A Generalized Sequential Bonferroni Procedure Using Smoothed Weights for Genome-Wide Association Studies Incorporating Information on Hardy-Weinberg Disequilibrium among Cases

Guimin Gao\textsuperscript{a} Guolian Kang\textsuperscript{b} Jiexun Wang\textsuperscript{a} Wenan Chen\textsuperscript{a} Huaizen Qin\textsuperscript{c} Bo Jiang\textsuperscript{d} Qizhai Li\textsuperscript{e} Chuanyu Sun\textsuperscript{a} Nianjun Liu\textsuperscript{d} Kellie J. Archer\textsuperscript{a} David B. Allison\textsuperscript{d}

\textsuperscript{a}Department of Biostatistics, Virginia Commonwealth University, Richmond, Va., \textsuperscript{b}Department of Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pa., \textsuperscript{c}Department of Biostatistics and Epidemiology, Case Western Reserve University, Cleveland, Ohio, and \textsuperscript{d}Department of Biostatistics, University of Alabama at Birmingham, Birmingham, Ala., USA; \textsuperscript{e}Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, PR China

Key Words
Generalized sequential Bonferroni procedure · Genome-wide association studies · Hardy-Weinberg disequilibrium · Multiple testing · Smoothed weights

Abstract

Background/Objectives: For genome-wide association studies (GWAS) with case-control designs, one of the most widely used association tests is the Cochran-Armitage (CA) trend test assuming an additive mode of inheritance. The CA trend test often has higher power than other association tests under additive and multiplicative disease models. However, it can have very low power under a recessive disease model in GWAS. Although tests (such as MAX3) robust to different genetic models have been developed, they often have relatively lower power than the CA trend test under additive and multiplicative models. The goal of this study is to propose an efficient method that not only has higher power than the CA trend test under dominant and recessive models but also maintains the power of the CA trend test under additive and multiplicative models. Methods: We employed the generalized sequential Bonferroni (GSB) procedure of Holm to incorporate information from a Hardy-Weinberg disequilibrium (HWD) test into the CA trend test based on estimating weights from the \( p \) values of the HWD test. We proposed to smooth the weights to reduce possible noise. Results and Conclusions: Results from extensive simulation studies showed that the proposed GSB procedure can achieve the goal described above.

1. Introduction

With the availability of a large number of datasets of genome-wide single-nucleotide polymorphisms (SNPs) for different human diseases, genome-wide association studies (GWAS) have become a useful approach to the identification of genes involved in common human diseases. For GWAS with case-control designs, one of the...
most widely used association tests is the Cochran-Armitage (CA) trend test \([1-3]\), which assumes an additive mode of inheritance. The CA trend test can control the type I error rate when there is departure from Hardy-Weinberg equilibrium (HWE) in the general population from which cases and controls are sampled \([3]\) and can have higher power for additive and multiplicative disease models than other association tests such as the genotype association test using the Pearson’s chi-square test statistic. However, this method can have very low power under a recessive disease model (e.g., \([4]\))


defined under recessive, additive, and dominant modes of inheritances have optimal power when the true underlying disease models are recessive, additive, and dominant, respectively \([9]\). (As stated earlier, the CA trend test statistic assuming additive genotypic values is widely used for GWAS.) The MAX3 of Freidlin et al. \([9]\) is the maximum of the absolute values of the three CA trend test statistics with genotypic values defined under recessive, additive, and dominant models, respectively. Since the asymptotic distribution of MAX3 is very complicated, approximation methods have been developed to estimate p values for MAX3 (e.g., \([10, 11]\)).

In addition to MAX3, other robust tests, adaptive approaches based on genetic model selection (GMS) and genetic model exclusion (GME), have been proposed under the assumption of HWE in the general population \([4, 11, 12]\). These adaptive approaches first select (or exclude) the underlying genetic model by using an HWD trend test of Song and Elston \([6]\) and then apply a CA trend test (or a linear combination of two CA trend tests), which uses genotypic values corresponding to the selected underlying model(s), to analysis of real data. MAX3 and the adaptive approaches have power robust to different genetic models. However, they can be time-consuming when applied to GWAS and can have relatively lower power than the widely used CA trend test statistic (assuming additive genotypic values) when underlying disease models are additive or multiplicative \([13]\), especially for small sample sizes. For example, from the simulation studies of Joo et al. \([13]\) with 500 cases and 500 controls, when the disease locus followed a multiplicative model, the genetic relative risk for the wild homozygote was 1.5, and the minor allele frequency was 0.1, the power of the CA trend test was 0.35, which is 18.6% higher than the power (0.295) of MAX3, in relative pattern.

Because the adaptive genetic model is often assumed for GWAS in the literature and most identified genetic variants follow an additive mode of inheritance, in the present study, we aimed to develop a computationally efficient method for GWAS that not only has increased power under dominant and recessive disease models in comparison to the CA trend test but also maintains the power of the CA trend test under additive and multiplicative models. Specifically, we estimated a weight for each marker by using the p value of an HWD test using cases only \([5]\), and then we employed the generalized sequential Benferroni (GSB) procedure of Holm \([14]\) to adjust the p values from the CA trend test by the weights. This step is to achieve higher power than the CA trend test under dominant and recessive models. Further, we proposed to...
smooth the weights used in the GSB procedure to make the power of the GSB procedure comparable to the CA trend test under additive and multiplicative models. Extensive simulation studies were performed to evaluate the GSB procedure and to compare it with several existing methods. Results from the simulation studies showed that in comparison to the CA trend test, the proposed GSB procedure had much higher power under a recessive model and could have higher power under a dominant model, and that the proposed procedure had power comparable to the CA trend test under additive and multiplicative models. Additionally, the results showed that the GSB procedure was robust to the departure from HWE in the general population. We applied the GSB procedure to the analysis of the genome-wide case-control data for seven complex diseases from the Wellcome Trust Case Control Consortium [15].

2. Methods

2.1. Notation

Herein we use notations similar to those of Zheng and Ng [4]. We consider a genetic case-control design with r cases and s controls genotyped at m independent SNP markers across the whole genome; therefore, n = r + s represents the total number of individuals in the case-control sample. We assume there are no genotyping errors in the sample. For a biallelic SNP marker with alleles A and a, where A is the minor allele, let gᵢ denote a genotype at the marker, where i denotes the number of minor alleles, and (gᵢ, gᵢ, gᵢ) = (aa, Aa, AA). We assume that cases and controls are sampled from a homogeneous general population in which HWE holds and no stratification exists. In this article, fᵢ, fᵢ, and fᵢ denote the frequencies of genotypes gᵢ in the case population, control population, and the general population, respectively. Similarly, fᵢ, fᵢ, and fᵢ denote the frequencies of allele A in the same three populations, respectively.

2.2 Association Tests for a Single Marker

Here we review two association tests for single marker analysis: the Cochran-Armitage trend test [1–3], and the HWD test using cases only [5]. In our proposed method, we used the HWD test among cases instead of the HWD trend test of Song and Elston [6], which compares the HWD coefficients between cases and controls, because our simulation studies suggested that the HWD test using cases only has slightly higher power than the HWD trend test when HWE holds and no stratification exists in the general population.

