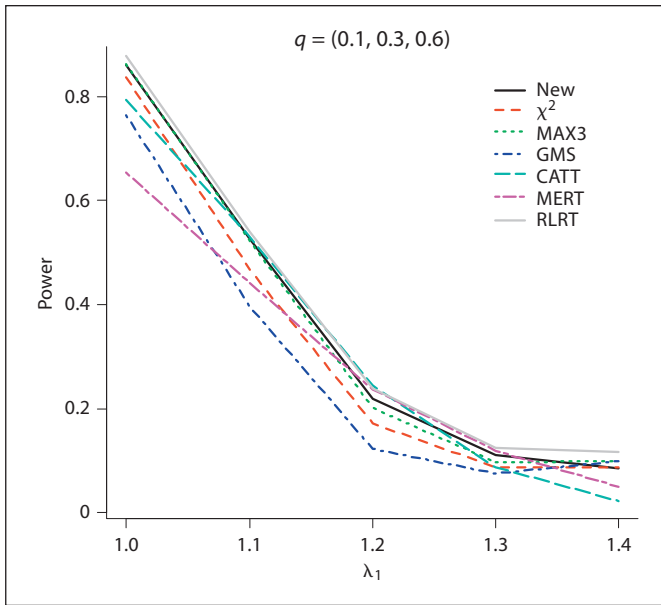
Fig. 3. The estimated power values of the different test procedures when probabilities for genotypes (*AA, Aa, aa*) are (0.1, 0.3, 0.6) in controls with $\lambda_1 = 1.0, 1.1, 1.2, 1.3, 1.4$, and $\lambda_2 = 1.4$.
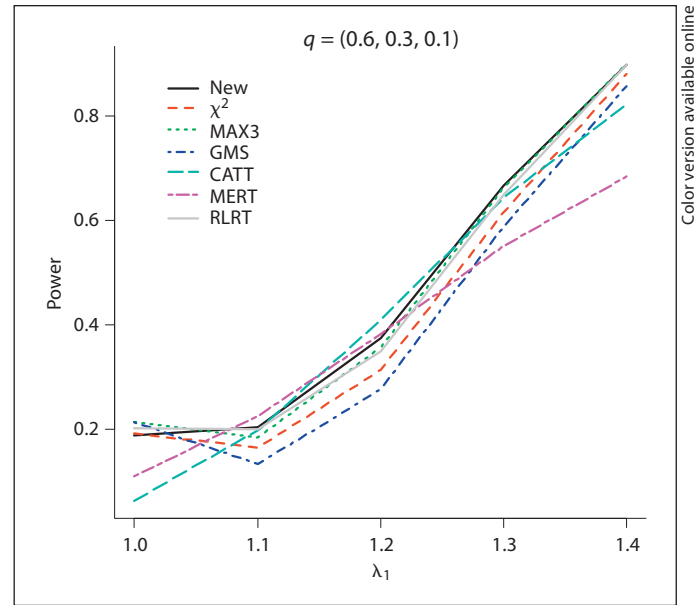

Fig. 4. The estimated power values of the different test procedures when probabilities for genotypes (*AA, Aa, aa*) are (0.6, 0.3, 0.1) in controls with $\lambda_1 = 1.0, 1.1, 1.2, 1.3, 1.4$, and $\lambda_2 = 1.4$.

in the simulation study. This robustness property is one of the merits of the proposed method because the underlying genetic models are seldom known in practice.

Figures 1–4 clearly show that the proposed methods have highest or close to highest powers for all the situations considered in the simulations. It should be noticed that for figures 1 and 2, when the estimated power values for MAX3 and GMS are very close to each other, the differences between the two lines for these two methods are not appreciable. From our simulation study, we observe that the power values of some methods not only depend on the genotypic frequencies, but also the genetic model. For example, in figure 2, except for CATT and MERT, the power values of all other methods decrease when λ_1 increases for $\lambda_1 < 1.2$ ($\lambda_1 = 1.2$ is the additive model) and the power values increase when λ_1 increases for $\lambda_1 > 1.2$.

