

## Comparison of two-phase analyses for case–control genetic association studies

Gang Zheng<sup>1,\*</sup>, Mark Meyer<sup>1,2</sup>, Wentian Li<sup>3</sup> and Yaning Yang<sup>4</sup>

<sup>1</sup>*Office of Biostatistics Research, National Heart, Lung and Blood Institute, Bethesda, MD, U.S.A.*

<sup>2</sup>*Department of Mathematics and Statistics, American University, Washington, DC, U.S.A.*

<sup>3</sup>*The Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, Manhasset, NY, U.S.A.*

<sup>4</sup>*Department of Statistics and Finance, University of Science and Technology of China, Hefei, Anhui, People's Republic of China*

### SUMMARY

To test for genetic association between a marker and a complex disease using a case–control design, Cochran–Armitage trend tests (CATTs) and Pearson's chi-square test are often employed. Both tests are genotype-based. Song and Elston (*Statist. Med.* 2006; **25**:105–126) introduced the Hardy–Weinberg disequilibrium trend test and combined it with CATT to test for association. Compared to using a single statistic to test for case–control genetic association (referred to as single-phase analysis), two-phase analysis is a new strategy in that it employs two test statistics in one analysis framework, each statistic using all available case–control data. Two such two-phase analysis procedures were studied, in which Hardy–Weinberg equilibrium (HWE) in the population is a key assumption, although the procedures are robust to moderate departure from HWE. Our goal in this article is to study a new two-phase procedure and compare all three two-phase analyses and common single-phase procedures by extensive simulation studies. For illustration, the results are applied to real data from two case–control studies. On the basis of the results, we conclude that with an appropriate choice of significance level for the analysis in phase 1, some two-phase analyses could be more powerful than commonly used test statistics. Copyright © 2008 John Wiley & Sons, Ltd.

**KEY WORDS:** adaptive approach; constrained likelihood ratio test; efficiency robustness; genetic model selection; MAX; two-phase analysis

\*Correspondence to: Gang Zheng, Office of Biostatistics Research, National Heart, Lung and Blood Institute, Room 9200, 6701 Rockledge Drive, MSC 7913, Bethesda, MD 20892, U.S.A.

†E-mail: zhengg@nhlbi.nih.gov

Contract/grant sponsor: NSF; contract/grant number: 10671189

Contract/grant sponsor: Chinese Academy of Science; contract/grant number: KJCX3-SYW-S02

Received 3 September 2007

Accepted 18 April 2008

## 1. INTRODUCTION

In case-control association studies to test for genetic association between a disease and a bi-allelic marker, such as single nucleotide polymorphism (SNP), Cochran-Armitage trend tests (CATTs) [1-3] and Pearson's chi-square with 2 degrees of freedom (df) are usually employed. To apply the CATT, one assumes that the underlying genetic model (mode of inheritance) is known, for example, the recessive, additive, multiplicative or dominant models. Three CATTs are available optimal for the recessive, additive (multiplicative) and dominant models, respectively. The CATT has an asymptotic chi-square distribution with 1 df. However, it is not robust to misspecification of the genetic model. Pearson's test, on the other hand, is robust to misspecification of the genetic model, but it follows a chi-square distribution with 2 df asymptotically, which might be less powerful when the genetic model is known. A test with greater efficiency robustness is the maximum of the three CATTs for the recessive, additive (multiplicative) and dominant models, denoted here as MAX3 [4, 5]. In the literature, it may also be referred to as MAX. The efficiency robustness is defined as follows: test 1 is said to have greater efficiency robustness than test 2 if the first test is more efficient than the second one in their respective worst situations across a family of scientifically plausible genetic models. Another robust test is the constrained likelihood ratio test (CLRT), which has comparable power with MAX3 [6].

We refer to the CATTs and Pearson's test as a *single-phase analysis*, i.e. there is only one test statistic with a pre-specified significance level  $\alpha$ . In genetic analysis using case-control data, two jointly normally distributed test statistics for association, which are statistically independent, may be available. In this situation, one can consider a two-phase analysis such that applying one statistic in the first phase with a significance level  $\alpha_1$  followed by a second statistic with a significance level  $\alpha_2$  only if the first analysis is significant. Because the statistical powers of the two analyses are different, we apply the test with lower power in the first phase at a less stringent level  $\alpha_1$  and the more powerful test in the second phase with a more stringent level  $\alpha_2$ . The overall significance level of the two analyses is controlled at  $\alpha$  by properly choosing  $\alpha_1$  and  $\alpha_2$ . We refer to this analysis strategy as *two-phase analysis*. To control the overall Type I error at  $\alpha$  for the two-phase analysis, the two levels  $(\alpha_1, \alpha_2)$  satisfy  $\alpha_1\alpha_2 = \alpha$  by the independence of the two statistics under  $H_0$ . For example, if  $\alpha = 0.05$ , we can choose  $\alpha_1 = 0.5$  and  $\alpha_2 = 0.1$ . The null hypothesis is rejected if the two tests have  $p$ -values less than 0.5 and 0.1, respectively. Although we choose  $\alpha_1 = 0.5$  for illustration, testing a null hypothesis of zero mean at the 0.5 level with a test statistic with a positive mean under the alternative, which has at least 50 per cent power, is different from flipping a coin, which has a constant 50 per cent power. More justification of using a relatively large  $\alpha_1$  is given in Section 3.2 and Figure 1. If we apply a single test, we have to test  $H_0$  at a more stringent level  $\alpha = 0.05$ . Although two independent test statistics should be applied in this two-phase analysis, testing them at less stringent significance levels may have power advantage for certain genetic models. The two-phase analysis is designed to take the advantage that both tests can be significant at less stringent significance levels given that a single test is not significant at the more stringent level. An illustrative example, some numerical comparisons and analytical justification using the two-phase analysis and some single-phase analysis are given in Sections 3.2 and 3.3.

Our two-phase analysis is different from the two-stage design of Skol *et al.* [7], who considered a two-stage cost-effective genotyping design with a joint analysis. In contrast, we focus on the two-phase analysis with a single-stage genotyping design. Thus, all markers are genotyped using all case-control samples. For review of other two-stage designs for association studies, see Elston *et al.* [8].

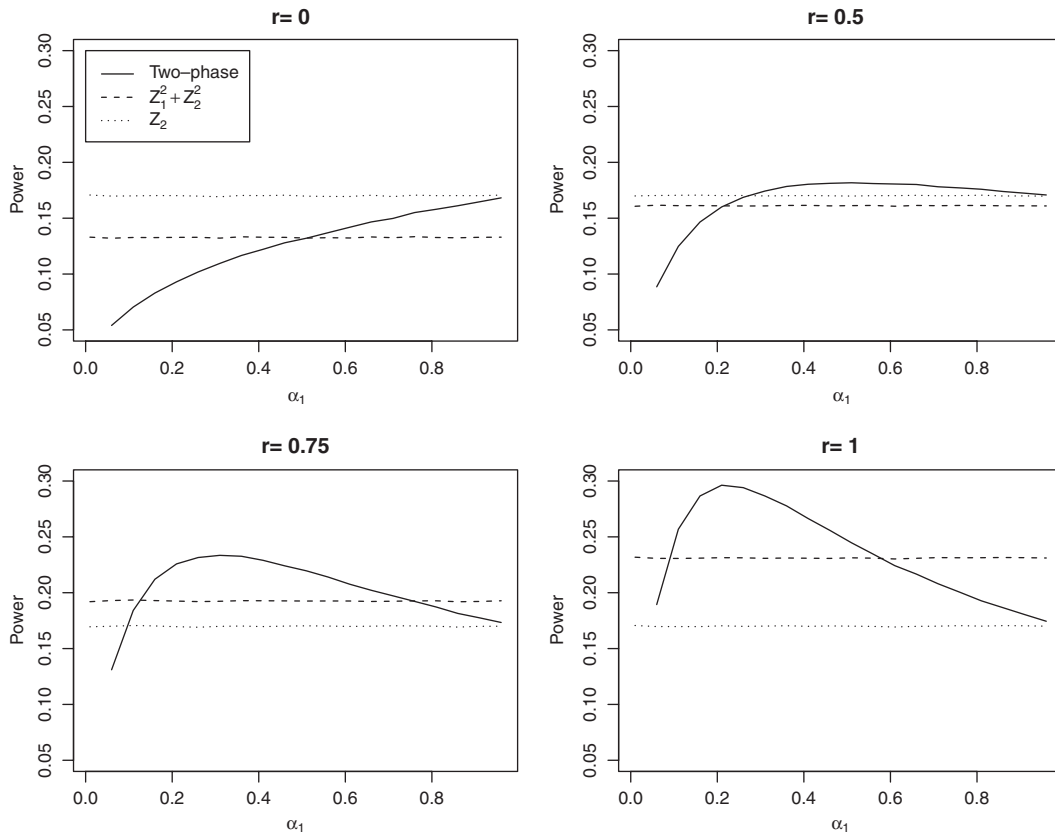


Figure 1. Power comparison of two-phase and single-phase analyses (see Section 3.2 for details).

Two different types of two-phase analysis procedures have been recently studied for case-control association studies [9, 10]. Both use the two trend tests, the CATT and Hardy-Weinberg disequilibrium trend test (HWDTT) [11], in different ways. They also differ in whether they are ordered. The two tests in the first type are ordered and applied sequentially. The two tests in the second type are not ordered; hence, each test can be applied in phase 1. However, we usually apply the less powerful test in phase 1. A brief review of these two-phase analyses will be given in Section 3.1. Although these two-phase procedures are proposed, their performance has been compared only with some single-phase analyses. One important assumption for these two-phase analyses is that Hardy-Weinberg equilibrium (HWE) holds in the population [9, 10], although the procedures are robust to the moderate deviation from HWE. In this article, we propose another two-phase analysis strategy, referred to as two-phase beta-test (TBT) and compare the three two-phase analyses and some common single-phase analyses by simulation studies and applications with real data. This new two-phase procedure does not require HWE in the population. In Section 2, commonly used single-phase analyses are reviewed. The existing two-phase analyses and new TBT are given in Section 3. Simulation studies are reported in Section 4 with applications to two case-control studies in Section 5. Discussion is given in Section 6.