2.2.1. Cochran-Armitage Trend Test

Zheng and Ng [4] described the CA trend test statistic based on Sasieni [3] as

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

(1)

where \((x_0, x_1, x_2) = (0, 1, 1)\) are prespecified values for the three genotypes (aa, Aa, AA); \(r_i\) and \(s_i\) denote the observed counts of genotype \(g_i\) in cases and controls, respectively, and \(n_i = r_i + s_i\) is the number of individuals with genotype \(g_i\) in the case-control sample. Under the null hypothesis of no association between a marker and the disease, \(Z_t\) asymptotically follows a standard normal distribution N(0, 1), and \(Z_{t+12}\) asymptotically follows a \(\chi^2\) distribution with one degree of freedom. For recessive, additive (or multiplicative), and dominant models, the optimal choices of \(x\) are 0, ½, and 1, and the corresponding statistics are denoted as \(Z_0\), \(Z_{12}\), and \(Z_1\), respectively.

Because the underlying disease model is unknown in reality, the additive model assuming additive genotypic values (\(x = ½\)) is often used in practice. As stated earlier, in this article, the CA trend test statistic always denotes the statistic assuming additive genotypic values (\(x = ½\)) if other types of genotypic values are not arbitrarily assumed. We use \(T_{trend} = Z_{12}\) to denote the CA trend test statistic with \(x = ½\). The statistic \(T_{trend}\) can have high power when the underlying genetic model is additive or multiplicative, but it has very low power when the underlying genetic model is recessive.

2.2.2. HWD Test Statistic

The HWD test among cases has been used to test the genetic association between markers and a disease (e.g., [5]) when assuming HWE in the general population. For testing HWD at a marker locus among cases, the test statistic [5, 16] based on comparing the observed counts of genotypes among cases with the corresponding expected counts can be written as

\[ T_{HWD} = \frac{\hat{\Delta}_D^2}{\text{var}(\hat{\Delta}_D)} \]  

(2)

where \(\hat{\Delta}_D = f_D^A - (f_D^A)^2\) is the estimate of the HWD coefficient in cases (\(\Delta_D = f_D^A - (f_D^A)^2\)) and \(\text{var}(\hat{\Delta}_D) = (f_D^A + f_D^{AA})^2(f_D^{AA} + f_D^{AA})^2r\), where hat notation denotes the estimate of the variable from the data. Under the null hypothesis of no association between the marker and the disease, if HWE holds in the general population, then \(\Delta_D = 0\) and \(T_{HWD}\) asymptotically follows a \(\chi^2\) distribution with one degree of freedom.

This HWD test can provide valuable information that can be complementary to or different from that provided by the CA trend test for detecting genetic association when the underlying genetic disease model is dominant or recessive [17] but it provides little and no information under the additive and multiplicative disease models, respectively [5]. However, the HWD test statistic \(T_{HWD}\) in (2) can have an inflated type I error rate when the minor allele frequency is low [18]. To control the type I error rate of the HWD test for a marker with low frequencies, a chi-square statistic corrected for continuity [19] can be used. A better choice is the HWD exact test [16]. Wigginton et al. [18] proposed a fast approach for the HWD exact test that can be applied to GWAS.

2.3. Multiple Testing for Genome-Wide Association Studies

GWAS involve multiple hypothesis testing, i.e., testing hundreds of thousands (or millions) of SNPs simultaneously. Here, we introduce some concepts on multiple hypothesis testing in the context of GWAS. Let \(p_1, p_2, ..., p_m\) be \(p\) values of association tests for \(m\) independent markers corresponding to \(m\) null hypotheses of \(H_1, H_2, ..., H_m\), where \(H_i\) is the null hypothesis of no association...
between a marker and a disease. The commonly used criterion for error control in multiple testing is the family-wise error rate (FWER), which is the probability of rejecting at least one true null hypothesis (e.g., [20]). The power of a multiple testing procedure can be measured by average power, taken to be the average power of individual tests corresponding to all false null hypotheses [21].

In GWAS, a traditional association test (such as the CA trend test) is often used to test each single marker for its association with a disease. To control the FWER for multiple testing, the significant threshold for each single marker is often adjusted using the Bonferroni procedure. In this study we use the GSB procedure of Holm [14] to control the FWER for GWAS, and more importantly, to adjust p values from the CA trend test by weights calculated from the p values of the HWD test using cases only.

2.3.1. Generalized Sequential Bonferroni Procedure

Here we describe the GSB procedure of Holm [14] in the context of GWAS. Let \( p_1, p_2, \ldots, p_m \) denote the \( m \) independent markers calculated by the CA trend test, where the \( c \) superscript simply denotes the CA trend test. Given weights \( w_1, w_2, \ldots, w_m \) for the tests with hypotheses \( H_1, H_2, \ldots, H_m \) (where we can let \( w_i = 1 \) or a constant for all \( i \) if no weights are provided), define B values as \( B_i = p_i / w_i \). Let \( B(1) \leq B(2) \leq \ldots \leq B(m) \) be the ordered B values, with \( w(1), w(2), \ldots, w(m) \) being the weights, and \( H(1), H(2), \ldots, H(m) \) being the hypotheses corresponding to the ordered B values. The GSB procedure can be described as follows (see also [20]): starting from \( i = 1 \), given \( H(1), H(2), \ldots, H(i-1) \) have been tested and rejected, if

\[
B(i) \leq \alpha / \sum_{k=1}^{i} w(k) ,
\]

reject \( H(i) \) otherwise, accept \( H(i), H(i+1), \ldots, H(m) \), and stop the GSB procedure. The GSB procedure controls the FWER at the nominal significance level of \( \alpha \) if the weights are independent of the CA trend tests. From this procedure, we can see that to reject the hypothesis \( H(i) \), the corresponding p value from the CA trend test \( p(i) \) must satisfy

\[
p(i) \leq w(i) \alpha / \sum_{k=1}^{i} w(k) \leq \alpha .
\]

In other words, if \( p(i) > \alpha \), irrespective of the size of the corresponding weight \( w(i) \), the GSB procedure will not reject the hypothesis \( H(i) \).

In the GSB procedure, if the weights are estimated appropriately by use of prior information, the procedure can have higher power than the corresponding Bonferroni procedure [14, 20]. Unfortunately, Holm [14] did not provide a method to determine weights for a GSB procedure.

2.3.2. Estimating Weights by Use of Information on HWD Among Cases

Here we propose a method to estimate weights for the GSB procedure in the context of GWAS by using information on HWD among cases. Let \( p^c_1, p^c_2, \ldots, p^c_m \) be the p values from HWD tests for \( m \) markers among cases based on equation (2). We define weight \( w_i \) for the \( i \)-th marker as \( w_i = 1 / p^c_i \), \( 1 \leq i \leq m \), and therefore the B value for the \( i \)-th marker \( B_i = p_i / (1/p^c_i) = p_i p^c_i \). We note that the Fisher’s method (online suppl. Appendix A, karger.com/doi/10.1159/000332916) is based on considering the null distribution of the logarithm of these B values. We call the GSB procedure with these weights \( B \), the GSB-HWD procedure.

To control the FWER of the GSB-HWD procedure, we need to prove that the HWD test statistic \( T_{HWD} \) and the CA trend test statistic \( T_{trend} \) at a marker are independent under the null hypothesis of no association in order to guarantee that the weights are calculated by using information independent of the CA trend tests. Zheng and Ng [4] proved that when HWE holds in the general population, the two statistics are asymptotically uncorrelated, i.e., the correlation coefficient of \( T_{trend} \) and \( T_{HWD} \) equals zero, under the null hypothesis. In this article we further prove that the two statistics \( T_{trend} \) and \( T_{HWD} \) are asymptotically independent (see online suppl. Appendix A) by demonstrating the vector \( (T_{trend}, T_{HWD}) \) follows a bivariate normal distribution when HWE holds in the general population.