4. Numerical Illustrations

In this section, we apply our proposed method with others to some real SNPs reported from four GWA studies with 100,000–500,000 SNPs for age-related macular degeneration (AMD) [25], two cancer studies [26, 27], and a hypertension study [28]. The datasets are summarized in table 3, which were taken from [17].

Table 3. Datasets for numerical illustrations (adapted from [17])

GWA study SNP ID	Case			Control		
	AA	*Aa*	*aa*	*AA*	*Aa*	*aa*
AMD						
rs380390	50	35	11	6	25	19
rs1329428	2	24	68	5	29	14
Prostate cancer						
rs1447295	25	283	864	10	218	929
rs6983267	223	598	351	301	579	277
rs7837688	27	283	861	11	206	939
Breast cancer						
rs10510126	10	180	955	14	272	854
rs12505080	50	477	608	99	408	628
rs17157903	18	316	777	26	220	862
rs1219648	250	543	352	170	538	433
rs7696175	187	605	353	249	496	396
rs2420946	242	546	357	165	537	440
Hypertension						
rs2820037	40	587	1,325	72	684	2,180
rs6997709	118	716	1,116	237	1,201	1,500
rs7961152	416	963	570	492	1,448	992
rs11110912	67	647	1,237	83	804	2,049
rs1937506	113	742	1,097	244	1,205	1,484
rs2398162	111	624	1,205	194	1,121	1,608

Table 4. p values for real SNP data from the various methods and observed statistics Z_1 and Z_2 from the proposed method

SNP	χ^2	MAX3	GMS	CATT	MERT	RLRT	Z_1	Z_2	W
rs380390	1.8×10^{-6}	1.0×10^{-6}	2.0×10^{-6}	3.1×10^{-7}	3.9×10^{-7}	3.1×10^{-7}	-4.04	-3.11	9.0×10^{-7}
rs1329428	3.6×10^{-6}	1.0×10^{-6}	1.0×10^{-6}	8.7×10^{-7}	8.5×10^{-6}	9.2×10^{-7}	1.92	4.60	2.0×10^{-6}
rs1447295	1.9×10^{-4}	8.3×10^{-5}	8.1×10^{-5}	4.5×10^{-5}	6.0×10^{-5}	4.7×10^{-5}	-2.46	-3.32	8.4×10^{-5}
rs6983267	3.5×10^{-5}	3.0×10^{-5}	2.9×10^{-5}	7.9×10^{-6}	7.4×10^{-6}	1.5×10^{-5}	3.67	2.66	1.5×10^{-5}
rs7837688	1.6×10^{-5}	5.0×10^{-6}	6.0×10^{-6}	2.7×10^{-6}	8.3×10^{-6}	3.5×10^{-6}	-2.51	-3.98	6.9×10^{-6}
rs10510126	3.7×10^{-6}	1.0×10^{-6}	3.0×10^{-6}	1.4×10^{-6}	1.7×10^{-4}	8.2×10^{-7}	0.78	4.94	2.9×10^{-6}
rs12505080	1.8×10^{-5}	7.4×10^{-5}	8.3×10^{-5}	0.32	0.039	5.3×10^{-5}	4.27	-1.90	2.4×10^{-4}
rs17157903	9.9×10^{-6}	5.3×10^{-5}	4.8×10^{-5}	6.3×10^{-4}	0.058	3.8×10^{-5}	1.31	-4.62	5.0×10^{-5}
rs1219648	7.5×10^{-6}	4.0×10^{-6}	7.0×10^{-6}	1.8×10^{-6}	1.4×10^{-6}	2.9×10^{-6}	-3.92	-2.87	3.2×10^{-6}
rs7696175	1.6×10^{-5}	2.1×10^{-3}	1.9×10^{-3}	0.59	0.40	1.8×10^{-3}	3.77	-2.80	1.7×10^{-3}
rs2420946	8.8×10^{-6}	7.0×10^{-6}	4.0×10^{-6}	1.9×10^{-6}	1.5×10^{-6}	3.5×10^{-6}	-3.82	-2.95	3.7×10^{-6}
rs2820037	7.7×10^{-7}	1.0×10^{-6}	2.0×10^{-6}	5.8×10^{-5}	0.013	2.3×10^{-6}	1.03	-5.21	2.9×10^{-6}
rs6997709	4.4×10^{-5}	2.8×10^{-5}	1.4×10^{-5}	7.9×10^{-6}	1.9×10^{-5}	1.4×10^{-5}	2.43	3.77	1.9×10^{-5}
rs7961152	3.0×10^{-5}	1.4×10^{-5}	1.6×10^{-5}	7.4×10^{-6}	6.0×10^{-6}	1.4×10^{-5}	-3.68	-2.70	1.3×10^{-5}
rs11110912	1.9×10^{-5}	5.0×10^{-6}	1.9×10^{-5}	9.2×10^{-6}	2.2×10^{-4}	5.8×10^{-6}	-1.08	-4.53	1.3×10^{-5}
rs1937506	4.5×10^{-5}	3.4×10^{-5}	2.6×10^{-5}	9.2×10^{-6}	8.5×10^{-6}	1.3×10^{-5}	3.13	3.19	1.9×10^{-5}
rs2398162	5.7×10^{-6}	1.0×10^{-6}	2.0×10^{-6}	7.9×10^{-6}	1.2×10^{-4}	1.7×10^{-6}	1.02	4.81	4.0×10^{-6}

