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

Zhongxue Chen\textsuperscript{a}    Hon Keung Tony Ng\textsuperscript{b}

\textsuperscript{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 \textsuperscript{b}Department of Statistical Science, Southern Methodist University, Dallas, Tex., USA

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, \textit{A} and \textit{a}, which is assumed to be at risk, there are three genotypes: \textit{AA}, \textit{Aa}, and \textit{aa}. Suppose that there are \textit{r} cases and \textit{s} controls in the study. In the \textit{r} cases, there are \textit{r}_0, \textit{r}_1, and \textit{r}_2 affected people with genotypes \textit{AA}, \textit{Aa}, and \textit{aa}, respectively. There are \textit{s}_0, \textit{s}_1, and \textit{s}_2 people with genotypes \textit{AA}, \textit{Aa}, and \textit{aa}, respectively, in the \textit{s} unaffected controls.

Testing whether there is an association between the genotype and the disease status is equivalent to testing the association in the $2 \times 3$ contingency table. Pearson’s $\chi^2$ 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...
with $Aa$ is between the risks associated with $AA$ and $aa$. However, Pearson’s $\chi^2$ 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 $\chi^2$ 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 $\chi^2$ 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 $\chi^2$ 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 maximin 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 $\chi^2$ 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.

### Table 1. SNP data in GWA studies

<table>
<thead>
<tr>
<th>Genotype</th>
<th>$AA$</th>
<th>$Aa$</th>
<th>$aa$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>$r_0$</td>
<td>$r_1$</td>
<td>$r_2$</td>
<td>$r$</td>
</tr>
<tr>
<td>Control</td>
<td>$s_0$</td>
<td>$s_1$</td>
<td>$s_2$</td>
<td>$s$</td>
</tr>
<tr>
<td>Total</td>
<td>$n_0$</td>
<td>$n_1$</td>
<td>$n_2$</td>
<td>$n$</td>
</tr>
</tbody>
</table>
The statistic for MAX3 is [7]:

\[ \text{MAX3} = \max \{ |Z_0|, |Z_{1/2}|, |Z_1| \} \]

The statistic for GMS is [16, 17]:

\[ \text{GMS} = Z_0 I(Z_{1/2} > 0)I(Z_{HWDTS} < c) + Z_{1/2} I(Z_{1/2} > 0)I(Z_{HWDTS} < c) + Z_{1/2} I(Z_{HWDTS} < c) - Z_0 I(Z_{1/2} \leq 0)I(Z_{HWDTS} < c) + Z_{1/2} I(Z_{1/2} \leq 0)I(Z_{HWDTS} < c) - Z_0 I(Z_{1/2} \leq 0)I(Z_{HWDTS} < c), \]

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

\[ Z_{HWDTS} = \left( \frac{r s}{n} \right)^{1/2} \left( \hat{\Delta}_p - \hat{\Delta}_Q \right) \]

where \( \hat{\Delta}_p = r_s / r - (r_s / r + s / (2r))^2 \), \( \hat{\Delta}_Q = s_2 / s - (s_2 / s + s_2 / (2s))^2 \), and c is a constant and usually chosen as 1.645.

The statistic for MERT is [19]:

\[ \text{MERT} = (Z_0 + Z_{1/2})/\sqrt{2(1 + \rho_0)}, \]

where \( \rho_0 = (n_0 n_2)/(n_0 + n_1)(n_1 + n_2) \)

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 \( q_0, q_1, q_2 \) 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

\[
\begin{align*}
\lambda_1 &= \frac{p_1 \Pr(AA)}{p_0 \Pr(Aa)} \\
\lambda_2 &= \frac{p_1 \Pr(AA)}{p_0 \Pr(Aa)}
\end{align*}
\]

where \( \Pr(AA) = \Pr(AA | \text{case}) \Pr(\text{case}) + \Pr(AA | \text{control}) \Pr(\text{control}) = k_p 0 + (1 - k) q_0 \) and \( k \) is the disease prevalence. Similarly, we have \( \Pr(Aa) = k_p 1 + (1 - k) q_1 \) and \( \Pr(\text{AA}) = k_p 2 + (1 - k) q_2 \).