2.3.3. GSB-HWD Procedure with Smoothed Weights

As will be shown in the simulation study section, in GWAS, the GSB-HWD procedure can have much higher power than the CA trend test (with Bonferroni correction) under the dominant and recessive models. However, like Fisher’s method, the GSB-HWD procedure has lower power than CA trend test under the additive and multiplicative models because in this situation the weights from the HWD test provides essentially no information but simply contributes noise to the genetic association studies.

The additive model is often assumed for GWAS in the literature when the true underlying genetic model is unknown. Therefore, we hope to modify the GSB-HWD procedure so that it can maintain the power of the CA trend test under the additive and multiplicative models, but still have increased power under the dominant and recessive disease models. We employ a smoothing method of Roeder et al. [22] to achieve this goal. We calculate a smoothed weight \( w_i' \) for each marker as a linear function of the original weight \( w_i \) and the average value \( \bar{w} \) of the weights for all markers. Therefore, \( w_i' = (1 - \lambda) w_i + \lambda \bar{w} \), where

\[
\bar{w} = \frac{1}{m} \sum_{i=1}^{m} w_i,
\]

and \( \lambda \) is a parameter that represents the level of smooth \( (0 \leq \lambda \leq 1) \). If \( \lambda = 1 \), then all weights \( w_i \) equal the constant \( \bar{w} \), no weight information from the HWD test is used, and the GSB-HWD procedure is approximately equivalent to the CA trend test and has the highest power for additive and multiplicative models. If \( \lambda = 0 \), then \( w_i' = w_i \), i.e., the weights are not smoothed, and the GSB-HWD procedure will have the highest power for recessive and dominant models. Roughly speaking, the larger the value of \( \lambda \), the higher the power of the GSB-HWD procedure will have for the additive and multiplicative models and the lower the power for recessive and dominant models. A problem is how to determine the optimal value of \( \lambda \). Since we hope to preserve the power for additive and multiplicative models, we suggest using \( \lambda \geq 0.6 \). In this paper, we used \( \lambda = 0.6 \). For simplicity, we call the GSB-HWD procedure using the smoothed weights the S-GSB-HWD procedure.

3. Simulation Design

To evaluate our proposed methods, we compared the FWER and average power of our methods (GSB-HWD procedure and S-GSB-HWD procedure) with those of several existing methods.
(HWD test, CA trend test, Fisher’s method and MAX3) by simulation studies. As in other previous studies [4, 23], we used Wright’s coefficient of inbreeding \( F \) to measure the deviation from HWE in the general population. In our simulation studies, we considered two scenarios: (1) HWE held \( (F = 0) \) in the general population, and (2) HWE was slightly violated (we considered \( F = 0.005 \) and 0.01).

Given the inbreeding coefficient \( F \) and frequency \( f_A \) of the minor allele \( A \) at an SNP marker in the general population, the genotype frequencies \( f_{AA} \) at the SNP marker in the population can be computed as [23]

\[
\begin{align*}
    f_{AA} &= (1 - f_A)^2 + 2Ff_A(1 - f_A), \\
    f_{Aa} &= 2(1 - F)f_A(1 - f_A), \\
    f_{aa} &= F^2 f_A^2 + 2F^2 f_A^2 (1 - f_A). \\
\end{align*}
\]

We considered two case-control designs: (1) 1,000 cases and 1,000 controls with 1,000 SNPs. This design corresponds to an association study design for candidate genes. (2) 2,000 cases and 3,000 controls with 400,000 SNPs. This design is based on the GWAS data from the Wellcome Trust Case Control Consortium (WTCCC) [15]. Our simulated datasets were generated using a process similar to that described by Joo et al. [13].

3.1. Evaluation of the FWER

To estimate the FWER for the methods being compared, we simulated 100,000 replicated datasets for each case-control design under the assumption that none of the \( m \) SNPs was associated with a disease, where \( m \) is the number of SNPs. We randomly sampled MAF \( f \) for the general population at each of the \( m \) SNPs from a uniform distribution \( U[0.1, 0.50] \) independently. To generate genotypes at an SNP for \( r \) cases and \( s \) controls, we followed a process of Joo et al. [13] and randomly generated genotype counts \( (r_{aa}, r_{Aa}, r_{AA}) \) for the cases, where \( r = r_{aa} + r_{Aa} + r_{AA} \), from the multinomial distribution with parameter \( r \) and \( (f_{aa}, f_{Aa}, f_{AA}) \), where \( f_{aa}, f_{Aa}, \) and \( f_{AA} \) are calculated by equation (3). We then randomly assign the genotypes \( aa, Aa, \) and \( AA \) to \( r_{aa}, r_{Aa}, \) and \( r_{AA} \) individually among the \( r \) cases, respectively. Similarly, we randomly generate genotype counts \( (s_{aa}, s_{Aa}, s_{AA}) \) and assigned genotypes to the \( s \) controls by using the multinomial distribution with parameter \( s \) and \( (f_{aa}, f_{Aa}, f_{AA}) \), where \( s = s_{aa} + s_{Aa} + s_{AA} \). FWERs were estimated as the proportion of 100,000 replicated datasets in which at least one of the \( m \) null hypotheses was rejected.

3.2. Evaluation of Power

For each case-control design, we chose 10 SNPs as disease SNPs (i.e., SNPs associated with the disease) with independent genetic effects. We set all disease SNPs to have the same MAFs in the general population. We considered three values for the MAFs: 0.1, 0.3 and 0.5. At each of the remaining non-disease SNPs, we generated genotypes for cases and controls in the same way as described in the previous section (on evaluation of FWER) for generating genotypes under the null hypothesis. Here we describe how we generate genotype data for each disease SNP. Let \( \psi_g = Pr(\text{disease} | g) \) be the penetrance of genotype \( g \) (\( g = AA, Aa, aa \)) at an SNP marker in the general population. Let \( \gamma_1 = \psi_{aA}/\psi_{aa} \) and \( \gamma_2 = \psi_{AA}/\psi_{aa} \) be the genotype relative risks (GRRs) [24]. For the additive, multiplicative, recessive, and dominant genetic models, \( \gamma_1 \) and \( \gamma_2 \) satisfy the equations \( \gamma_2 = 2\gamma_1 - 1 \), \( \gamma_2 = \gamma_1 \), \( \gamma_1 = 1 \) and \( \gamma_1 = \gamma_2 \), respectively. Therefore, in our simulation studies, we used \( \gamma_1 \) as the GRR parameter for the additive, multiplicative and dominant models and \( \gamma_2 \) for the recessive model. For each replicated dataset, we set all disease SNPs to have the same GRR values.