## 2. COMMON SINGLE-PHASE ANALYSES

## 2.1. Notation and models

For an SNP with two alleles  $A$  and  $B$ , denote genotype by  $G_0=AA$ ,  $G_1=AB$  or  $G_2=BB$ . We assume that  $B$  is the risk allele so that the risk of disease increases with the number of risk allele  $B$  in the genotypes  $G_0$ ,  $G_1$  and  $G_2$ . Assume that  $r$  cases and  $s$  controls are randomly sampled from the population. We observe genotype counts  $(r_0, r_1, r_2)$  and  $(s_0, s_1, s_2)$  for  $(G_0, G_1, G_2)$  in cases and controls, respectively. Denote their corresponding probabilities for  $(G_0, G_1, G_2)$  by  $(p_0, p_1, p_2)$  and  $(q_0, q_1, q_2)$ . Then  $(r_0, r_1, r_2)$  and  $(s_0, s_1, s_2)$  follow independent multinomial distributions  $\text{mul}(r; p_0, p_1, p_2)$  and  $\text{mul}(s; q_0, q_1, q_2)$ , respectively. The penetrances are denoted by  $f_i = \Pr(\text{case}|G_i)$  for  $i=0, 1, 2$ . The prevalence of disease is denoted by  $K = \Pr(\text{case}) = \sum_{i=0}^2 g_i f_i$ , where  $g_i = \Pr(G_i)$  in the population. Then  $p_i = f_i g_i / K$  and  $q_i = (1 - f_i) g_i / (1 - K)$  for  $i=0, 1, 2$  [12].

Under the null hypothesis of no association,  $H_0: p_i = q_i = g_i$  for  $i=0, 1, 2$ , i.e.  $f_0 = f_1 = f_2 = K$ . The alternative hypothesis, when  $B$  is the risk allele, is given by  $H_1: f_2 \geq f_1 \geq f_0$  and  $f_2 > f_0$ . Four common genetic models are defined in terms of penetrances. The model is recessive (REC) when  $f_0 = f_1$  (i.e.  $\lambda_1 = f_1/f_0 = 1$ ), additive (ADD) when  $f_1 = (f_0 + f_2)/2$  (i.e.  $2\lambda_1 = 1 + \lambda_2$  where  $\lambda_2 = f_2/f_0$ ), multiplicative (MUL) when  $f_2 f_0 = f_1^2$  (i.e.  $\lambda_2 = \lambda_1^2$ ) and dominant (DOM) when  $f_1 = f_2$  (i.e.  $\lambda_1 = \lambda_2$ ). Here  $\lambda_i$ ,  $i=1, 2$ , are referred to as genotype relative risks (GRRs).

## 2.2. Single-phase analysis

Given the  $2 \times 3$  contingency table with  $(r_0, r_1, r_2)$  cases and  $(s_0, s_1, s_2)$  controls, Pearson's chi-square test, denoted as  $\chi_2^2$ , can be expressed as

$$\chi_2^2 = \sum_{i=0}^2 (r_i - rn_i/n)^2 / (rn_i/n) + \sum_{i=0}^2 (s_i - sn_i/n)^2 / (sn_i/n) \quad (1)$$

where  $n_i = r_i + s_i$  and  $n = r + s$ . Under  $H_0$ ,  $\chi_2^2$  has an asymptotic chi-square distribution with 2 df. Note that to apply  $\chi_2^2$ , one does not need to know the risk allele.

Denote  $x_0 = 0, x_1 = x, x_2 = 1$  and  $0 \leq x \leq 1$ . The CATT can be expressed as [4, 12]

$$T_x = \frac{U_x}{\{\widehat{\text{var}}(U_x)\}^{1/2}} = \frac{n^{-1} \sum_{i=0}^2 x_i (sr_i - rs_i)}{\{\widehat{\text{var}}[n^{-1} \sum_{i=0}^2 x_i (sr_i - rs_i)]\}^{1/2}} \quad (2)$$

where  $x = 0, \frac{1}{2}, 1$  for the REC, ADD and DOM models, respectively, when  $B$  is the risk allele [13]. The score  $x = \frac{1}{2}$  is also used for the MUL model. For a given  $x$ , under  $H_0$ ,  $T_x$  has an asymptotic normal distribution  $N(0, 1)$  ( $E_{H_1}(T_x) > 0$  under the alternative when  $B$  is the risk allele). To apply  $T_x$ , the variance in the denominator of (2) can be consistently estimated under  $H_0$  as

$$\widehat{\text{var}}(U_x) = n\phi(1-\phi) \left\{ \sum_{i=0}^2 x_i^2 (n_i/n) - \left[ \sum_{i=0}^2 x_i (n_i/n) \right]^2 \right\}$$

where  $\phi = r/n$  is the proportion of cases [14]. When the genetic model is known (REC, ADD, MUL or DOM), the corresponding CATT is a directed test with 1 df and is more powerful than  $\chi_2^2$  [15]. When the genetic model is unknown,  $\chi_2^2$  is more robust than the single CATT. On the other

hand, as the genetic models are unknown for many complex diseases, a more efficiency robust test,  $\text{MAX3} = \max(T_0, T_{1/2}, T_1)$  has been studied (when  $B$  is the risk allele) and applied to real data [4, 5, 16, 17]. When the risk allele is unknown, we consider  $\text{MAX3} = \max(|T_0|, |T_{1/2}|, |T_1|)$ . The theory of efficiency robustness was studied by Gastwirth [18] and Freidlin *et al.* [19].  $\text{MAX3}$  has been included in the latest version of SAS JMP Genomics software.

An alternative robust test is the CLRT proposed by Wang and Sheffield [6], given by

$$T_{\text{CLRT}} = 2 \left\{ \sup_{(f_0, f_1, f_2) \in \Omega} l_2(f_0, f_1, f_2) - l_0 \right\} \quad (3)$$

where  $l_2 = \sum_{i=0}^2 \{r_i \log f_i + s_i \log(1 - f_i)\}$ ,  $\Omega = \{(f_0, f_1, f_2) : f_0 \leq f_1 \leq f_2; f_2 > f_0\}$  (given that  $B$  is the risk allele),  $l_0 = l_2(\hat{f}_0, \hat{f}_0, \hat{f}_0) = r \log(\hat{f}_0) + s \log(1 - \hat{f}_0)$ , and  $\hat{f}_0 = r/n$ . The statistic  $T_{\text{CLRT}}$  can be easily obtained as follows [6]: Estimate  $f_i$  in  $l_2$  by  $\hat{f}_i = r_i/n_i$  for  $i = 0, 1, 2$ . If  $\hat{f}_0 \leq \hat{f}_1 \leq \hat{f}_2$  and  $\hat{f}_2 > \hat{f}_0$ , calculate  $l_2$  and  $T_{\text{CLRT}}$ . Otherwise, estimate  $f_i$  in  $l_2$  under the REC model by  $\hat{f}_0 = \hat{f}_1 = (r_0 + r_1)/(n_0 + n_1)$  and  $\hat{f}_2 = r_2/n_2$ . Then calculate  $T_{\text{CLRT}}$ , denoted by  $T_{\text{CLRT,REC}}$ . Also estimate them under the DOM model by  $\hat{f}_0 = r_0/n_0$  and  $\hat{f}_1 = \hat{f}_2 = (r_1 + r_2)/(n_1 + n_2)$ . Then calculate  $T_{\text{CLRT}}$ , denoted by  $T_{\text{CLRT,DOM}}$ . Finally,  $T_{\text{CLRT}}$  is given by  $T_{\text{CLRT}} = \max(T_{\text{CLRT,REC}}, T_{\text{CLRT,DOM}})$ . The asymptotic distribution for  $T_{\text{CLRT}}$  is a mixture of chi-squared distributions [6], which can also be obtained by the parametric bootstrap (see Appendix B and simulation studies later).

Zaykin and Nielsen [20] considered using the difference in Hardy–Weinberg disequilibrium (HWD) in cases and controls to test for association. Formally, Song and Elston [11] introduced the HWDTT, which compares HWD coefficients in cases ( $\Delta_1 = p_2 - (p_2 + p_1/2)^2$ ) and controls ( $\Delta_0 = q_2 - (q_2 + q_1/2)^2$ ). The HWDTT can be expressed as

$$T_{\text{HWDTT}} = \frac{(rs/n)^{1/2}(\hat{\Delta}_1 - \hat{\Delta}_0)}{\{1 - n_2/n - n_1/(2n)\}\{n_2/n + n_1/(2n)\}} \quad (4)$$

where  $\hat{\Delta}_1 = r_2/r - \{r_2/r + r_1/(2r)\}^2$  and  $\hat{\Delta}_0 = s_2/s - \{s_2/s + s_1/(2s)\}^2$ .  $T_{\text{HWDTT}}$  asymptotically follows  $N(0, 1)$  under  $H_0$ . Numerical and analytical results [10, 11] showed that  $T_{1/2}$  and  $T_{\text{HWDTT}}$  are asymptotically independent under  $H_0$  when HWE holds in the population. Song and Elston [11] proposed a linear combination of the two trend tests, e.g. a weighed average or using Fisher's combination of their  $p$ -values:  $-2 \log(p_{1/2}) - 2 \log(p_{\text{HWDTT}}) \sim \chi_4^2$  under  $H_0$ , where  $p_{1/2}$  and  $p_{\text{HWDTT}}$  are  $p$ -values of  $T_{1/2}$  and  $T_{\text{HWDTT}}$ , and  $\chi_4^2$  is the chi-square distribution with 4 df. After combining the two trend tests, the combined test is still a single-phase analysis.

### 3. TWO-PHASE TESTS

#### 3.1. Existing two-phase analysis

The two-phase analysis that we study contains the two trend tests: HWDTT and CATT. Let  $n$  be the number of individuals and  $p$  be the risk allele frequency. From Zheng and Ng [10], when HWE holds in the population, the asymptotic null correlations between the HWDTT and CATTs are given by  $\rho_0 = \text{corr}_{H_0}(T_{\text{HWDTT}}, T_0) = \{(1-p)/(1+p)\}^{1/2} + O(n^{-1})$ ,  $\rho_{1/2} = \text{corr}_{H_0}(T_{\text{HWDTT}}, T_{1/2}) = O(n^{-1})$  and  $\rho_1 = \text{corr}_{H_0}(T_{\text{HWDTT}}, T_1) = -\{p/(2-p)\}^{1/2} + O(n^{-1})$ , where  $p$  is the frequency of allele  $B$ . Thus, because the HWDTT and  $T_{1/2}$  follow a bivariate normal distribution, they are

asymptotically independent under  $H_0$  and  $\rho_{1/2}=0$ . This property is not sensitive to moderate departure from HWE [10].