Then, equation 1 can be written as

\[
\begin{align*}
\lambda_1 &= \frac{p_1 \{k_p_0 + (1 - k) q_0\}}{p_0 \{k_p_1 + (1 - k) q_1\}} \\
\lambda_2 &= \frac{p_1 \{k_p_0 + (1 - k) q_0\}}{p_0 \{k_p_2 + (1 - k) q_2\}}
\end{align*}
\]

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

\[
\begin{align*}
[p_1 q_0 \geq p_0 q_1] \\
[p_1 q_0 \geq p_0 q_1]
\end{align*}
\]

Let

\[
\begin{align*}
T_1 &= p_0 q_0 - p_0 q_1 \\
T_2 &= p_1 q_0 - p_1 q_1
\end{align*}
\]

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

\[
\begin{align*}
T_1 &= \frac{r_s q_0 - r_s q_1}{rs} \\
T_2 &= \frac{r_s q_0 - r_s q_1}{rs}
\end{align*}
\]

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 \( T_1 \) and \( T_2 \) are zeros (i.e. \( E(T_1) = E(T_2) = 0 \) ) and the variance-covariance matrix of \( T_1 \) and \( T_2 \) is

\[
\sum = \begin{bmatrix}
\text{Var}(T_1) & \text{Cov}(T_1, T_2) \\
\text{Cov}(T_1, T_2) & \text{Var}(T_2)
\end{bmatrix}
\]

where

\[
\begin{align*}
\text{Var}(T_1) &= p_0 p_1 (2 p_2 + n (p_0 + p_1))/rs \\
\text{Var}(T_2) &= p_0 p_1 (2 p_2 + n (p_0 + p_1))/rs \\
\text{Cov}(T_1, T_2) &= p_0 p_1 (2 - n) /rs
\end{align*}
\]

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

\[
\hat{\Sigma} = \frac{n_i}{rs n} \begin{bmatrix}
2 n_2 + n (n_0 + n_1) & n_0 n_2 (2 - n) \\
(n_0 + n_1) & n_0 n_2 (2 - n)
\end{bmatrix}
\]

Because \( \hat{\Sigma} \) is a positive definite square matrix, eigen-decomposition of \( \hat{\Sigma} \) gives \( \hat{\Sigma} = P \hat{D} P' \), 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

\[
Z = \begin{bmatrix}
Z_1 \\
Z_2
\end{bmatrix} = \hat{\Sigma}^{-1} \begin{bmatrix}
T_1 \\
T_2
\end{bmatrix}
\]

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

\[
\hat{\Sigma}^{-1} = P \begin{bmatrix}
d_1 & 0 \\
0 & d_2
\end{bmatrix} P'
\]

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 Z₁ and Z₂ are still valid with E(Z₁) and E(Z₂) 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

\[
\begin{align*}
C₁ &= -2 \log(\Phi(Z₁)\Phi(Z₂)) \\
C₂ &= -2 \log(\Phi(Z₁-Z₂))
\end{align*}
\]

where \(\Phi(·)\) is the cumulative distribution function of the standard normal distribution. The asymptotic distributions of \(C₁\) and \(C₂\) are as follows.

**Theorem 1**
Under the null hypothesis, \(C₁\) and \(C₂\) are asymptotically \(\chi^2\) distributed with 4 df.

The proposed test statistic is the maximum of \(C₁\) and \(C₂\) denoted as

\[W = \max(C₁, C₂).
\]

It should be noted that \(C₁\) and \(C₂\) 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

\[2\gamma - \gamma^2 \leq Pr(W > w) \leq 2\gamma,
\]

where \(\gamma = 1 - \chi^2(w)\) and \(\chi^2(·)\) is the cumulative distribution function of the \(\chi^2\) distribution with 4 df.

Theorem 2 suggests that we can estimate the p value of the test procedure based on \(W\) by \(2\gamma\), and with small \(\gamma\), 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:

\[RLRT = 2 \sum_{i=1}^{3} r_i \log\left(\hat{\pi}_i / \hat{p}\right) + \sum_{i=1}^{3} s_i \log\left((1 - \hat{\pi}_i) / (1 - \hat{p})\right),
\]