Given the population prevalence \( (k) \), the inbreeding coefficient \( F \), population MAF \( f_A \), and GRRs under a disease model, the penetrances \( \psi_g \) can be written as [25]: \( \psi_{aa} = k / (2\gamma_A^2 + 2\gamma_f A(1 - f_A) + (1 - f_A)^2), \psi_{aA} = \psi_{aa} \gamma_1, \psi_{AA} = \psi_{aa} \gamma_2 \). The frequencies of genotype \( g \) among cases and controls for a disease SNP can be calculated by \( f_g^C = Pr(g | \text{disease}) = \psi_g f_A / k \), and \( f_g^C = Pr(g | \text{non-disease}) = (1 - \psi_g) f_A / (1 - k) \), where \( f_A \) is genotype frequency in the general population as described above in equation (3). We randomly generated genotypes at each disease SNP marker for \( r \) cases and \( s \) controls by using the multinomial distributions \( Mult(r, f_{aa}^C, f_{Aa}^C, f_{AA}^C) \) and \( Mult(s, f_{aa}^C, f_{Aa}^C, f_{AA}^C) \), respectively, in a similar way as described in the previous section on evaluating the FWER.

For each case-control design and each set of parameters, we estimated the average power of each method as the average power of individual tests for the 10 disease SNPs based on 1,000 replicated datasets.

3.3. Applications to GWAS of WTCCC Data

The WTCCC [15] conducted GWAS for seven complex human diseases. The dataset for each disease consists of about 2,000 cases and 3,000 shared controls (genotyped with the Affymetrix 500K SNP chip). After quality control filtering, less than 400,000 SNPs that mapped on the autosomes remained in the dataset for each disease. In this article, we considered only the analysis of autosomal SNPs. By using the threshold \( \alpha = 5 \times 10^{-7} \) for each single SNP, the WTCCC identified a total of 21 disease-susceptible regions with significant SNPs for the seven diseases by either the genotype association test (GAT) using Pearson’s chi-square test statistic or the CA trend test (see online suppl. table 5). To make comparisons with these results at the same level of threshold, when we applied the GSB procedures to \( m \) autosomal SNPs in each dataset, we used a nominal FWER (overall significance) level of \( 5 \times 10^{-7} \cdot m \).

4. Results

We applied the six methods (HWD test, CA trend test, Fisher’s method, GSB-HWD procedure, S-GSB-HWD procedure and MAX3) to the simulated datasets to evaluate FWERs and power. For MAX3, we used the asymptotic distribution approach as of Zang et al. [11] implemented in the R package Rassoc to estimate the \( p \) values.

4.1. Evaluation of FWER

The empirical FWER at the nominal significance levels of 0.01, 0.05, and 0.1, for the two case-control designs based on 100,000 replicates assuming that no SNPs are associated with the disease are reported in table 1. When HWE holds \( (F = 0) \) or is slightly violated \( (F = 0.005, 0.01) \) in the general population, we can see that CA trend test, MAX3 and our two GSB-HWD procedures controlled
the FWER quite well. The HWD method had an inflated FWER, especially when HWE did not hold. Fisher’s method had a slightly inflated FWER when HWE held and had FWERs much higher than the corresponding nominal levels when HWE did not hold. To control the FWER, our S-GSB-HWD procedure requires only that the weights estimated from the HWD test are independent of the CA trend test, no matter whether the HWD test controls the FWER or not (see also [14]).

4.2. Power Comparisons

Since the HWD test using cases only and Fisher’s method can have inflated FWERs, especially when HWE does not hold in the general population, we only listed the power of these methods in the tables when HWE holds in the general population to show how much information the HWD test can provide and to show that Fisher’s method has lower power than the CA trend test under the additive and multiplicative models.

Results when HWE Held in the General Population

When HWE holds in the general population (F = 0), the estimated average power of the methods being compared is reported in tables 2–5. From tables 2 and 3, we can see that under the additive and multiplicative models, MAX3 always had lower power than the CA trend test, especially for small sample size, whereas our S-GSB-HWD procedure (using smoothed weights) had power always very close to that of the CA trend test. For example, for the case-control design with 1,000 SNPs, when GRR = 1.25 (i.e., $\gamma_1 = 1.25, \gamma_2 = 1.5$) and MAF = 0.3, if the disease SNPs followed the additive model, the power of the CA trend test, our S-GSB-HWD, and MAX3 was 0.2732, 0.2612 and 0.2273, respectively (table 2); the power of MAX3 is 16.8% lower than that of CA trend test, in the relative pattern. From table 4, we can see that under the recessive model, our S-GSB-HWD procedure had much higher power than the CA trend test; we also noted that MAX3 had higher power than our S-GSB-HWD procedure. For example, when GRR = 1.6 (i.e., $\gamma_1 = 1.0, \gamma_2 = 1.6$), and MAF = 0.3, for

| Table 1. Empirical FWERs based on 100,000 replicates (prevalence $k = 0.1$) |
| Methods | $F^a$ | $n_d = 1,000, n_c = 1,000, m = 1,000^c$ | $n_d = 2,000, n_c = 3,000, m = 400,000^d$ |
| | 0.01^b | 0.05 | 0.1 | 0.05 | 0.1 | 0.2 | 0.05 | 0.1 | 0.2 |
| HWD | 0 | 0.0174 | 0.0751 | 0.1393 | 0.0696 | 0.1264 | 0.2270 |
| | 0.005 | 0.0278 | 0.1112 | 0.1976 | 0.1804 | 0.3003 | 0.4743 |
| | 0.01 | 0.0514 | 0.1883 | 0.3163 | 0.4501 | 0.6500 | 0.8408 |
| CATT | 0 | 0.0086 | 0.0444 | 0.0857 | 0.0500 | 0.0953 | 0.1764 |
| | 0.005 | 0.0085 | 0.0436 | 0.0848 | 0.0501 | 0.0961 | 0.1784 |
| | 0.01 | 0.0087 | 0.0455 | 0.0882 | 0.0502 | 0.0966 | 0.1789 |
| Fisher | 0 | 0.0100 | 0.0510 | 0.0993 | 0.0503 | 0.1013 | 0.1957 |
| | 0.005 | 0.0132 | 0.0610 | 0.1171 | 0.0992 | 0.1821 | 0.3092 |
| | 0.01 | 0.0215 | 0.0915 | 0.1696 | 0.2340 | 0.3788 | 0.5772 |
| GSB-HWD | 0 | 0.0070 | 0.0377 | 0.0755 | 0.0444 | 0.0911 | 0.1793 |
| | 0.005 | 0.0059 | 0.0309 | 0.0632 | 0.0405 | 0.0854 | 0.1670 |
| | 0.01 | 0.0052 | 0.0278 | 0.0583 | 0.0391 | 0.0816 | 0.1618 |
| S-GSB-HWD | 0 | 0.0080 | 0.0408 | 0.0818 | 0.0475 | 0.0951 | 0.1780 |
| | 0.005 | 0.0074 | 0.0377 | 0.0758 | 0.0454 | 0.0920 | 0.1749 |
| | 0.01 | 0.0070 | 0.0377 | 0.0751 | 0.0438 | 0.0909 | 0.1731 |
| MAX3 | 0 | 0.0079 | 0.0449 | 0.0888 | 0.0500 | 0.0884 | 0.1856 |
| | 0.005 | 0.0084 | 0.0450 | 0.0877 | 0.0492 | 0.0956 | 0.1900 |
| | 0.01 | 0.0083 | 0.0454 | 0.0905 | 0.0420 | 0.0956 | 0.1892 |