*3.1.1. Adaptive power approach.* Song and Elston [11] noted that  $T_{\text{HWDTT}}$  is less powerful than  $T_{1/2}$ , in particular for the ADD and MUL models. Under the ADD model, the HWDTT has little power whereas under the MUL model, it has no power. Thus, in a two-phase analysis, Zheng *et al.* [9] first applied the HWDTT at the level  $\alpha_1$  followed by a more powerful CATT at the level  $\alpha_2$  in the second phase when the HWDTT is significant. Instead of choosing fixed  $\alpha_1$  and  $\alpha_2$  given  $\alpha_1\alpha_2=\alpha$ , an alternative approach studied by Zheng *et al.* [9] was to choose  $\alpha_1$  so that the asymptotic power of the HWDTT estimated using the observed data,  $\pi_{\text{HWDTT}}(\hat{p}_i, \hat{q}_i, \alpha_1)$ , is at least 80 per cent in the first phase. Here the asymptotic power for the HWDTT,  $\pi_{\text{HWDTT}}$ , is a function of  $(p_i, q_i, \alpha_1)$ , whose explicit expression was derived in [9] and also given in Appendix A. The parameters in the power function can be estimated consistently under the null using cases and controls separately as  $\hat{p}_i = r_i/r$  and  $\hat{q}_i = s_i/s$ . Denote  $\alpha_1^* = \inf\{\alpha_1 : \pi_{\text{HWDTT}}(\hat{p}_i, \hat{q}_i, \alpha_1) \geq 0.80\}$ . The level for the second phase is  $\alpha_2^* = \alpha/\alpha_1^*$ . However, using  $(\alpha_1^*, \alpha_2^*)$  only controls Type I error asymptotically at the level  $\alpha$ . Numerical results [9] showed that the Type I error was inflated using  $\alpha_1^*$  and  $\alpha_2^* = \alpha/\alpha_1^*$ . Because the analyses are not independent for finite sample sizes when  $\alpha_1^*$  and  $\alpha_2^*$  are functions of data. Using a bootstrap correction procedure (Appendix B), Zheng *et al.* [9] found an adjusted level  $\alpha^* < \alpha$  such that using  $\alpha_2^* = \alpha^*/\alpha_1^*$  in the second phase would control the overall Type I error at  $\alpha$  if we still treat the two analyses are independent, where  $\alpha_1^*$  is the same as before. This approach is referred to as adaptive power approach (APA). The  $p$ -value of the APA is the  $p$ -value of the second phase multiplied by a factor  $\alpha/\alpha_2^*$  if the first phase analysis is significant (see discussion in Appendix C). The APA is not ordered, but we apply the HWDTT in phase 1 as it is often less powerful than the CATT.

*3.1.2. Genetic model selection.* In the APA,  $T_{1/2}$  is always used in the second phase no matter what the true genetic model is. Genetic model selection (GMS) is a two-phase analysis in which the ADD model is assumed unless the REC or DOM models are selected using the HWDTT in the first phase, followed by testing association using the CATT corresponding to the selected genetic model in the second phase [10]. Wittke-Thompson *et al.* [21] studied the four genetic models using HWD coefficient  $\Delta = p_2 - (p_2 + p_1/2)^2$  in cases ( $\Delta_1$ ) and controls ( $\Delta_0$ ). When HWE holds in the population, Zheng and Ng [10] divided the alternative space into four mutually disjoint regions corresponding to the four genetic models. Using the signs of  $(\Delta_1, \Delta_0)$ , they further showed that both ADD and MUL models have the signs  $(-, -)$  for  $(\Delta_1, \Delta_0)$ , but the REC and DOM models have the signs  $(+, -)$  and  $(-, +)$ , respectively. Thus, the REC and DOM models imply that  $\Delta_1 - \Delta_0 > 0$  and  $\Delta_1 - \Delta_0 < 0$ , respectively. A similar discussion on genetic model and HWD coefficient was also given in Suh and Li [22]. Note that when the genetic model is unknown, one would use  $T_{1/2}$  to test association. Hence, we would use  $T_{1/2}$  to test for association unless data suggest a strong REC or DOM model. Then we use  $T_0$  or  $T_1$ , respectively. Using the HWDTT in the first phase of the GMS [10], a strong REC (DOM) model is indicated if  $T_{\text{HWDTT}} > c$  ( $T_{\text{HWDTT}} < -c$ ), where  $c = \Phi^{-1}(0.95) = 1.645$  is used and  $\Phi$  is the distribution function for  $N(0, 1)$ . Note that the GMS in phase 1 based on HWDTT is independent of the risk allele. However, which CATT will be chosen in phase 2 given a selected genetic model requires the risk allele to be known. In the second phase,  $T_{\text{model}}$  is used as a test statistic for association. When  $B$  is the risk allele,  $T_{\text{model}} = T_0$  if  $T_{\text{HWDTT}} > c$ ,  $T_1$  if  $T_{\text{HWDTT}} < -c$ , and  $T_{1/2}$  if otherwise.

Since the model selection is correlated with  $T_{\text{model}}$ , the critical value for the second phase depends on the correlation. Suppose in the second phase the null hypothesis is rejected at the level  $\alpha^*$  using the normal distribution. Let  $z_{1-\alpha}$  be the  $100(1-\alpha)$ th percentile of  $N(0, 1)$ . Then, when  $B$  is the risk allele, to control the overall Type I error at  $\alpha$  in the GMS,  $\alpha^*$  satisfies  $\Pr_{H_0}(T_{\text{model}} > z_{1-\alpha^*/2}) = \alpha/2$ , which can be expressed as

$$\alpha - 0.9\alpha^* = 2 \sum_{x=0}^1 \int_{\Omega_x} \left\{ \Phi \left( \frac{-z_{1-\alpha^*/2} + \rho_x u}{(1-\rho_x^2)^{1/2}} \right) \right\} d\Phi(u) \quad (5)$$

where  $\Omega_0 = \{u : u > c\}$ ,  $\Omega_1 = \{u : u < -c\}$  and  $\rho_x = \text{corr}_{H_0}(T_{\text{HWDTT}}, T_x)$ . Note that (5) is slightly different from the one derived in Zheng and Ng [10], who used the minor allele as the risk allele or the risk allele is known. We assume that the risk allele is unknown, which could have frequency greater than 0.5 in the samples. To apply the above GMS, we apply GMS twice using alleles  $A$  and  $B$  as the risk allele, respectively, each at the level  $\alpha/2$ . The parameter  $p$  in (5) is estimated by  $\hat{p} = (2n_2 + n_1)/(2n)$ . A SAS IML program to find  $\alpha^*$  is available upon request. Its  $p$ -value can be expressed as  $p_{\text{model}} = 2Pr_{H_0}(T_{\text{model}} > t)$ , where  $t$  is the observed  $|T_{\text{model}}|$ , and is given by

$$p_{\text{model}} = 1.8\Phi(-t) + 2 \sum_{x=0}^1 \int_{\Omega_x} \left\{ \Phi \left( \frac{-t + \hat{\rho}_x u}{(1-\hat{\rho}_x^2)^{1/2}} \right) \right\} d\Phi(u)$$

Note that the GMS is ordered. In phase 2, the CATT is chosen based on the results of phase 1.

### 3.2. Power advantage of the two-phase analysis

To show advantage of using two-phase analyses, let us consider unordered situations. We use a real example from a genome-wide association study for age-related macular degeneration using 103 611 SNPs [23, 24]. Applying a single test statistic comparing allele frequencies in cases and controls with all the SNPs, Klein *et al.* [23] identified the top two SNPs (rs380390 and rs1329428), which had observed  $p$ -values  $4.1 \times 10^{-8}$  and  $1.35 \times 10^{-6}$ , respectively (Bonferroni-corrected  $p$ -values are 0.0043 and 0.14). Note that their  $p$ -values were based on the allele-based test, which was equivalent to  $T_{1/2}$  when HWE holds [3]. We focus on the second top SNP as it has Bonferroni-corrected  $p$ -value greater than 0.05. Using the second SNP with  $(r_0, r_1, r_2) = (2, 24, 68)$  and  $(s_0, s_1, s_2) = (5, 29, 14)$  as an example, we compare the CATTs  $(T_0, T_{1/2}, T_1)$ , Pearson's 2 df test, Fisher's combination with a simple two-phase analysis with HWDTT tested at the first phase with  $\alpha_1 = 0.50$  followed by using the CATT  $T_{1/2}$  at the second phase with  $\alpha_2 = 2\alpha/103611$  to control the overall significance level at  $\alpha/103611 = 4.83 \times 10^{-7}$ .

Numerical results show that  $(T_0^2, T_{1/2}^2, T_1^2) = (24.273, 24.196, 4.658)$  with observed  $p$ -values  $(8.36 \times 10^{-7}, 8.70 \times 10^{-7}, 3.09 \times 10^{-2})$ . Neither is less than  $4.83 \times 10^{-7}$ . The HWDTT ( $T_{\text{HWDTT}}^2$ ) and Pearson's test are 3.28 and 25.045 with observed  $p$ -values 0.0701 and  $3.64 \times 10^{-6}$ , respectively. If we apply the HWDTT at  $\alpha_1 = 0.50 (> 0.0701)$ , we will reject the null and conduct the second test using  $T_{1/2}$  at  $\alpha_2 = 2 \times 4.83 \times 10^{-7} = 9.66 \times 10^{-7}$ , which is greater than the observed  $p$ -value of  $T_{1/2}$ . Thus, using a simple two-phase analysis, the second top SNP of Klein *et al.* [23] would be significant. On the other hand, if we apply Fisher's combination,  $-2\log(p_{1/2}) - 2\log(p_{\text{HWDTT}}) = 33.23$  is less than 34.919, the critical value for  $\chi_4^2$ . Applying the APA to the same SNP, Zheng *et al.* [9] obtained  $\alpha_1^* = 0.33$  and  $\alpha^* = 0.031$ . Thus,  $\alpha_2^* = 0.031/(103611\alpha_1^*) = 9.06 \times 10^{-7}$ , which is greater than the observed  $p$ -value of  $Z_{1/2}$ . Thus, the null hypothesis of no association for the second

top SNP is rejected with Bonferroni-corrected  $p$ -value 0.048. Zheng and Ng [10] also applied the GMS to the second top SNP. They showed  $T_{\text{HWDTT}} = 1.930 > c$ . Hence,  $T_{\text{model}} = T_0 = 4.9268$  with observed  $p$ -value  $8.36 \times 10^{-7}$ . The Bonferroni-corrected  $p$ -value is 0.0867, which is greater than the adjusted nominal level  $\alpha^* = 0.0374$ . Thus, this SNP is not significant when GMS is applied. However, the model selection actually chose the largest CATT as MAX3 does (the second SNP is not significant when MAX3 is applied). To apply CLRT of Wang and Sheffield [6], we obtain  $\hat{f}_0 = 0.2857$ ,  $\hat{f}_1 = 0.4528$ ,  $\hat{f}_2 = 0.8293$  and  $T_{\text{RLRT}} = 25.348$ . From Wang and Sheffield [6], the asymptotic null distribution for  $T_{\text{RLRT}}$  is a mixture of a chi-squared distribution of 2 df with probability  $2k = 0.8285$  and a bivariate normal distribution with correlation coefficient  $\gamma = 0.2662$  with probability  $1 - 2k = 0.1715$ . Here  $\gamma = (g_0 g_2)^{1/2} / \{(1 - g_0)(1 - g_2)\}^{1/2}$  and  $k = \arccos(\gamma) / (2\pi)$  [6]. A simulation with 10 000 000 replicates from the above asymptotic null distribution shows that the estimated  $p$ -value for  $T_{\text{RLRT}}$  is  $3.2 \times 10^{-6}$ , which is not significant after the Bonferroni correction.