Table 4 reports the p values for each SNP from the different methods. It can be observed that when the genetic model is between recessive and dominant ($Z_1 > 0$ and $Z_2 > Z_1$, or $Z_1 < Z_2$ and $Z_2 < 0$), the proposed method has similar p values with those from CATT, which is usually more powerful than other methods. However, when the genetic model is close to recessive or dominant, CATT performs poorly, but GMS, MAX3 and our proposed method have similar p values and better than CATT. This indicates that GMS, MAX3 and our proposed method perform similar under this situation. For SNPs having large absolute values but with different signs of Z_1 and Z_2, their genetic models do not belong to the generalized genetic model. Under those situations, Pearson's χ^2 test has the smallest p value as expected, since it is more robust than any other method, while CATT and MERT have large p values. The p values of the proposed method are similar to those from GMS and MAX3. Note that the p values from RLRT are usually similar to or smaller than those from our proposed test. However, we found that ten out of the seventeen estimated MAF values from controls are less than 0.3; these small p values may be due to the liberal nature of the RLRT for highly unbalanced data as mentioned before.

In these illustrations, we can also see that the two observed statistics Z_1 and Z_2 can be used to determinate the genetic model and the at-risk allele. For example, for SNP rs380390 in table 4, $Z_1 = -4.04$, $Z_2 = -3.11$, since both Z_1 and Z_2 are negative, the at-risk allele is A instead of a; the genetic model should be neither recessive nor dominant, but between the two since the absolute values of Z_1 and Z_2 are much larger than 1. From table 4, we also see that some SNPs have Z_1 and Z_2 with different signs. Three out of six SNPs from the breast cancer data fall into this category. This situation deserves special attention. It is possible that the underlying genetic models are over- or under-dominant. But it is also possible that this happened merely due to chance or something else, such as population substructure, which needs further investigation. The breast cancer SNP data in table 3 were taken from the Nurses' Health Study (NHS) and three additional studies have been conducted by the authors [26]. The p values for SNP rs17157903 from the other three studies are 0.72, 0.49 and 0.92, respectively, using the χ^2 test. Therefore the association between the SNP rs17157903 and breast cancer needs to be validated by future studies.