HWD, CATT, and Fisher denote HWD test among cases, CA trend test, and Fisher’s method, respectively. Each of these methods was applied to the simulated data with Bonferroni correction. S-GSB-HWD = GSB-HWD procedure using smoothed weights with $\lambda = 0.6$. Wright’s coefficient of inbreeding. Overall significance level (nominal FWER). Case-control design with 1,000 cases, 1,000 controls, and 1,000 SNPs. 2,000 cases and 3,000 controls with 400,000 SNPs. Bolded numbers denote inflated FWER.
From table 5, we can see that when the underlying disease model is dominant, our S-GSB-HWD procedure had power higher than the CA trend test, except when MAF = 0.1, the two methods had comparable power. We can also see that our S-GSB-HWD procedure had power slightly higher than MAX3 (when MAF is equal to either 0.1 or 0.3) and had power lower than MAX3 (when MAF = 0.5 and GRR >1.35). For example, when GRR = 1.3 (i.e., $\gamma_1 = \gamma_2 = 1.3$), and MAF = 0.3, for the design with 400,000 SNPs, the power of the CA trend test, our S-GSB-HWD procedure and MAX3 was 0.2281, 0.3608 and 0.3384, respectively. As stated earlier, MAX3 had relatively lower power than the CA trend test under the additive and multiplicative models.

### Results when HWE Is Violated in the General Population

When HWE was not held in the general population ($F = 0.005$ and 0.01), the estimated average power of the methods being compared is reported in online supplementary tables 1–4. When $F = 0.005$ and 0.01, in most situations, the power of our S-GSB-HWD method was very close to the corresponding power of the method with $F = 0$ (reported in tables 2–5), except that for the dominant model, the S-GSB-HWD method had decreased power in some cases when $F = 0.01$ (see online suppl. table 4). However, our S-GSB-HWD method still had power higher than the CA trend test under the recessive and dominant models.
4.3. Application to GWAS of the WTCCC Data
We applied our S-GSB-HWD procedure (see online suppl. table 5) to the analyses of the WTCCC data. Since the HWD test using cases only and Fisher’s method cannot control the FWER, and the original GSB-HWD procedure has relatively lower power than the CA trend test under the additive and multiplicative models, we did not apply these three methods to the WTCCC data. The WTCCC group applied the CA trend test and the genotype association test (GAT) using the Pearson’s chi-square test statistic (with a threshold $5 \times 10^{-7}$ for each SNP) to the dataset for each of the seven diseases. The CA trend test identified significant SNPs in 18 regions in total for the seven diseases. All these regions were identified by our S-GSB-HWD procedure using smoothed weights. Our method identified an additional significant SNP rs17042725 for bipolar disease that was not located in the regions identified by WTCCC with the CA trend test or GAT. This SNP may be used for replication or follow-up molecular or bioinformatics studies.

For comparison, we also applied MAX3 to the data (Max was implemented by the asymptotic distribution approach asy of Zang et al. [11]). For Crohn’s disease, with the threshold $5 \times 10^{-7}$ (used by the WTCCC) for single SNPs, the SNP rs10761659 was identified by the CA trend test (p value = 2.677 $\times 10^{-14}$) and our S-GSB-HWD procedure but not by MAX3 (p value = 7.516 $\times 10^{-7}$). This might be because this SNP may follow the additive or multiplicative model, and MAX3 has relatively lower power under the additive and multiplicative models. We also note that some SNPs identified by MAX3 were not identified by the CA trend test and our S-GSB-HWD pro-

<table>
<thead>
<tr>
<th>MAF</th>
<th>GRR</th>
<th>HWD</th>
<th>CATT</th>
<th>Fisher</th>
<th>GSB-HWD</th>
<th>S-GSB-HWD</th>
<th>MAX3</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_1 = 1,000, n_c = 1,000, m = 1,000^a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>1.3</td>
<td>0.0002</td>
<td>0.1244</td>
<td>0.0729</td>
<td>0.0804</td>
<td>0.1136</td>
<td>0.0983</td>
</tr>
<tr>
<td>1.4</td>
<td>0</td>
<td>0.3971</td>
<td>0.2666</td>
<td>0.2915</td>
<td>0.3737</td>
<td>0.3371</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>0</td>
<td>0.7227</td>
<td>0.5877</td>
<td>0.6122</td>
<td>0.7047</td>
<td>0.6642</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>0</td>
<td>0.9245</td>
<td>0.85</td>
<td>0.8632</td>
<td>0.9131</td>
<td>0.8947</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1.2</td>
<td>0</td>
<td>0.1415</td>
<td>0.0784</td>
<td>0.0924</td>
<td>0.1299</td>
<td>0.107</td>
</tr>
<tr>
<td>1.25</td>
<td>0</td>
<td>0.3392</td>
<td>0.2297</td>
<td>0.2518</td>
<td>0.3219</td>
<td>0.2828</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>0</td>
<td>0.6055</td>
<td>0.4594</td>
<td>0.4877</td>
<td>0.5854</td>
<td>0.5401</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>0</td>
<td>0.9379</td>
<td>0.8706</td>
<td>0.8841</td>
<td>0.9283</td>
<td>0.9104</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.2</td>
<td>0</td>
<td>0.1955</td>
<td>0.1143</td>
<td>0.1268</td>
<td>0.1811</td>
<td>0.1505</td>
</tr>
<tr>
<td>1.25</td>
<td>0</td>
<td>0.4393</td>
<td>0.31</td>
<td>0.3357</td>
<td>0.4179</td>
<td>0.3752</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>0</td>
<td>0.7052</td>
<td>0.5644</td>
<td>0.5886</td>
<td>0.6828</td>
<td>0.6379</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>0</td>
<td>0.9673</td>
<td>0.924</td>
<td>0.93</td>
<td>0.9623</td>
<td>0.9499</td>
<td></td>
</tr>
<tr>
<td>n_2 = 2,000, n_c = 3,000, m = 400,000^b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>1.25</td>
<td>0</td>
<td>0.0795</td>
<td>0.042</td>
<td>0.0466</td>
<td>0.0719</td>
<td>0.0665</td>
</tr>
<tr>
<td>1.3</td>
<td>0</td>
<td>0.2386</td>
<td>0.1543</td>
<td>0.162</td>
<td>0.2218</td>
<td>0.2059</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>0</td>
<td>0.7522</td>
<td>0.6339</td>
<td>0.6473</td>
<td>0.7329</td>
<td>0.7075</td>
<td></td>
</tr>
<tr>
<td>1.45</td>
<td>0</td>
<td>0.9149</td>
<td>0.8436</td>
<td>0.8519</td>
<td>0.9046</td>
<td>0.8881</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1.2</td>
<td>0</td>
<td>0.2639</td>
<td>0.1661</td>
<td>0.1757</td>
<td>0.2447</td>
<td>0.2123</td>
</tr>
<tr>
<td>1.25</td>
<td>0</td>
<td>0.4599</td>
<td>0.3241</td>
<td>0.3376</td>
<td>0.4344</td>
<td>0.4034</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>0</td>
<td>0.6624</td>
<td>0.5308</td>
<td>0.5425</td>
<td>0.6434</td>
<td>0.6081</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>0</td>
<td>0.9289</td>
<td>0.8633</td>
<td>0.8719</td>
<td>0.9215</td>
<td>0.9009</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.175</td>
<td>0</td>
<td>0.1778</td>
<td>0.1076</td>
<td>0.1146</td>
<td>0.163</td>
<td>0.1445</td>
</tr>
<tr>
<td>1.2</td>
<td>0.0001</td>
<td>0.3646</td>
<td>0.2408</td>
<td>0.2516</td>
<td>0.3405</td>
<td>0.3135</td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>0</td>
<td>0.778</td>
<td>0.6554</td>
<td>0.668</td>
<td>0.7584</td>
<td>0.7294</td>
<td></td>
</tr>
<tr>
<td>1.275</td>
<td>0</td>
<td>0.8983</td>
<td>0.8218</td>
<td>0.8293</td>
<td>0.8882</td>
<td>0.8703</td>
<td></td>
</tr>
</tbody>
</table>

MAF = Minor allele frequency; GRR = genotype relative risk $\gamma_i$; S-GSB-HWD = GSB-HWD procedure using smoothed weights with $\lambda = 0.6$.