Although the above real example indicates the advantage of using the two-phase analysis, motivation of applying the two-phase analysis in general setting has not been clearly justified. Here we provide further justification of the two-phase analysis for hypothesis testing. Suppose we test for a null hypothesis of  $\mu = 0$  with two independent test statistics  $Z_1$  and  $Z_2$ . Suppose  $(Z_1, Z_2)$  has a bivariate normal distribution with means  $(\mu_1, \mu_2)$ , unit variances and correlation  $\rho$ . Under the null,  $\mu_1 = \mu_2 = 0$  and  $\rho = 0$ , and under the alternative  $\mu_2 \geq \mu_1 \geq 0$  ( $\mu_2 > 0$ ) and  $\rho = \mu_1 / (\mu_1^2 + \mu_2^2)^{1/2} = r / (1 + r^2)^{1/2} > 0$ , where  $r = \mu_1 / \mu_2$ . Note that  $Z_2$  is more powerful than  $Z_1$ . Two common test statistics are  $Z_2^2$  and  $Z_1^2 + Z_2^2$ , which follow chi-squared distributions with 1 and 2 df, respectively, under the null. Consider an alternative approach with a two-phase analysis based on the pair  $(Z_1, Z_2)$ . We compared their analytical power functions of the above three tests for a given  $\alpha_1$  and  $r = 0, 0.5, 0.75$  and 1, which corresponds to  $\rho = 0, 0.447, 0.6$  and 0.707. Figure 1 shows the plots of their power functions for different situations. Note that, in our application,  $r = 0$  corresponds to the MUL model, under which  $E(T_{\text{HWDTT}}) = 0$  under the alternative. Under the MUL,  $Z_1$  has no information, and  $Z_2$  is always more powerful than the two-phase analysis, which could still be more powerful than  $Z_1^2 + Z_2^2$  when  $\alpha_1 > 0.5$ . Note that there is no single test dominates the other two tests. When  $r > 0$ , the two-phase analysis could be more powerful than  $Z_2$  alone and  $Z_1^2 + Z_2^2$  when  $\alpha_1$  is in the range of 0.2–0.5. The power of the two-phase analysis decreases when  $\alpha_1$  becomes smaller, because phase 1 is more difficult to reject, which reduces the overall rejection of the null. On the other hand, when  $\alpha_1$  is close to 1, the information in phase 1 is not fully used; therefore, the two-phase analysis is less powerful than  $Z_1^2 + Z_2^2$ , but it is always more powerful than  $Z_2$  when  $\alpha_1$  is close to 1, because  $Z_2$  does not use any information of  $Z_1$ .

Our analyses using the real example and analytical powers justify the use of the two-phase analysis in hypothesis testing with appropriately chosen  $\alpha_1$ .

### 3.3. Two-phase beta test

In this section, we first study a new two-phase analysis based on  $T_{1/2}$  and Pearson's test  $\chi_2^2$ . After studying its power performance, we modify it to incorporate the HWDTT.

**3.3.1. A naive TBT.** Denote  $Z_1 = T_{1/2}^2 / \chi_2^2$  as the additive component of  $\chi_2^2$  and  $Z_2 = \chi_2^2$ . In Appendix D, we prove that these two test statistics are asymptotically independent under  $H_0$ . This does not depend on whether HWE holds in the population. Further, under  $H_0$ ,  $Z_1$  has an asymptotic beta distribution  $\text{Beta}(\frac{1}{2}, \frac{1}{2})$  with the density function  $f(x; \frac{1}{2}, \frac{1}{2}) = \{x(1-x)\}^{-1/2} / \pi$ .

To determine which test will be applied in the first phase, we conduct a simulation study to compare their power performance under the four genetic models with a common allele frequency  $p=0.3$ . The test with lower power will be used in the first phase with level  $\alpha_1=0.5$  and the test with higher power in the second phase with level  $\alpha_2=\alpha/\alpha_1=2\alpha$  only if the first test is significant. (Note that the order of  $Z_1$  and  $Z_2$  can be reversed with the same power if we choose different  $\alpha_1$  and  $\alpha_2$ .) In the simulation, we chose sample size  $n=500$  with 250 cases and 250 controls. For a given genetic model, the alternative is specified by a single GRR  $\lambda_2$  ( $\lambda_1$  can be calculated given the genetic model, prevalence  $K$  and  $\lambda_2$ ). The disease prevalence is  $K=0.1$  with the risk allele frequency  $p=0.3$ . Assuming HWE holds in the population,  $g_0=\Pr(G_0)=(1-p)^2$ ,  $g_1=\Pr(G_1)=2p(1-p)$  and  $g_2=\Pr(G_2)=p^2$ . The penetrances are calculated by  $f_0=K/(g_0+\lambda_1g_1+\lambda_2g_2)$ ,  $f_1=\lambda_1f_0$  and  $f_2=\lambda_2f_0$ . Then genotype frequencies in cases ( $p_i$ ) and controls ( $q_i$ ) can be obtained (see Section 2). Genotype counts in cases and controls can be simulated from  $\text{mul}(r; p_0, p_1, p_2)$  and  $\text{mul}(s; q_0, q_1, q_2)$ , respectively. These simulation procedures will be used later for power comparison. For comparison, Fisher's combination of the  $p$ -values of  $Z_1$  and  $Z_2$  is also considered. The empirical powers of  $Z_1$ ,  $Z_2$  and Fisher's combination based on 10 000 replications are first plotted in Figure 2. The results indicate that  $Z_1$  has little power compared with  $Z_2$ , in particular under the REC and DOM models. This is not surprising because  $Z_1$  can be interpreted as the proportion of  $Z_2$  explained by the additive component. The results also show that Pearson's test ( $Z_2=\chi_2^2$ ) is more powerful than Fisher's combination test under the REC and DOM models, whereas the latter outperforms  $Z_2$  under the ADD and MUL models, although the difference is small.

Based on Figure 2, in a two-phase analysis, we test  $Z_1$  in the first phase with  $\alpha_1=0.5$  followed by  $\chi_2^2$  at the level  $\alpha_2=\alpha/\alpha_1=2\alpha$  if the first phase is significant. This approach is referred to as a naive TBT (TBT0). Table I reports the empirical power of  $Z_1$  with level  $\alpha_1=0.5$  and  $p=0.1, 0.3$ . Under  $H_0$  (GRR=1), the power of  $Z_1$  is expected to be about 50 per cent. Table I shows that the power of  $Z_1$  is lower for the REC model and relatively higher for the ADD, MUL and DOM models. When MAF=0.1, the power is about 54–65 per cent for the ADD or MUL models and 58–82 per cent for the DOM model. When MAF=0.3, the power is about 59–80 per cent for the ADD or MUL models and 62–78 per cent for the DOM model. In Figure 2, the TBT0 with  $(\alpha_1, \alpha_2)=(0.5, 0.1)$  is also plotted. The TBT0 outperforms the other tests under the ADD and MUL models, and is more powerful than  $Z_2=\chi_2^2$  and Fisher's combination under the DOM model when GRR is less than 1.5. Under the DOM model when GRR is greater than 1.5, TBT0 loses less than 10 per cent power compared with  $\chi_2^2$ . The TBT0, however, is not powerful under the REC model.

**3.3.2. A modified TBT.** Since the HWDTT is more powerful under the REC and DOM models than the ADD and MUL models, we use the HWDTT to modify the TBT0. A modified TBT, denoted by TBT\*, is to combine TBT0 and the two-phase analysis with  $T_{\text{HWDTT}}$  and  $T_{1/2}$  as follows. We apply two different two-phase analyses. For the first two-phase analysis, we apply  $T_{1/2}/\chi_2^2$  in phase 1 at the level  $\alpha_1=0.5$  and  $\chi_2^2$  in phase 2 at the level  $\alpha_2=0.1$  only if  $T_{1/2}/\chi_2^2$  is significant. Denote the  $p$ -value for this first two-phase analysis by  $p_{\text{TBT0}}$ , which is given by  $p_{\text{TBT0}}=\alpha_2 p_{\chi_2^2}=p_{\chi_2^2}/2$ , where  $p_{\chi_2^2}$  is the  $p$ -value of stage 2 analysis (Appendix C). For the second two-phase analysis, we apply  $T_{\text{HWDTT}}$  in phase 1 at the same level  $\alpha_1=0.5$ , and  $T_{1/2}$  in phase 2 at the level  $\alpha_2=0.1$  only if phase 1 analysis is significant. Denote the  $p$ -value by  $p_{\text{TT}}=\alpha_1 p_{T_{1/2}}=p_{T_{1/2}}/2$ , where  $p_{T_{1/2}}$  is the  $p$ -value of stage 2 analysis (Appendix C). We reject the null hypothesis if  $1.5 \times \min(p_{\text{TT}}, p_{\text{TBT0}}) < \alpha$ , where