where \(\hat{p} = r/n\), and \(\hat{\pi}_i = r_i/n_i\), \(i = 0, 1, 2\), are order-restricted MLEs satisfying \(\hat{\pi}_0 \leq \hat{\pi}_1 \leq \hat{\pi}_2\). Usually \(\hat{\pi}_i\) is estimated using the pool adjacent violators algorithm (PAVA) [23], and the above statistic has a weighted \(\chi^2\) distribution \(\left(\chi^2\right)\) under the null hypothesis [24]. Its p value is \(Pr(\text{RLRT} > q) = w_1 Pr(\chi^2_1 > q) + w_2 Pr(\chi^2_2 > q)\). For the SNP data, \(k = 3\) and the weights can be estimated using [22] \(w_1 = 0.5, w_2 = 0.5 - \cos^{-1}(r/r_2)(r-r_1)(r-r_2)/2m\). 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

\[
\begin{align*}
P_0 &= \frac{p_0}{q_0}, & P_1 &= \frac{p_1}{q_1}, & P_2 &= \frac{p_2}{q_2},
\end{align*}
\]

where \(\lambda_1\) and \(\lambda_2\) are the relative risks. Since \(p_0 + p_1 + p_2 = 1\), we have

\[
\begin{align*}
u &= \frac{1}{q_0 + \lambda_1 q_1 + \lambda_2 q_2}, & \phi &= \frac{q_0}{q_0 + \lambda_1 q_1 + \lambda_2 q_2}, & \psi &= \frac{\lambda_1 q_1}{q_0 + \lambda_1 q_1 + \lambda_2 q_2},
\end{align*}
\]

Therefore, for given \(q_0\), \(\lambda_1\), \(\lambda_2\), the values of the corresponding \(p\) 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 \(\lambda_1\) and \(\lambda_2\) to compare the performance of our proposed method with those of GMS, MERT, MAX3, Pearson’s \(\chi^2\) 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
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 $2.3 \times 10^{-5}$, which is statistically significantly different than the nominal level $1 \times 10^{-5}$ 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 $2.5 \times 10^{-5}$ and $2.8 \times 10^{-5}$, 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 dominant ($\lambda_1 = \lambda_2 = 1.4$) 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.
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 $\lambda_1$ increases for $\lambda_1 < 1.2$ ($\lambda_1 = 1.2$ is the additive model) and the power values increase when $\lambda_1$ 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>
<thead>
<tr>
<th>GWA study</th>
<th>SNP ID</th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD</td>
<td>rs380390</td>
<td>50</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>rs1329428</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>rs1447295</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>rs6983267</td>
<td>233</td>
<td>310</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>rs7837688</td>
<td>27</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>rs10510126</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>rs12505080</td>
<td>50</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>rs7961152</td>
<td>333</td>
<td>239</td>
</tr>
<tr>
<td></td>
<td>rs1219648</td>
<td>250</td>
<td>170</td>
</tr>
<tr>
<td></td>
<td>rs7696175</td>
<td>187</td>
<td>249</td>
</tr>
<tr>
<td></td>
<td>rs2420946</td>
<td>242</td>
<td>165</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hypertension</th>
<th>SNP ID</th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2820037</td>
<td>40</td>
<td>72</td>
<td>684</td>
</tr>
<tr>
<td>rs6997709</td>
<td>118</td>
<td>237</td>
<td>1,201</td>
</tr>
<tr>
<td>rs7961152</td>
<td>416</td>
<td>492</td>
<td>1,448</td>
</tr>
<tr>
<td>rs11110912</td>
<td>67</td>
<td>83</td>
<td>804</td>
</tr>
<tr>
<td>rs1937506</td>
<td>113</td>
<td>244</td>
<td>1,205</td>
</tr>
<tr>
<td>rs2398162</td>
<td>111</td>
<td>194</td>
<td>1,121</td>
</tr>
</tbody>
</table>

Table 3. Datasets for numerical illustrations (adapted from [17])
Table 4. p values for real SNP data from the various methods and observed statistics $Z_1$ and $Z_2$ from the proposed method