^a 1,000 cases and 1,000 controls with 1,000 SNPs. ^b 2,000 cases and 3,000 controls with 400,000 SNPs.
procedure. In addition, MAX3 took a much longer time than the other methods for the GWAS analysis of the WTCCC data. For example, for the GWAS analysis of the coronary artery disease dataset, MAX3 took about 566 minutes whereas the CA trend test and our S-GSB-HWD procedure each took less than one minute. We used a Beowulf computer cluster [Intel® Xeon® X5550, 2.66 GHz, 8 M cache, turbo, HT, 32 GB memory (8 × 4 GB), 1333 MHz dual ranked RDIMMs for 2 processors, advanced ECC].

### 5. Discussion

We have proposed a simple and fast S-GSB-HWD procedure that uses smoothed weights for GWAS. The S-GSB-HWD procedure is based on application of the generalized sequential Bonferroni procedure of Holm to the incorporation of information of the HWD test using cases only into the CA trend test. When HWE holds or is only slightly violated in the general population, our S-GSB-HWD procedure can control the FWER very well. Compared to the widely used CA trend test (which assumes an additive mode of inheritance), our S-GSB-HWD procedure has much higher power for the recessive disease model and higher or comparable power for the dominant model but also roughly maintains the power of the CA trend test for the additive and multiplicative models. In addition, compared to the CA trend test, for case-control data with larger sample sizes, our S-GSB-HWD procedure can gain much more increased power under the recessive and dominant disease models compared to the CA trend test (tables 2–5). Therefore, our proposed

### Table 4. Empirical average power (for 10 disease SNPs in the data) based on 1,000 replicates (data were simulated under the recessive model assuming Hardy-Weinberg equilibrium with prevalence k = 0.1, nominal FWER level α = 0.05)

<table>
<thead>
<tr>
<th>MAF</th>
<th>GRR</th>
<th>HWD</th>
<th>CATT</th>
<th>Fisher</th>
<th>GSB-HWD</th>
<th>S-GSB-HWD</th>
<th>MAX3</th>
</tr>
</thead>
<tbody>
<tr>
<td>n_d = 1,000, n_c = 1,000, m = 1,000^a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>2.5</td>
<td>0.3795</td>
<td>0.0049</td>
<td>0.3784</td>
<td>0.0637</td>
<td>0.0313</td>
<td>0.0583</td>
</tr>
<tr>
<td>3.0</td>
<td>0.6898</td>
<td>0.0145</td>
<td>0.7193</td>
<td>0.1328</td>
<td>0.0614</td>
<td>0.2119</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>0.8955</td>
<td>0.0437</td>
<td>0.9185</td>
<td>0.2701</td>
<td>0.1121</td>
<td>0.4959</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>0.9734</td>
<td>0.095</td>
<td>0.9837</td>
<td>0.4251</td>
<td>0.1741</td>
<td>0.7561</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1.5</td>
<td>0.1205</td>
<td>0.0343</td>
<td>0.2626</td>
<td>0.1654</td>
<td>0.1004</td>
<td>0.1284</td>
</tr>
<tr>
<td>1.6</td>
<td>0.2431</td>
<td>0.0778</td>
<td>0.4879</td>
<td>0.3268</td>
<td>0.1834</td>
<td>0.2817</td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td>0.4183</td>
<td>0.1528</td>
<td>0.7282</td>
<td>0.541</td>
<td>0.2828</td>
<td>0.4928</td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>0.5845</td>
<td>0.2644</td>
<td>0.8794</td>
<td>0.7634</td>
<td>0.4067</td>
<td>0.6849</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.3</td>
<td>0.0188</td>
<td>0.0505</td>
<td>0.1263</td>
<td>0.118</td>
<td>0.0935</td>
<td>0.0892</td>
</tr>
<tr>
<td>1.4</td>
<td>0.0736</td>
<td>0.1739</td>
<td>0.4173</td>
<td>0.3971</td>
<td>0.3028</td>
<td>0.317</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>0.1808</td>
<td>0.4085</td>
<td>0.7465</td>
<td>0.7308</td>
<td>0.5502</td>
<td>0.6281</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>0.3353</td>
<td>0.657</td>
<td>0.9284</td>
<td>0.9298</td>
<td>0.7651</td>
<td>0.8591</td>
<td></td>
</tr>
<tr>
<td>n_d = 2,000, n_c = 3,000, m = 400,000^b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>2.4</td>
<td>0.3816</td>
<td>0.0011</td>
<td>0.4472</td>
<td>0.1723</td>
<td>0.0696</td>
<td>0.1411</td>
</tr>
<tr>
<td>2.6</td>
<td>0.5767</td>
<td>0.0032</td>
<td>0.663</td>
<td>0.2961</td>
<td>0.1019</td>
<td>0.2819</td>
<td></td>
</tr>
<tr>
<td>2.8</td>
<td>0.7492</td>
<td>0.0054</td>
<td>0.8265</td>
<td>0.4209</td>
<td>0.1224</td>
<td>0.4749</td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>0.861</td>
<td>0.0139</td>
<td>0.9199</td>
<td>0.5367</td>
<td>0.155</td>
<td>0.669</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1.4</td>
<td>0.03</td>
<td>0.0085</td>
<td>0.1404</td>
<td>0.135</td>
<td>0.0936</td>
<td>0.0844</td>
</tr>
<tr>
<td>1.5</td>
<td>0.1143</td>
<td>0.0392</td>
<td>0.4497</td>
<td>0.4328</td>
<td>0.2732</td>
<td>0.3091</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>0.3047</td>
<td>0.1299</td>
<td>0.7838</td>
<td>0.7719</td>
<td>0.4486</td>
<td>0.6366</td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td>0.549</td>
<td>0.2973</td>
<td>0.952</td>
<td>0.9516</td>
<td>0.5803</td>
<td>0.8845</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>1.3</td>
<td>0.0075</td>
<td>0.0638</td>
<td>0.2222</td>
<td>0.2271</td>
<td>0.1862</td>
<td>0.1866</td>
</tr>
<tr>
<td>1.35</td>
<td>0.027</td>
<td>0.1662</td>
<td>0.4813</td>
<td>0.4898</td>
<td>0.4101</td>
<td>0.4055</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>0.0607</td>
<td>0.3429</td>
<td>0.7413</td>
<td>0.7484</td>
<td>0.6415</td>
<td>0.6676</td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>0.2064</td>
<td>0.7441</td>
<td>0.9727</td>
<td>0.9739</td>
<td>0.9009</td>
<td>0.9522</td>
<td></td>
</tr>
</tbody>
</table>

- MAF = Minor allele frequency; GRR = genotype relative risk \( \gamma_2 \); S-GSB-HWD = GSB-HWD procedure using smoothed weights with \( \lambda = 0.6 \).
- ^a1,000 cases and 1,000 controls with 1,000 SNPs.
- ^b2,000 cases and 3,000 controls with 400,000 SNPs.
method is more suitable for data with large sample sizes (such as \( \geq 1,000 \) cases and \( \geq 1,000 \) controls).