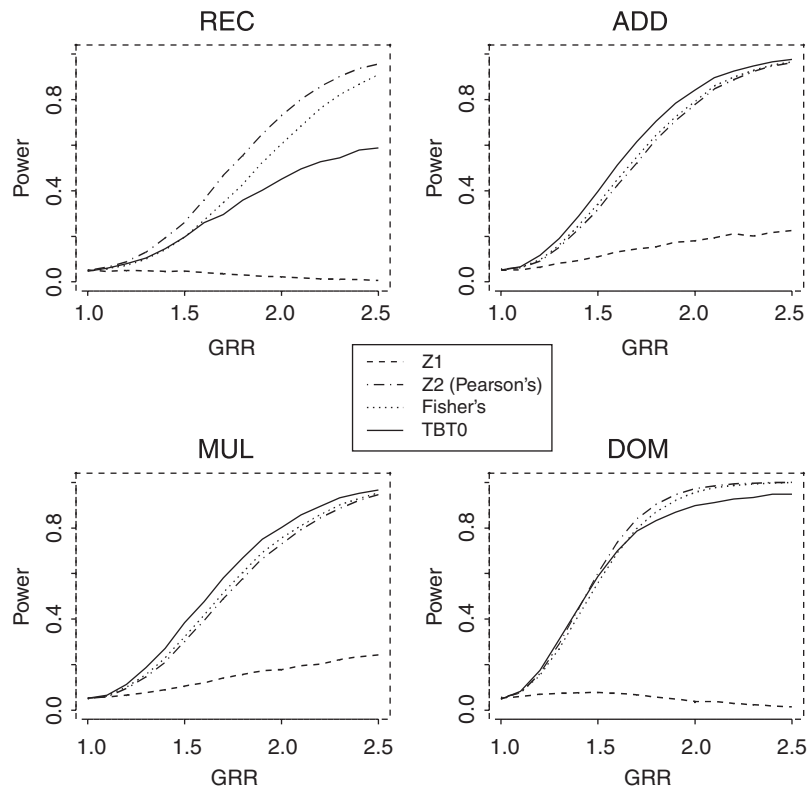


Figure 2. Empirical power of  $Z_1$ ,  $Z_2$  ( $\chi^2$ ), Fisher's combination of  $p$ -values of  $Z_1$  and  $Z_2$  and TBT0 ( $\alpha_1=0.5$ ) with  $n=500$  case-control samples, prevalence  $K=0.1$ , and risk allele frequency  $p=0.3$  under four genetic models. GRR is  $\lambda_2$  for a given genetic model.

$1.5 \times \min(p_{TT}, p_{TBT0})$  is the approximated  $p$ -value of TBT\*. Since both two-phase procedures are applied and the correlation between them is not available, we need to correct for multiple testing empirically. The true  $p$ -value of TBT\* should be between  $\min(p_{TT}, p_{TBT0})$  (without correction) and  $2 \times \min(p_{TT}, p_{TBT0})$  (with Bonferroni correction). Thus, a suitable 'ad hoc' correction could be  $1.5 \times \min(p_{TT}, p_{TBT0})$ . Figure 3 plots the power of  $\chi^2$ , TBT0 and TBT\*. Results show that Type I error rates for TBT\* are well controlled (see also Tables II and III for more results of Type I error rates for TBT\*). In addition, TBT\* is always more powerful than TBT0. It is also more powerful than  $\chi^2$  under the ADD, MUL and DOM models, but loses up to 10 per cent power compared with  $\chi^2$  under the REC model.

**3.3.3. Examples.** We apply TBT0 and TBT\* to the top two SNPs of Klein *et al.* [23]. For the top SNP with  $(r_0, r_1, r_2) = (50, 35, 11)$  and  $(s_0, s_1, s_2) = (6, 25, 19)$ , the  $T_{\text{HWDTT}}^2 = 1.1617$  with observed  $p$ -value  $p = 0.281 < 0.50$ , and  $T_{1/2}^2 / \chi^2 = 0.9877$  with the observed  $p$ -value  $p < 0.50$ . Both tests are significant in the first phase. In the second phase,  $\chi^2 = 26.510$  but the critical value for the level  $\alpha_2 = 1.5 \times \alpha / 103611 = 7.23 \times 10^{-7}$  is 28.2773. However, the CATT  $T_{1/2}^2 = 26.185$  with observed

Table I. Empirical power (per cent) of  $Z_1$  with level 0.5 with  $n=500$  case-control samples, prevalence  $K=0.1$  and risk allele frequency  $p=0.1, 0.3$  under four genetic models.

MAF $p$	Model	GRR	Power	Model	GRR	Power
0.1	REC	1.1	50.7	MUL	1.1	50.7
		1.2	50.4		1.2	53.7
		1.5	49.5		1.5	64.2
		2.0	46.5		2.0	84.3
	ADD	1.1	50.7	DOM	1.1	51.3
		1.2	53.3		1.2	57.9
		1.5	67.4		1.5	82.3
		2.0	88.6		2.0	97.6
0.3	REC	1.1	50.4	MUL	1.1	52.8
		1.2	51.1		1.2	59.4
		1.5	50.6		1.5	79.1
		2.0	54.3		2.0	96.1
	ADD	1.1	53.3	DOM	1.1	54.3
		1.2	58.9		1.2	61.5
		1.5	80.3		1.5	77.9
		2.0	96.5		2.0	90.8

$p$ -value =  $3.1 \times 10^{-7} < \alpha_2$ . Thus, we will reject the null hypothesis when TBT\* is used but not when TBT0 is used. The approximated observed  $p$ -value of the TBT\* is  $1.5 \times 3.1 \times 10^{-7} = 4.65 \times 10^{-7}$  (with Bonferroni-corrected  $p$ -value 0.0482). For the second top SNP,  $T_{1/2}^2/\chi_2^2 = 0.9661 > 0.50$ , the median of  $\text{Beta}(\frac{1}{2}, \frac{1}{2})$  ( $p$ -value 0.1161 from  $\text{Beta}(\frac{1}{2}, \frac{1}{2})$ ); hence,  $H_0$  is rejected in the first phase. In the second phase, we test the association using  $\chi_2^2$  at the level  $2\alpha/103611 = 9.65 \times 10^{-7}$  with a critical value 27.702 from the chi-square distribution with 2 df. The observed  $\chi_2^2 = 25.045$ . Thus, we cannot reject the null for the second top SNP. To apply TBT\*, given that TBT0 is not significant and the  $p$ -value of the HWDTT  $< 0.50$ , we only need to test the association using  $T_{1/2}$  at the level  $1.5\alpha/103611 = 7.32 \times 10^{-7}$  with the critical value 24.551. The observed  $T_{1/2}^2$  is 24.196, slightly greater than the critical value.

#### 4. SIMULATION STUDIES

Simulation studies were conducted to compare the sizes and power of single-phase analyses (CATT  $T_{1/2}$ , MAX3, CLRT, and Pearson's test  $\chi_2^2$ ) and two-phase analyses (APA, GMS and the modified TBT (TBT\*)) under the null and alternative hypotheses.

##### 4.1. Type I error rate

In simulation, equal number of cases and controls was used with a total sample size of  $n = 1000$  ( $r = s = 500$ ). Under  $H_0$ , genotype counts of cases and controls were simulated from  $\text{mul}(r; p_0, p_1, p_2)$  and  $\text{mul}(s; q_0, q_1, q_2)$ , where  $p_i = q_i = \text{Pr}(G_i)$ ,  $\text{Pr}(G_0) = (1-p)^2 + p(1-p)F$ ,  $\text{Pr}(G_1) = 2p(1-p)(1-F)$  and  $\text{Pr}(G_2) = p^2 + p(1-p)F$ , where  $F$  is Wright's inbreeding coefficient and  $p$  is the risk allele frequency. Case-control samples were simulated following the simulation procedures

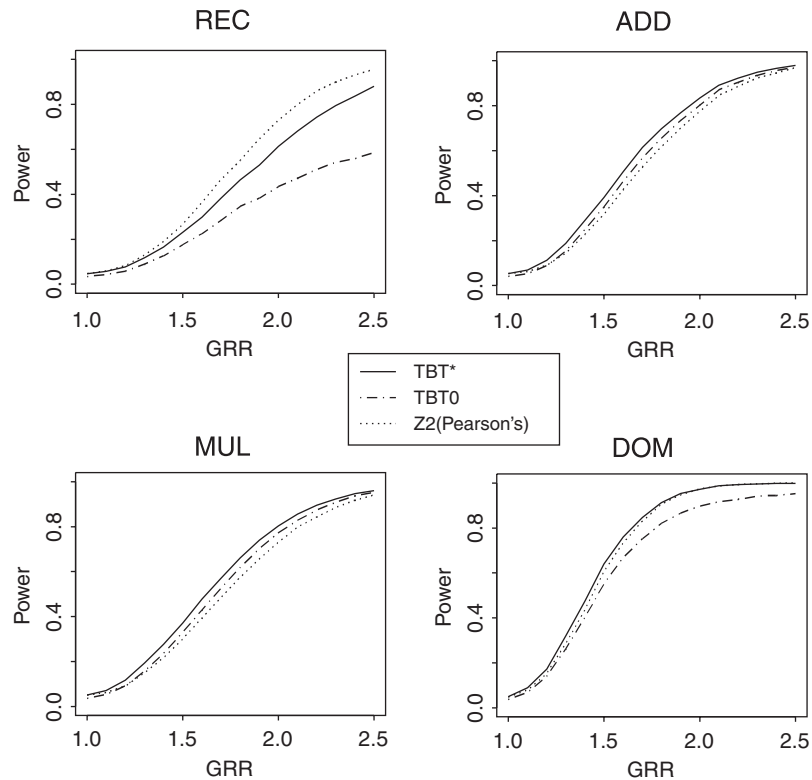


Figure 3. Empirical power of  $Z_2$  ( $\chi^2_2$ ), TBT0 ( $\alpha_1=0.5$ ) and TBT\* with  $n=500$  case-control samples, prevalence  $K=0.1$ , and risk allele frequency  $p=0.3$ . The nominal level for all tests is  $\alpha=5$  per cent.

used in Section 3. We chose  $F=0$  for HWE in the population and  $F=0.05$  without HWE in the population. In practice,  $F \leq 0.05$ . Parametric bootstrap was applied to APA, MAX3 and CLRT with 1000 bootstrap replications and each bootstrap replication was repeated 10 times. The estimated false-positive rates were based on 10 000 replications. The simulated Type I error rates are reported in Table II with nominal level 5 per cent. Table II shows that the Type I error rates of all the approaches are close to the nominal levels, including that for TBT\*. There is no impact of departure from HWE in the population after applying the parametric bootstrap correction. Since an approximation was used for TBT\*, we further examined the accuracy of this approximation by some independent simulations with significance level  $\alpha=0.0001$ . This significance level was used for genome-wide scans [17]. Results are reported in Table III. All the simulated Type I errors are close to  $\alpha=0.0001$ .