5. Conclusions

In GWA studies, since the underlying genetic models are usually unknown, choosing a powerful statistic test is desirable. There is no single test performing uniformly better than the other competitors, and most of the existing methods may suffer from serious power lost under some models. Through Monte Carlo simulations and the study of real SNP data, we have seen that our proposed

method is more robust and powerful than existing methods in many situations.

The robustness of the proposed test is expected since, unlike the CATT test, it does not require a complete genetic model specification except that we assume the model belongs to a generalized ORRR model. Therefore it is not very sensitive to model mis-specification. Meanwhile, the proposed test correctly incorporates the property of the monotonicity of the relative risks for SNP data in GWA studies which results in power gains. Moreover, based on the simulation results, we observed that when the genetic models are not one of the perfect models (i.e. recessive, additive, and dominant models), the proposed test usually has the highest or second highest power. In real world applications, the perfect models may be rare if not impossible; hence the proposed method is certainly preferable. Finally, through simulation (data not shown) and real data (see table 4), when the genetic model is not ORRR, e.g. over- or under-dominant, the proposed method has reasonable power. Beside the robustness of the proposed method, another advantage of the proposed method is that the p value can be easily approximated with very high accuracy. Although sometimes RLRT can also be applied to generalized ORRR models, it should be used with caution as it inflates type I error rates when the data are highly unbalanced (e.g. HWE with small MAF).

Since some of SNPs are highly correlated due to linkage disequilibrium, the p values obtained from individual SNPs are also correlated. Traditional multiple tests correction methods, such as the Bonferroni procedure, are not appropriate. One may choose instead to use the recently proposed method which is based on the concept of effective number [29].

Appendix

Proof of Theorem 1
For large sample sizes, which are usually available for GWA studies, we can assume Z_1 and Z_2 are independently and identically distributed as standard normal, so that $\Phi(Z_1)$ and $\Phi(Z_2)$ are independently and identically distributed uniformly between 0 and 1. Therefore, according to Fisher [30], C_1 is χ^2 distributed with df = 4. Similarly, C_2 is χ^2 distributed with df = 4.

Proof of Theorem 2
For a large sample size, from Theorem 1, C_1 and C_2 are both χ^2 distributed with df = 4. Using the concept of associated random variables by Esary et al. [31] and Theorem 2 by Owen [32], we have

$$\Pr(C_1 > w) + \Pr(C_2 > w) - \Pr(C_1 > w)\Pr(C_2 > w) \leq \Pr(W > w) \leq \Pr(C_1 > w) + \Pr(C_2 > w), \text{ i.e. } 2\gamma - \gamma^2 \leq \Pr(W > w) \leq 2\gamma.$$

Acknowledgements

References

1 Armitage P: Tests for linear trends in proportions and frequencies. Biometrics 1955; 11:375–386.
2 Balding DJ: A tutorial on statistical methods for population association studies. Nat Rev Genet 2006;7:781–791.
3 Cochran W: Some methods for strengthening the common chi-square tests. Biometrics 1954;10:417–451.
4 Zheng G, Freidlin B, Gastwirth JL: Comparison of robust tests for genetic association using case-control studies; in Rojo J (ed): Optimality: The Second Erich L. Lehmann Symposium, May 19–22, 2004, Rice University. IMS Lecture Notes – Monograph Series. Beachwood, Institute of Mathematical Statistics, 2006, vol 49, pp 253–265.
5 Chen Z, Zheng G: Exact robust tests for detecting candidate-gene association in case-control trio design. J Data Sci 2005;3:19–33.