In the S-GSB-HWD procedure, the smoothed weights depend on the smooth parameter \( \lambda \); how to determine the optimal value of \( \lambda \) is an open question. Based on our simulation studies, it appears the optimal value of \( \lambda \) does not depend on the number of SNPs but only depends on the underlying disease models. When the underlying disease models are unknown, setting \( \lambda = 0.6 \) is a good choice.

We note that when the MAF is low (such as \( \leq 10\% \)), in comparison to the CA trend test, our S-GSB-HWD procedure does not have much increased power for all GRR values under the dominant model and for GRR \( \leq 2.0 \) under the recessive model because the procedure does not gain much information from the HWD test using cases only.

From section 2.3.1. on the GSB procedure, we know that one required condition to reject the null hypothesis of no association at a marker is that the p value from the CA trend test \( (p^c) \) is less than the nominal FWER level \( \alpha \) (such as 0.05), i.e., \( p^c < \alpha \) (this means that the marker has at least some weak association signal). In other words, if \( p^c \geq \alpha \), no matter how small the p value from the HWD test (or how big the weight for the marker), the S-GSB-HWD cannot reject the null hypothesis. In some situations, genotyping error can cause very small p values of the HWD test using cases only. This feature of the GSB procedure described here can prevent some false positive findings due to the very small p values caused by genotyping errors.
In our S-GSB-HWD procedure, we used p values from the HWD test statistic \( T_{HWD} \) by using cases only to estimate weights. When HWE holds or is slightly violated, although the HWD test cannot control the FWER for multiple testing, this does not influence that the S-GSB-HWD procedure controls the FWER well. To control the FWER, our S-GSB-HWD procedure requires only that the weights are independent of the p values from the CA trend test. For data with a small sample size or with very small MAF (such as <10%), it might be better to use the HWD exact test statistic \([18]\) to replace the traditional HWD test statistic \( T_{HWD} \) for more accurate results. When HWE holds in the population, the p values from the exact test may not follow a uniform distribution \([0, 1]\) under the null hypothesis of no association \([18, 23]\) depending on allele frequencies and sample sizes. Therefore these p values cannot be used in Fisher’s method, which requires that p values from each test follow \([0, 1]\) distribution. However, the p values from the HWD exact test can be used to estimate weights in our S-GSB-HWD procedure. We also note that when HWE holds in the general population, if the sample size is small (such as with 1,000 cases), Fisher’s method often has slightly violated type I error rate (see our simulation results). When HWE is violated in the general population, Fisher’s method has type I error rate much higher than the nominal levels.

In the S-GSB-HWD procedure, we calculate weights on the basis of reciprocal function of p values from the HWD test. We did test other functions, such as \(-\log, \text{square root, and entropy of the p values. It seems that the reciprocal function is optimal because it generated the highest power in our simulation studies. When using the reciprocal function of p values in the S-GSB-HWD procedure, we essentially compare the product of a p value from the HWD test and a p value from the CA trend test with the corresponding threshold. We note that Fisher’s method is based on using the distribution of the logarithm of this product. However Fisher’s method cannot control the FWER well while our S-GSB-HWD procedure can.}

Our S-GSB-HWD procedure is based on the assumptions of independence, i.e., absence of linkage disequilibrium between markers, and cannot handle LD among markers. Wu et al. [26] proposed an SNP-set analysis method which can account for LD and interaction among a set of SNPs and reduce the number of tests in GWAS. Therefore using SNP-set analysis in GWAS can achieve higher power than using individual SNP tests. In our future research, we hope to incorporate the SNP-set analysis into our S-GSB-HWD procedure.

Our S-GSB-HWD procedure also assumes no population stratification and no genotyping error. These issues often exist in the analysis of genetic data for both candidate association studies and GWAS. Since the S-GSB-HWD procedure estimates weights by use of p values from the HWD test using cases only, our S-GSB-HWD procedure may be sensitive to genotyping errors. Song and Elston [6] showed their HWD trend test (detecting the difference of HWD coefficients between cases and controls) can be robust to genotyping errors and population stratification. In addition, Price et al. [27] suggested that a generalization of the CA trend test can be used to control the stratification based on principal components analysis. Our future research will include extending our S-GSB-HWD procedure to handle genotyping errors and stratification. We plan to replace the HWD test using cases only and the CA trend test by the HWD trend test and generalized CA trend test, respectively, in the S-GSB-HWD procedure. However, when there is no obvious stratification in the case-control sample data, especially for data from a homogenous population, use of the HWD test (using cases only) and CA trend test as in our S-GSB-HWD procedure can generate higher power. We note that it may be difficult to extend the (non-permutation-based) approximation methods of Zang et al. [11] for MAX3 to handle population stratification.

Our methods in the present study for GWAS are based on controlling the FWER for multiple testing. It is not difficult to extend the method to control of the false discovery rate.

Acknowledgements

This research was supported by grants GM073766, GM077490, and GM081488 from the National Institute of General Medical Sciences. Q. Li was partially supported by NSFC No. 10901155. This study makes use of data generated by the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113 and 085475. We thank Xiangning Chen, Chun-Yu Liu, and Fengjiao Hu for helpful discussion.
Appendix A: Proof of the Null Asymptotic Independence Between the Statistics of CA Trend Test and Case-Only HWD Test

Here we use the notations defined in Section 2.1. For simplicity, let \( g \in \{0, 1, 2\} \) stands for the genotype that carries \( g \) copies of minor allele \( A \) at an SNP marker. We define an indicator variable \( r_{g} = 1 \) if the \( i \)-th case has \( g \) copies of allele \( A \) and \( r_{g} = 0 \) otherwise. Similarly, \( s_{g} = 1 \) if the \( j \)-th control has \( g \) copies of allele \( A \) and \( s_{g} = 0 \) otherwise. Then the numbers of individuals with genotype \( g \) in cases and controls can be calculated as \( r_{g} = \sum_{i=1}^{r} r_{g_{i}} \) and \( s_{g} = \sum_{j=1}^{s} s_{g_{j}} \), respectively. Let \( n_{g} = r_{g} + s_{g} \). We assume that HWE holds in the general population. Under the null hypothesis \( H_{0} \): the marker is not associated with the disease, the hypothesis \( H_{01} \) is true, where \( H_{01} \) is that HWE holds in the cases at the marker.