#### 4.2. Empirical power performance

Simulation procedure under the alternative hypothesis of association was similar to that given in Section 3 except that  $F$  was used here. We also chose  $n=2000$  and 1000 with equal numbers of cases and controls when GRR  $\lambda_2=1.1$  and  $\lambda_2=1.5$ , respectively, for small to moderate genetic effects. The risk allele frequencies were chosen as  $p=0.1, 0.3, 0.5$  and disease prevalences as

Table II. Type I error rates (per cent) from simulation with 10 000 replications, various risk allele frequencies  $p$  and disease prevalence  $K$ .

$F$	$K$	$p$	$T_{1/2}$	$\chi_2^2$	MAX3	APA	GMS	TBT*	CLRT	
0	0.01	0.1	5.01	3.96	5.57	5.18	4.08	4.89	5.70	
		0.2	5.30	5.35	5.64	5.22	5.46	5.56	5.60	
		0.3	4.77	4.96	5.01	4.99	5.14	5.37	5.09	
		0.4	5.40	5.31	5.60	5.34	5.53	5.56	5.49	
		0.5	5.21	5.10	4.97	5.04	5.17	5.49	5.08	
	0.10	0.1	5.17	3.80	4.87	5.07	3.87	5.14	5.34	
		0.2	4.97	4.84	4.90	4.78	4.87	5.03	5.01	
		0.3	4.80	4.78	4.65	4.59	4.71	5.18	4.76	
		0.4	5.21	5.55	4.66	4.24	5.33	5.11	4.62	
		0.5	5.30	5.45	5.36	5.14	5.40	5.51	5.38	
	0.05	0.01	0.1	5.43	4.80	5.90	5.84	4.79	5.81	5.91
			0.2	4.62	4.90	4.85	4.67	4.81	4.96	4.72
			0.3	5.27	5.29	5.13	5.10	5.35	5.37	5.30
			0.4	5.08	5.14	4.94	5.00	5.37	5.33	4.91
			0.5	5.01	4.82	4.86	4.70	4.91	4.90	4.90
0.10		0.1	4.98	4.48	4.76	4.92	4.45	5.11	4.54	
		0.2	4.89	4.97	4.99	4.96	4.96	4.85	5.10	
		0.3	5.30	5.20	5.21	5.07	5.26	5.39	5.20	
		0.4	4.58	4.77	4.56	4.38	4.72	4.67	4.50	
		0.5	5.53	5.34	5.18	5.08	5.37	5.55	5.24	

The total sample size is  $n=1000$  with equal number of cases and controls. The nominal level is 5 per cent.  $F$  is Wright's inbreeding coefficient.

Table III. Type I error rates for TBT from simulation with 100 000 replications, various risk allele frequencies  $p$  and disease prevalence  $K$ .

$F$	$K$	$p$				
		0.1	0.2	0.3	0.4	0.5
0	0.01	0.000080	0.000110	0.000101	0.000101	0.000124
	0.1	0.000094	0.000110	0.000105	0.000123	0.000130
0.05	0.01	0.000110	0.000106	0.000105	0.000110	0.000103
	0.1	0.000097	0.000111	0.000105	0.000106	0.000095

The total sample size is  $n=1000$  with equal number of cases and controls. The nominal level is 0.0001.  $F$  is Wright's inbreeding coefficient.

$K=0.01, 0.1$ . Four genetic models were considered. Under HWE, results were reported in Table III for  $K=0.01$  and Table IV for  $K=0.1$ .

First, we compare MAX3, CLRT, GMS and  $\chi_2^2$ . Tables III and IV show that MAX3 and CLRT have comparable power, and MAX3 is slightly more powerful than GMS, which is slightly more powerful than  $\chi_2^2$  under various genetic models, different allele frequencies and disease prevalences. In addition, MAX3 is always more powerful than the APA. Next, we compare  $T_{1/2}$ , MAX3 and

Table IV. Empirical power (per cent) from simulation with 10 000 replications, various risk allele frequencies  $p$  and various genetic models.

$p$	$\lambda_2$	Model	$T_{1/2}$	MAX3	$\lambda_2^*$	APA	GMS	CLRT	TBT*
0.1	1.1	REC	4.94	5.24	5.18	4.76	5.17	5.20	5.19
		ADD	7.77	7.33	6.55	7.83	6.63	6.94	7.70
		MUL	7.00	7.04	6.21	6.55	6.56	6.83	7.47
		DOM	13.68	11.62	11.18	11.90	11.13	11.20	13.30
	1.5	REC	6.16	8.63	7.67	7.35	7.98	8.71	6.80
		ADD	33.26	30.30	25.00	28.39	26.65	30.26	32.23
		MUL	29.60	26.07	22.23	23.66	23.92	26.61	29.00
		DOM	74.28	70.42	67.06	65.05	68.17	68.66	74.11
0.3	1.1	REC	6.50	9.30	8.00	8.10	7.90	9.70	7.50
		ADD	11.25	9.91	8.90	9.71	9.81	10.06	10.80
		MUL	10.72	9.49	8.49	8.94	9.39	9.45	10.40
		DOM	16.35	15.77	14.78	15.35	14.63	15.80	16.56
	1.5	REC	29.73	42.82	41.59	40.90	43.21	43.47	33.76
		ADD	59.77	55.44	49.63	49.94	53.80	55.93	57.62
		MUL	56.65	52.10	46.54	47.09	50.52	52.79	54.87
		DOM	82.94	85.08	83.22	82.20	84.03	84.39	84.99
0.5	1.1	REC	11.73	12.74	12.47	12.12	12.92	12.79	12.51
		ADD	11.84	10.87	9.86	10.48	10.84	10.79	11.65
		MUL	12.12	11.26	9.62	10.64	10.88	11.33	11.72
		DOM	11.72	12.59	12.19	12.06	12.35	12.51	12.07
	1.5	REC	68.71	77.34	74.99	73.72	75.24	77.20	72.40
		ADD	61.81	57.28	51.42	49.47	56.29	57.73	60.05
		MUL	62.38	57.69	52.08	49.61	57.16	57.91	60.70
		DOM	56.02	68.45	66.78	64.97	66.83	68.46	60.59

The total sample sizes are  $n=2000$  (1000) with equal number of cases and controls when  $\lambda_2=1.1$  (1.5). HWE holds in the population. The disease prevalence is  $K=0.01$ . The nominal level is  $\alpha=0.05$ . In APA,  $(\alpha_1^*, \alpha_2^*)$  are used.

TBT\*. We use +, ++ and +++ to indicate from least powerful to most powerful among the three tests (Table VI). Then the following table summarizes the comparison of the three tests under various situations. The results show that  $T_{1/2}$  is most powerful under the ADD and MUL models, whereas MAX3 is most powerful under the REC and DOM models. TBT\*, however, is neither most powerful nor least powerful. Finally, we compare among the two-phase analyses (TBT\* and GMS). The GMS is more powerful under the REC and DOM models (when the allele frequency  $p$  is moderate). The TBT\* is slightly more powerful than the GMS when  $p=0.1$  under the ADD, MUL and DOM models, whereas the GMS is more powerful under other situations. We plotted the empirical powers of the three two-phase analyses in Figure 4 under the four genetic models when HWE holds in the population. We chose GRR  $\lambda_2=1.0$  to 1.5 with an increment of 0.1. The empirical power was estimated based on 10 000 replications with risk allele frequency  $p=0.3$  and disease prevalence  $K=0.1$ . Figure 4 and numerical results show that (i) under the REC model, the GMS is most powerful and the TBT\* is least powerful; (ii) under the ADD and MUL models, the TBT\* is most powerful and the APA is least powerful; and (iii) under the DOM model, the APA is slightly more powerful when GRR is 1.1 and 1.2, and TBT\* is slightly more powerful when GRR is 1.3–1.5.

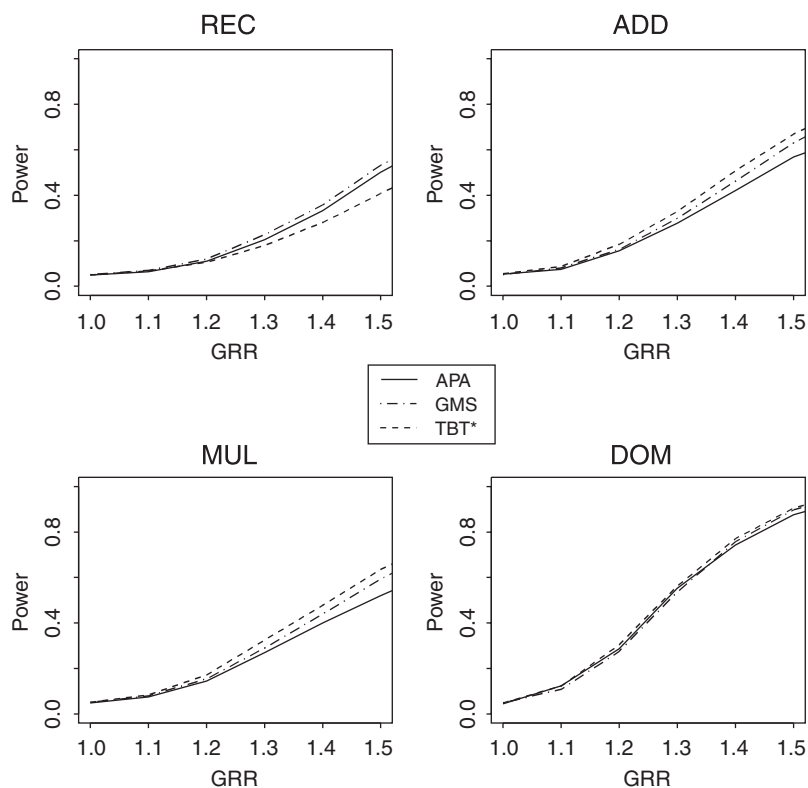


Figure 4. Empirical power of APA, GMS and TBT\* with  $n=1000$  case-control samples, prevalence  $K=0.1$  and risk allele frequency  $p=0.3$ .