<table>
<thead>
<tr>
<th>SNP</th>
<th>$\chi^2$</th>
<th>MAX3</th>
<th>GMS</th>
<th>CATT</th>
<th>MERT</th>
<th>RLRT</th>
<th>$Z_1$</th>
<th>$Z_2$</th>
<th>$W'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs380390</td>
<td>$1.8 \times 10^{-6}$</td>
<td>$1.0 \times 10^{-6}$</td>
<td>$2.0 \times 10^{-6}$</td>
<td>$3.1 \times 10^{-7}$</td>
<td>$3.9 \times 10^{-7}$</td>
<td>$3.1 \times 10^{-7}$</td>
<td>$-4.04$</td>
<td>$-3.11$</td>
<td>$9.0 \times 10^{-7}$</td>
</tr>
<tr>
<td>rs1329428</td>
<td>$3.6 \times 10^{-6}$</td>
<td>$1.0 \times 10^{-6}$</td>
<td>$1.0 \times 10^{-6}$</td>
<td>$8.7 \times 10^{-7}$</td>
<td>$8.5 \times 10^{-7}$</td>
<td>$9.2 \times 10^{-7}$</td>
<td>$1.92$</td>
<td>$1.86$</td>
<td>$2.0 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs1447295</td>
<td>$1.9 \times 10^{-4}$</td>
<td>$8.3 \times 10^{-5}$</td>
<td>$8.1 \times 10^{-5}$</td>
<td>$4.5 \times 10^{-5}$</td>
<td>$6.0 \times 10^{-5}$</td>
<td>$4.7 \times 10^{-5}$</td>
<td>$-2.46$</td>
<td>$-3.32$</td>
<td>$8.4 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs6983267</td>
<td>$3.5 \times 10^{-5}$</td>
<td>$3.0 \times 10^{-5}$</td>
<td>$2.9 \times 10^{-5}$</td>
<td>$7.9 \times 10^{-6}$</td>
<td>$7.4 \times 10^{-6}$</td>
<td>$1.5 \times 10^{-5}$</td>
<td>$3.67$</td>
<td>$2.66$</td>
<td>$1.5 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs787688</td>
<td>$1.6 \times 10^{-5}$</td>
<td>$5.0 \times 10^{-6}$</td>
<td>$6.0 \times 10^{-6}$</td>
<td>$2.7 \times 10^{-6}$</td>
<td>$8.3 \times 10^{-6}$</td>
<td>$3.5 \times 10^{-6}$</td>
<td>$-2.51$</td>
<td>$-3.98$</td>
<td>$6.9 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs1051026</td>
<td>$3.7 \times 10^{-6}$</td>
<td>$1.0 \times 10^{-6}$</td>
<td>$3.0 \times 10^{-6}$</td>
<td>$1.4 \times 10^{-6}$</td>
<td>$1.7 \times 10^{-4}$</td>
<td>$8.2 \times 10^{-7}$</td>
<td>$0.78$</td>
<td>$-4.94$</td>
<td>$2.9 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs12505080</td>
<td>$1.8 \times 10^{-5}$</td>
<td>$7.4 \times 10^{-5}$</td>
<td>$8.3 \times 10^{-5}$</td>
<td>$0.32$</td>
<td>$0.039$</td>
<td>$5.3 \times 10^{-5}$</td>
<td>$4.27$</td>
<td>$-1.90$</td>
<td>$2.4 \times 10^{-4}$</td>
</tr>
<tr>
<td>rs17157903</td>
<td>$9.9 \times 10^{-6}$</td>
<td>$5.3 \times 10^{-5}$</td>
<td>$4.8 \times 10^{-5}$</td>
<td>$6.3 \times 10^{-4}$</td>
<td>$0.058$</td>
<td>$3.8 \times 10^{-5}$</td>
<td>$1.31$</td>
<td>$-4.62$</td>
<td>$5.0 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs1219648</td>
<td>$7.5 \times 10^{-6}$</td>
<td>$4.0 \times 10^{-6}$</td>
<td>$7.0 \times 10^{-6}$</td>
<td>$1.8 \times 10^{-6}$</td>
<td>$1.4 \times 10^{-6}$</td>
<td>$2.9 \times 10^{-6}$</td>
<td>$-3.92$</td>
<td>$-2.87$</td>
<td>$3.2 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs7696175</td>
<td>$1.6 \times 10^{-5}$</td>
<td>$2.1 \times 10^{-3}$</td>
<td>$1.9 \times 10^{-3}$</td>
<td>$0.59$</td>
<td>$0.40$</td>
<td>$1.8 \times 10^{-3}$</td>
<td>$3.77$</td>
<td>$-2.80$</td>
<td>$1.7 \times 10^{-3}$</td>
</tr>
<tr>
<td>rs2420946</td>
<td>$8.8 \times 10^{-6}$</td>
<td>$7.0 \times 10^{-6}$</td>
<td>$4.0 \times 10^{-6}$</td>
<td>$1.9 \times 10^{-6}$</td>
<td>$1.5 \times 10^{-6}$</td>
<td>$3.5 \times 10^{-6}$</td>
<td>$-3.82$</td>
<td>$-2.95$</td>
<td>$3.7 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs2820037</td>
<td>$7.7 \times 10^{-7}$</td>
<td>$1.0 \times 10^{-6}$</td>
<td>$2.0 \times 10^{-6}$</td>
<td>$5.8 \times 10^{-5}$</td>
<td>$0.013$</td>
<td>$2.3 \times 10^{-6}$</td>
<td>$1.03$</td>
<td>$-5.24$</td>
<td>$2.9 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs6997709</td>
<td>$4.4 \times 10^{-5}$</td>
<td>$2.8 \times 10^{-5}$</td>
<td>$1.4 \times 10^{-5}$</td>
<td>$7.9 \times 10^{-6}$</td>
<td>$1.9 \times 10^{-5}$</td>
<td>$1.4 \times 10^{-5}$</td>
<td>$2.43$</td>
<td>$3.77$</td>
<td>$1.9 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs7961152</td>
<td>$3.0 \times 10^{-5}$</td>
<td>$1.4 \times 10^{-5}$</td>
<td>$1.6 \times 10^{-5}$</td>
<td>$7.4 \times 10^{-6}$</td>
<td>$6.0 \times 10^{-6}$</td>
<td>$1.4 \times 10^{-5}$</td>
<td>$-3.68$</td>
<td>$-2.70$</td>
<td>$1.3 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs1110919</td>
<td>$1.9 \times 10^{-5}$</td>
<td>$5.0 \times 10^{-6}$</td>
<td>$1.9 \times 10^{-5}$</td>
<td>$9.2 \times 10^{-6}$</td>
<td>$2.2 \times 10^{-4}$</td>
<td>$5.8 \times 10^{-6}$</td>
<td>$-1.08$</td>
<td>$-4.53$</td>
<td>$1.3 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs1937506</td>
<td>$4.5 \times 10^{-5}$</td>
<td>$3.4 \times 10^{-5}$</td>
<td>$2.6 \times 10^{-5}$</td>
<td>$9.2 \times 10^{-6}$</td>
<td>$8.5 \times 10^{-6}$</td>
<td>$1.3 \times 10^{-5}$</td>
<td>$3.13$</td>
<td>$3.19$</td>
<td>$1.9 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs2398162</td>
<td>$5.7 \times 10^{-6}$</td>
<td>$1.0 \times 10^{-6}$</td>
<td>$2.0 \times 10^{-6}$</td>
<td>$7.9 \times 10^{-6}$</td>
<td>$1.2 \times 10^{-4}$</td>
<td>$1.7 \times 10^{-6}$</td>
<td>$1.02$</td>
<td>$4.81$</td>
<td>$4.0 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