6 Freidlin B, Podgor MJ, Gastwirth JL: Efficiency robust tests for survival or ordered categorical data. Biometrics 1999;55:883–886.
7 Freidlin B, Zheng G, Li Z, Gastwirth JL: Trend tests for case-control studies of genetic markers: power, sample size and robustness. Hum Hered 2002;53:146–152.
8 Gonzalez JR, Carrasco JL, Dudbridge F, Armengol L, Estivill X, Moreno V: Maximizing association statistics over genetic models. Genet Epidemiol 2008;32:246–254.
9 Kwak M, Joo J, Zheng G: A robust test for two-stage design in genome-wide association studies. Biometrics 2009;65:1288–1295.
10 Sasieni PD: From genotypes to genes: doubling the sample size. Biometrics 1997;53: 1253–1261.

11 Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, Boutin P, Vincent D, Belisle A, Hadjadj S, Balkau B, Heude B, Charpentier G, Hudson TJ, Montpetit A, Pshezhetsky AV, Prentki M, Posner BI, Balding DJ, Meyre D, Polychronakos C, Froguel P: A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature 2007;445:881–885.
12 Slager SL, Schaid DJ: Case-control studies of genetic markers: power and sample size approximations for Armitage's test for trend. Hum Hered 2001;52:149–153.
13 Song K, Elston RC: A powerful method of combining measures of association and Hardy-Weinberg disequilibrium for fine-mapping in case-control studies. Stat Med 2006; 25:105–126.
14 Wang K, Sheffield VC: A constrained-likelihood approach to marker-trait association studies. Am J Hum Genet 2005;77:768–780.

15 Zheng G, Freidlin B, Li Z, JL G: Choice of scores in trend tests for case-control studies of candidate gene associations. Biometrical J 2003;45:335–348.

16 Zheng G, Ng HKT: Genetic model selection in two-phase analysis for case-control association studies. Biostatistics 2008;9:391–399.

17 Zang Y, Fung WK, Zheng G: Simple algorithms to calculate the asymptotic null distributions of robust tests in case-control genetic association studies in R. J Stat Software 2010;33:1–24.

18 Gastwirth JL: On robust procedures. J Am Stat Assoc 1966;61:929–948.

19 Gastwirth JL: The use of maximin efficiency robust tests in combining contingency tables and survival analysis. J Am Stat Assoc 1985;80:380–384.

20 Song K, Elston RC: A powerful method of combining measures of association and Hardy-Weinberg disequilibrium for fine mapping in case-control studies. Stat Med 2006;25:105–126.

21 Agresti A, Coull BA: The analysis of contingency tables under inequality constraints. J Stat Pplan Inference 2002;107:45–73.

22 Silvapulle MJ, Sen PK: Constrained Statistical Inference: Inequality, Order, and Shape Restrictions. Hoboken, NJ, John Wiley & Sons, 2005.

23 Ayer M, Brunk H, Ewing GM, Reid W, Silverman E: An empirical distribution function for sampling with incomplete information. Ann Math Stat 1955:641–647.

24 Barlow R, Bartholomew D, Bremner J, Brunk H: Statistical Inference under Order Restrictions. New York, Wiley, 1972.

25 Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J: Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385–389.

26 Hunter DL, Kraft P, Jacobs KB, Cox DG, Yeager N, Hankinson SE, Wacholder S, Wang Z, Welch R, et al: A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenpausal breast cancer. Nat Genet 2007;39:870–874.

27 Yeager M, Orr N, Hayes R, Jacobs K, Kraft P, Wacholder S, Minichiello M, Fearnhead P, Yu K, et al: Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007;39:645–649.

28 The Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678.

29 Chen Z, Liu Q: A new approach to account for the correlations among single nucleotide polymorphisms in genome-wide association studies. Hum Hered 2011;72:1–9.

30 Fisher RA (ed): Statistical Methods for Research Workers. Edinburgh, Oliver and Boyd, 1932.

31 Esary JD, Proschan F, Walkup DW: Association of random variables, with applications. Ann Math Stat 1967;38:1466–1474.

32 Owen AB: Karl Pearson's meta-analysis revisited. Ann Stat 2009;37:3867–3892.