## 5. APPLICATIONS TO REAL DATA

### 5.1. Genome-wide association study for type 2 diabetes

The first example is a genome-wide association study for type 2 diabetes [17]. The TBT0 is also included for comparison. Using 661 cases and 614 controls, eight SNPs identified were confirmed by the later replication analysis [17]. These eight SNPs are given in Table V (see also Table I and Tables 4S–5S of Sladek *et al.* [17]). Two different platforms were used: Illumina Infinium Human1 BeadArrays for 109 365 SNPs and Human Hap300 BeadArrays for 317 503 SNPs. In Table V, one SNP (rs1111875) was genotyped by both platforms; hence, both of them are included. Sladek *et al.* [17] applied MAX3 for genome-wide scans. The  $p$ -values of MAX3 and CLRT in Table V were obtained based on 10 000 000 parametric bootstrap simulations except for one SNP (rs7903146) whose  $p$ -value of MAX3 was taken from Sladek *et al.* [17] and  $p$ -value of CLRT could not be estimated using 10 million replicates. The observed  $p$ -values of all two-phase analysis procedures are also reported in Table V. For TBT0, if the first-phase analysis was not significant, the  $p$ -value was not reported. For APA, GMS and TBT\*, their first-phase analyses were all significant at their given levels. From Table V, the observed  $p$ -values of GMS are slightly less than those of MAX3 except for the last two SNPs. MAX3 is also slightly more powerful than CLRT except for the last

Table V. Empirical power (per cent) from simulation with 10 000 replications, various risk allele frequencies  $p$  and various genetic models.

$p$	$\lambda_2$	Model	$T_{1/2}$	MAX3	$\chi_2^2$	APA	GMS	CLRT	TBT*
0.1	1.1	REC	4.88	6.14	5.31	4.83	5.31	5.81	5.01
		ADD	7.57	7.34	6.75	7.10	6.86	7.36	7.57
		MUL	8.34	7.11	7.19	7.39	7.12	7.20	8.16
		DOM	15.05	13.37	12.45	14.41	12.14	12.74	14.73
	1.5	REC	6.04	9.30	8.77	8.61	8.77	9.64	6.77
		ADD	39.87	34.54	30.29	34.06	31.88	33.86	38.39
		MUL	35.66	30.34	26.53	28.43	28.87	29.92	34.63
		DOM	81.82	79.86	75.92	74.30	76.94	78.60	82.05
0.3	1.1	REC	7.56	8.70	8.49	8.90	8.62	8.68	7.74
		ADD	12.43	11.06	10.21	10.81	10.86	11.11	11.99
		MUL	12.35	11.64	10.37	11.93	11.04	11.69	12.51
		DOM	18.41	18.05	16.86	17.10	16.87	17.90	19.16
	1.5	REC	35.91	51.88	50.32	49.49	52.02	52.24	40.16
		ADD	68.73	64.51	58.19	56.83	62.29	64.49	66.34
		MUL	65.42	59.80	54.99	52.41	59.48	60.46	63.38
		DOM	88.63	90.84	89.21	87.51	89.64	90.68	90.34
0.5	1.1	REC	13.56	14.20	14.07	13.99	14.62	14.10	13.97
		ADD	13.07	11.07	10.56	11.50	11.60	11.23	12.64
		MUL	13.42	11.97	10.56	11.30	11.98	12.07	13.21
		DOM	13.10	13.22	13.01	13.18	13.53	13.22	13.56
	1.5	REC	77.93	85.39	83.91	82.33	84.03	85.17	81.30
		ADD	70.41	64.16	59.97	56.63	65.20	64.76	68.42
		MUL	70.69	65.72	60.59	56.59	65.34	66.22	68.44
		DOM	63.37	75.12	73.85	72.77	74.35	75.15	68.64

The total sample sizes are  $n=2000$  (1000) with equal number of cases and controls when  $\lambda_2=1.1$  (1.5). HWE holds in the population. The disease prevalence is  $K=0.1$ . The nominal level is  $\alpha=0.05$ . In APA,  $(\alpha_1^*, \alpha_2^*)$  are used.

Table VI. Comparison of performance among  $T_{1/2}$ , MAX3 and TBT\*.

Model	$T_{1/2}$	MAX3	TBT*
REC	+	+++	++
ADD	+++	+	++
MUL	+++	+	++
DOM	+	+++	++

The symbols +, ++, +++ indicate from least powerful to most powerful.

SNP. From this application, we show that MAX3, GMS and CLRT have similar performance and are often more powerful than  $\chi_2^2$ . The TBT\* has smallest  $p$ -values for rs7480010, rs7903146 and rs7923837. Note that the  $p$ -values of TBT0 are only half of those of Pearson's chi-square test if its first phase test is significant. The first-phase analysis of the TBT0 is usually not significant if the CATT has a larger  $p$ -value. Among the two-phase analysis procedures (including the TBT0), no single approach outperforms the others for these eight SNPs (Table VII).

Table VII. Observed  $p$ -values ( $\times 10^{-5}$  except for rs7903146 which has  $\times 10^{-17}$ ) of confirmed association results in Sladek *et al.* [17].

SNP	CATT	$\chi_2^2$	MAX3	APA	GMS	TBT*	TBT0	CLRT
rs11037909*	382	4.2	1.5	15.3	1.78	286	‡	1.68
rs1111875†	0.26	1.5	0.57	0.34	0.69	1.1	0.76	0.75
rs1111875*	0.27	1.6	0.53	0.36	0.72	1.2	0.81	0.69
rs1113132*	430	9.1	3.68	12.3	3.96	322	‡	4.25
rs13266634†	30.9	1.9	1.62	2.2	1.79	1.5	0.97	2.38
rs3740878*	436	4.2	1.47	8.7	1.77	327	‡	1.61
rs7480010†	2.2	4.4	2.02	1.7	5.36	1.6	2.2	2.45
rs7903146†	0.039	0.13	3.2	0.041	0.114	0.029	0.065	§
rs7923837†	0.086	0.28	0.26	0.071	0.23	0.064	0.14	0.11

The  $p$ -value of MAX3 for rs7903146 was given in Sladek *et al.* [17]. The  $p$ -value of CLRT for rs7903146 could not be estimated with 10 million replicates (its  $p$ -value  $< 10^{-7}$ ).

\*Using Hap300 BeadArrays.

†Using Human1 BeadArrays.

‡The first phase is not significant.

§The  $p$ -value cannot be estimated.

## 5.2. Association study of candidate gene for breast cancer

The second example is an association study of the progesterone receptor gene with breast cancer [25]. Using the CATT for the additive model, Pooley *et al.* [25] could not reject the association for SNP rs660149 with  $(r_0, r_1, r_2) = (327, 1719, 2421)$  and  $(s_0, s_1, s_2) = (334, 1855, 2353)$  at the level  $\alpha = 0.05$ . The three CATTs are  $(|T_0|, |T_{1/2}|, |T_1|) = (2.274, 1.832, 0.061)$ . The  $p$ -value of  $T_{1/2}$  is 0.0669. The  $T_{\text{HWDTT}}$  is 1.502 with a  $p$ -value 0.133. Using GMS, since  $|T_{\text{HWDTT}}| < 1.645$ ,  $T_{1/2}$  is chosen and the  $p$ -value of the GMS is greater than that of  $Z_{1/2}$ . Pearson's chi-square test is  $\chi_2^2 = 5.594$  with a  $p$ -value 0.061. Applying the APA,  $\alpha_1^* = 0.48$  and  $\alpha^* = 0.03$ . Thus, the first analysis of the APA is significant. But the second phase with  $T_{1/2}$  is not significant at the level  $\alpha_2^* = \alpha^* / \alpha_1^* = 0.0625$ . The  $p$ -value of the APA is 0.054. With 1 000 000 replications, MAX3 = 2.274 has  $p$ -value 0.05025, and CLRT = has  $p$ -value 0.05383. Applying TBT0,  $T_{1/2}^2 / \chi_2^2 = 0.60$ , significant at  $\alpha_1 = 0.5$ . Then, since the  $p$ -value of  $\chi_2^2$  is 0.061, the  $p$ -value of the TBT0 is  $0.061/2 = 0.0305 < 0.05$ . To apply TBT\*, the  $p$ -value of apply  $T_{\text{HWDTT}}$  and  $T_{1/2}$  is  $0.0669/2 = 0.0335$ . Thus, the approximated  $p$ -value for TBT\* is  $1.5 \times \min(0.0305, 0.0335) = 0.0458 < 0.05$ . Note that, in this application, only TBT0 and TBT\* are significant, whereas MAX3 is marginally significant. This application also shows that the GMS might not select the correct genetic model.

## 6. DISCUSSION

Although Van Steen *et al.* [26] first proposed a two-phase type of analysis for family-based association studies, the term 'two-phase analysis' is clearly defined in this article. The idea of analyses of Van Steen *et al.* [26], compared with the two-stage genotyping design of Skol *et al.* [7], has been studied for case-control association studies [8–10]. In this article, we have compared three two-phase analysis procedures to the existing robust test statistic MAX3, CLRT, the trend test

for the additive model, and Pearson's chi-square test with 2 df for case-control association studies. Our extensive simulation studies and applications demonstrated that MAX3, CLRT and two-phase analysis (GMS and TBT\*) often outperform other approaches (e.g. APA and Pearson's test with 2 df). The GMS is also more powerful than the APA. Comparing the GMS and the TBT\*, no approach outperforms the another. Among the existing two-phase analysis procedures, the TBT\* and APA are simplest to apply whereas the GMS needs moderate computation. On the basis of the results, we conclude that with an appropriate choice of significance level for phase 1, some two-phase analyses could be more powerful than commonly used single-phase test statistics.

The statistics that we studied can be applied to the candidate gene analysis and single-marker analysis for genome-wide association studies as well. For genome-wide association studies, single-marker genome-wide scan is usually the first step. For example, we applied these two-phase analysis procedures to confirmed SNPs from a genome-wide association study of diabetes [17]. The Wellcome Trust Case Control Consortium [27] used the better result of the CATT  $T_{1/2}$  and  $\chi^2_2$  for genome-wide scans, and Zheng *et al.* [28] used MAX3 for genome-wide scans. Our results showed that, in addition to candidate gene analysis, two-phase analyses (e.g. TBT\*) may also be used for genome-wide scans. The performance of applying these tests for genome-wide scans, however, is not studied here. Research has shown that a more powerful test usually results in better rankings of SNPs with true association [29, 30].