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 $\chi^2$ 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 determine 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 $\chi^2$ 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 rate 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 \( \chi^2 \) distributed with \( df = 4 \). Similarly, \( C_2 \) is \( \chi^2 \) distributed with \( df = 4 \).

Proof of Theorem 2

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

\[
Pr(C_1 > w) + Pr(C_2 > w) - Pr(C_1 > w)Pr(C_2 > w) \leq Pr(W > w)
\]

We have \( Pr(C_1 > w) \) and \( Pr(C_2 > w) \), i.e. \( 2\gamma - \gamma^2 \leq Pr (W > w) \leq 2\gamma \).

Acknowledgements

The first author would like to thank for the support from the NIH grant (UL1 RR024148) awarded to the University of Texas Health Science Center at Houston. The authors would also like to thank Professor Wayne Woodward for his suggestions which resulted in a much improved version of the manuscript.

References

In the article 'Genome-Wide Association Studies for Bivariate Sparse Longitudinal Data' by Das et al. [Hum Hered 2011;72:110–120], the following acknowledgement needs to be added:

Acknowledgements

The Framingham Heart Study project is conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University (N01 HC25195). The Genetic Analysis Workshop is supported by NIH grant R01 GM31575. The simulated data was supported by the Washington University Institute of Clinical and Translational Sciences, NIH grant 1U54RR023496. The GAW16 Framingham and simulated data used for the analyses described in this paper were obtained through dbGaP (accession numbers). The authors acknowledge the investigators that contributed the phenotype, genotype, and simulated data for this study. The manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or the NHLBI.