Apparently, the HWD test statistic \( U_{\text{HWD}} = \hat{U}_{\text{HWD}}^{2} \) (see section 2.2.2 in the main text), where \( U_{\text{HWD}} = D_{\text{HWD}} / \sqrt{\text{var}(D_{\text{HWD}})} \), \( D_{\text{HWD}} = \sqrt{\hat{\Delta}_{D}} \), and \( \hat{\Delta}_{D} = \hat{f}_{a}^{D} - \left( \hat{f}_{a}^{D} \right)^{T} \) is the estimate of HWD coefficient in the cases \( \Delta_{D} = f_{a}^{D} - \left( f_{a}^{D} \right)^{T} \). Then sample variance of \( D_{\text{HWD}} \) in the cases only \( \text{var}(D_{\text{HWD}}) = \hat{\Delta}_{D}^{T} \hat{\Delta}_{D} \)

\[
\hat{\Delta}_{D} = \frac{1}{T} \sum_{i=1}^{r} f_{a_{i}}^{D} + \frac{1}{S} \sum_{j=1}^{s} f_{a_{j}}^{D} - \frac{1}{T} \left( f_{a}^{D} \right)^{T} - \frac{1}{S} \left( f_{a}^{D} \right)^{T},
\]

where \( f_{a_{i}}^{D} \) stands for the transpose of \( f_{a} \) for the \( i \)-th case.

Define \( d = (d_{1}, d_{2}, d_{3})^{T} = \left( \sqrt{\hat{f}_{a}^{D} - f_{a}^{D}}, \sqrt{\hat{f}_{a}^{D} - f_{a}^{D}}, \sqrt{\hat{f}_{a}^{D} - f_{a}^{D}} \right)^{T} \), where \( \cdot^{T} \) stands for the transpose of the underlying vector and \( \hat{f}_{a}^{D} - f_{a}^{D} \) is the difference between the estimated and actual frequencies of genotype \( AA \) in the cases. Then we have

\[
\begin{bmatrix}
D_{\text{HWD}} \\
D_{\text{HWD}}
\end{bmatrix} = \begin{bmatrix}
\sqrt{1 - mn^{-1} \times D_{1} - \sqrt{mn^{-1} \times D_{2} - \sqrt{mn^{-1} \times D_{3}}}} - \sqrt{mn^{-1} \times \left( f_{a}^{D} - f_{a}^{D} \right)} \\
-2f_{a}^{D} - 2f_{a}^{D} + \sqrt{r^{-1/2} D_{i}^{2} + \sqrt{r} \Delta_{D}}
\end{bmatrix}
\]

\[ (A1) \]

Under \( H_{0} \) and \( H_{0} \) we observe that

\[
\begin{bmatrix}
D_{\text{HWD}} \\
D_{\text{HWD}}
\end{bmatrix} = \begin{bmatrix}
\sqrt{1 - mn^{-1}} \\
-2f_{a}^{D}
\end{bmatrix} \begin{bmatrix}
0 \\
1
\end{bmatrix} \begin{bmatrix}
d_{1} \\
r^{-1/2} d_{i}^{2} / \sqrt{\text{var}(D_{\text{HWD}})}
\end{bmatrix},
\]

\[ (A2) \]

\[
\begin{bmatrix}
U_{\text{HWD}} \\
U_{\text{HWD}}
\end{bmatrix} = \hat{H} d + \begin{bmatrix}
0 \\
r^{-1/2} d_{i} / \sqrt{\text{var}(D_{\text{HWD}})}
\end{bmatrix},
\]

\[ (A3) \]

where

\[
\hat{H} = \begin{bmatrix}
\sqrt{1 - mn^{-1}} / \text{var}(D_{\text{HWD}}) & 0 & -\sqrt{mn^{-1} / \text{var}(D_{\text{HWD}})} \\
-2f_{a}^{D} / \sqrt{\text{var}(D_{\text{HWD}})} & 1 / \sqrt{\text{var}(D_{\text{HWD}})} & 0
\end{bmatrix}.
\]

\[ (A4) \]

From the definition, we have \( \hat{f}_{a}^{D} = r^{-1} \sum_{i=1}^{r} u_{i}, \hat{f}_{a}^{D} = r^{-1} \sum_{i=1}^{r} u_{i}, \hat{f}_{a}^{D} = s^{-1} \sum_{j=1}^{s} v_{j} \) with \( u_{i} = r_{1} + \frac{1}{n} r_{1} \), \( r_{i} \) and \( s_{j} \) are two set of independent bounded random variables. Let \( \xi = (u_{1}, r_{1}, v_{1})^{T} \). Under \( H_{01} \) and \( H_{0} \), we have

\[
\begin{bmatrix}
E(\xi) \\
\text{cov}(\xi)
\end{bmatrix} = \begin{bmatrix}
\frac{f_{a}}{f_{a}} \\
\frac{2f_{a}}{f_{a}} & \frac{2f_{a}}{f_{a}} & \frac{2f_{a}}{f_{a}} & 0
\end{bmatrix} \begin{bmatrix}
\Delta \Omega
\end{bmatrix}.
\]

\[ (A5) \]

By the multivariate version of the Lindeberg-Levy Central Limit Theorem [28], we obtain the joint asymptotic normality of \( d \):

\[
d \rightarrow N(0, \Omega) \text{ as } r \rightarrow \infty \text{ and } s \rightarrow \infty,
\]

\[ (A6) \]
where \( \rightarrow^d \) reads as ‘convergences in distribution to’. By the laws of large numbers, we have \( f_A^D \rightarrow f_A^D \)
 as \( r \to \infty \), where \( \rightarrow^p \) reads as ‘convergences in probability to’. Similarly, we have \( f_A^C \rightarrow f_A^C \) and \( s \to \infty \).
Under \( H_{10} \) and \( H_0 \), as \( (r, s) \to (\infty, \infty) \) we have \( f_A^D \rightarrow f_A^C \), \( f_A^C \rightarrow f_A^C \), \( f_s \rightarrow f_s \), and \( f_g \) for each genotype \( g \in \{0, 1, 2\} \), \( \nu r \overset{\text{a.s.}}{\to} f_s^2 f_A^2 \), and \( \nu r \overset{\text{a.s.}}{\to} f_s^2 f_A^2 \). Thus, as \( (r, s) \to (\infty, \infty) \) and \( mn^{-1} \to c \), under \( H_{10} \) and \( H_0 \) we obtain that
\[
\mathbf{H} \overset{\text{a.s.}}{\to} \left[ \begin{array}{cc} \sqrt{2(1-c)} f_s^{-1/2} f_A^{-1/2} & 0 \\ -2 f_s^{-1} f_A^{-1} & 0 \end{array} \right] = \mathbf{H}.
\] (A7)

Since \( d_i \overset{\text{a.s.}}{\to} N(0, \frac{1}{2} f_s f_A) \) and \( \nu r \overset{\text{a.s.}}{\to} f_s^2 f_A^2 \), we have \( d_i^2 / \nu \overset{\text{a.s.}}{\to} \frac{1}{2} \chi^2_1 \)
and
\[
r^{-1/2} d_i / \nu \overset{\text{a.s.}}{\to} N(0, \frac{1}{r} A) \quad \text{as} \quad (\infty, \infty) \quad \text{and} \quad r^{-1/2} d_i / \nu \overset{\text{a.s.}}{\to} N(0, A).
\] (A8)

Substituting (A6)–(A8) into (A3) and applying Slutsky’s theorem [28], we obtain the joint asymptotic
normality of \( \left[ U_{\text{H0D}}, U_{\text{H0C}} \right] \overset{\text{a.s.}}{\to} N(0, A) \), where \( A = H\Omega H^t \). It follows from (A5) and (A7) that \( A = I_b \), the 2 \( \times \) 2 identity matrix.

References

15. The Wellcome Trust Case Control Consortium (WTCCC): Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–673.