#### APPENDIX A: ASYMPTOTIC POWER OF THE HWDTT

Let  $\phi$  and  $\Phi$  be density and distribution functions of  $N(0, 1)$  and  $z_{1-\alpha}$  be the  $(1-\alpha)$ th fractile of  $N(0, 1)$ . Denote  $\sigma_0^2 = \lim_{n \rightarrow \infty} \{n^2/(rs)\} f(n_1/n, n_2/n)$ , where  $f(s, t) = (1-2t-s)^2 t(1-t) + 2st(t+s/2)(1-2t-s) + (t+s/2)^2 s(1-s)$  and  $\sigma_a^2 = n\sigma_{\text{HWDTT}}^2$ . Assume  $r/n \rightarrow \phi \in (0, 1)$  as  $n \rightarrow \infty$ . Then  $n^2/(rs) \rightarrow 1/\{\phi(1-\phi)\}$  and  $n_i/n \rightarrow \phi p_i + (1-\phi)q_i$ . The following asymptotic power of the HWDTT,  $\pi_{\text{HWDTT}} = \pi_{\text{HWDTT}}(p_i, q_i, \alpha_1)$ , was given in Zheng *et al.* [9]:

$$\pi_{\text{HWDTT}} = \Phi\left(\frac{-z_{1-\alpha_1/2}\sigma_0 - n^{1/2}(\Delta_1 - \Delta_0)}{\sigma_a}\right) + 1 - \Phi\left(\frac{z_{1-\alpha_1/2}\sigma_0 - n^{1/2}(\Delta_1 - \Delta_0)}{\sigma_a}\right)$$

#### APPENDIX B: ALGORITHMS FOR BOOTSTRAP CORRECTION

When the nominal level is  $\alpha$ , Type I error of the APA is inflated. The size of the APA is an increasing function of  $\alpha$ , and it is zero when  $\alpha=0$ . Thus, there exists a unique  $\alpha^* \in (0, \alpha)$  such that the size of the APA is inflated to  $\alpha$  when the nominal level is  $\alpha^*$  and when we treat the two analyses are independent for finite sample sizes. The parametric bootstrap procedure to find such  $\alpha^*$  for the case-control design is given below (see also [9]).

For a given  $\tilde{\alpha} \in (0, \alpha)$ , calculate the genotype frequencies using the pooled data:  $\hat{p}_i = \hat{q}_i = n_i/n$  for  $i=0, 1, 2$ . At the  $b$ th bootstrap ( $b=1, \dots, B$ ), generate genotype counts in cases and controls from  $\text{mul}(r; \hat{p}_0, \hat{p}_1, \hat{p}_2)$  and  $\text{mul}(s; \hat{p}_0, \hat{p}_1, \hat{p}_2)$ . The APA procedure is applied to all  $B$  bootstrap case-control samples. The false-positive rate at the nominal level  $\tilde{\alpha}$  is obtained, which is greater than  $\tilde{\alpha}$ . If this false-positive rate is close to the original nominal level  $\alpha$ , then the adjusted level

$\alpha^* = \tilde{\alpha}$ . The adjusted level  $\alpha^*$  is found by a grid search over  $(0, \alpha)$  with a small increment. The increment that we chose for the simulations in this article was 0.005.

### APPENDIX C: $p$ -VALUES FOR SOME TWO-PHASE ANALYSES

First, we consider the  $p$ -value for the two-phase analysis with a fixed  $\alpha_1$  for the first phase and  $\alpha_2 = \alpha/\alpha_1$  for the second phase. The null hypothesis is rejected when analyses in both phases are significant. Since  $\alpha_1$  is fixed and the analyses in two phases are independent, the  $p$ -value in the first-phase analysis does not contribute to the overall  $p$ -value in the two-phase analysis given  $p_1 < \alpha_1$ . Suppose we use  $Z_1$  and  $Z_2$  for the two phases and observe  $z_1$  and  $z_2$  from the two statistics. Then the  $p$ -value in the two-phase analysis is given by  $p = \Pr_{H_0}(|Z_2| > |z_2|, p_1 < \alpha_1)$  when the first phase is significant, where  $p_1$  is the  $p$ -value of the first phase. Thus, due to independence of the analyses in the two phases,  $p = \alpha_1 \Pr_{H_0}(|Z_2| > |z_2|) = \alpha_1 p_2$ , where  $p_2$  is the  $p$ -value of the second phase. Note that  $\alpha = \alpha_1 \alpha_2$ . Thus, the Type I error using this  $p$ -value is  $\Pr_{H_0}(p < \alpha, p_1 < \alpha_1) = \Pr_{H_0}(p_2 < \alpha_2, p_1 < \alpha_1) = \alpha_1 \alpha_2 = \alpha$ . The  $p$ -value  $p = \alpha_1 p_2$  can be applied to TBTO.

Second, we consider the  $p$ -value for the APA. The analyses in the two phases are correlated when  $\alpha_1^*$  and  $\alpha_2^* = \alpha/\alpha_1^*$  are used. When the correlation between the two analyses is ignored,  $\Pr_{H_0}(|T_{1/2}| > z_{1-\alpha/(2\alpha_1^*)}) > \alpha/\alpha_1^*$ . If the correlation is taken into account, then

$$\Pr_{H_0}(|T_{1/2}| > z_{1-\alpha/(2\alpha_1^*)} | \text{phase 1 analysis}) = \alpha/\alpha_1^*$$

The parametric bootstrap procedure is to find  $\alpha^*$  such that

$$\Pr_{H_0}(|T_{1/2}| > z_{1-\alpha^*/(2\alpha_1^*)}) = \alpha/\alpha_1^*$$

Thus, we use  $\alpha_2^* = \alpha^*/\alpha_1^*$  for the second phase, where  $\alpha^* < \alpha$  is the adjusted level. Using the adjusted level  $\alpha_2^* = \alpha^*/\alpha_1^*$  and ignore the correlation of the two phases, we would have the Type I error rate inflated close to the original nominal level  $\alpha$ . Therefore, the  $p$ -value has to take into account this inflation, which can be expressed as  $p = \alpha_1^* p_2(\alpha/\alpha^*)$ , where  $\alpha/\alpha^*$  is the inflation factor. To check this  $p$ -value, we can calculate the Type I error using this  $p$ -value as

$$\begin{aligned} \Pr_{H_0}(p < \alpha, p_1 < \alpha_1^*) &= \Pr_{H_0}(p_2 < \alpha^*/\alpha_1^*, p_1 < \alpha_1^*) = \Pr_{H_0}(p_2 < \alpha_2^*, p_1 < \alpha_1^*) \\ &= \Pr_{H_0}(p_2 < \alpha_2^*) \Pr_{H_0}(p_1 < \alpha_1^*) = (\alpha/\alpha_1^*) \alpha_1^* = \alpha \end{aligned} \quad (C1)$$

In (C1), the events  $\{p_1 < \alpha_1^*\}$  and  $\{p_2 < \alpha_2^*\}$  are treated as independent when  $\alpha_2^* = \alpha^*/\alpha_1^*$  is used.

### APPENDIX D: DISTRIBUTION THEORY FOR TBT

Denote  $g_i = \Pr(G_i)$  for  $i = 0, 1, 2$ . Let  $\delta = (\hat{p}_1 - \hat{q}_1, \hat{p}_2 - \hat{p}_2)^t$ , where t stands for transpose. Then, under the null  $H_0$ ,

$$\Sigma = \text{cov}(\delta) = \frac{1}{n\phi(1-\phi)} \begin{pmatrix} g_1 - g_1^2 & -g_1 g_2 \\ -g_1 g_2 & g_2 - g_2^2 \end{pmatrix}$$

where  $\phi = r/n$ . Its estimate under  $H_0$  is denoted by  $\hat{\Sigma}$  by substituting  $g_i$  by the estimate  $\hat{g}_i = n_i/n = \phi \hat{p}_i + (1 - \phi) \hat{q}_i$ . It can be shown that Pearson's chi-square test for the  $2 \times 3$  table, (1), can be asymptotically expressed as

$$\chi_2^2 = n\phi(1 - \phi) \left( \frac{(\hat{p}_0 - \hat{q}_0)^2}{\hat{g}_0} + \frac{(\hat{p}_1 - \hat{q}_1)^2}{\hat{g}_1} + \frac{(\hat{p}_2 - \hat{q}_2)^2}{\hat{g}_2} \right) = \delta^t \hat{\Sigma}^{-1} \delta$$

and the CATT indexed by  $x$  as

$$T_x^2 = n\phi(1 - \phi) \frac{(x\hat{p}_1 + \hat{p}_2 - x\hat{q}_1 - \hat{q}_2)^2}{x^2\hat{g}_1 + \hat{g}_2 - (x\hat{g}_1 + \hat{g}_2)^2} = \frac{(a^t \delta)^2}{a^t \hat{\Sigma} a}$$

where  $a = (x, 1)^t$ .

Let  $W_2 = W_2(\Sigma) = \delta^t \Sigma^{-1} \delta$  and  $W_x = W_x(\Sigma) = (a^t \delta)^2 / a^t \Sigma a$  for  $x = 0, 1/2, 1$ . Apparently, Pearson's chi-square test is  $\chi_2^2 = W_2(\hat{\Sigma})$  and the CATT is  $T_x^2 = W_x(\hat{\Sigma})$ . Let  $y = \Sigma^{-1/2} \delta$  and  $\tilde{a} = \Sigma^{1/2} a / \sqrt{a^t \Sigma a}$ . Then  $W_2(\Sigma) = y^t y$  and  $W_x(\Sigma) = (\tilde{a}^t y)^2$ . Note that  $\tilde{a}^t \tilde{a} = 1$ . Thus  $W_2 \geq W_x$ . Furthermore,  $W_2 - W_x = (\tilde{b}^t y)^2$ , where  $\tilde{b} = (\tilde{a}_2, -\tilde{a}_1)^t$ , which is orthogonal to  $\tilde{a}$  with norm 1. This implies that, under  $H_0$ ,  $W_2 - W_x$  is asymptotically chi-square distributed with 1 df and it is asymptotically independent of  $W_x$ .

It is well known that if  $X$  and  $Y$  are statistically independent and  $X, Y \sim \chi_1^2$ , then

$$\frac{X}{X + Y} \sim \text{Beta}(1/2, 1/2)$$

a beta distribution with density  $f(x; \alpha, \beta) = \Gamma(\alpha + \beta) / \{\Gamma(\alpha)\Gamma(\beta)\} x^{\alpha-1} (1-x)^{\beta-1}$ . Therefore, asymptotically,

$$W_x / W_2 = W_x / [W_x + (W_2 - W_x)] = W_x(\Sigma) / W_2(\Sigma) \sim \text{Beta}(1/2, 1/2)$$

Substituting  $\hat{\Sigma}$ , we have asymptotically

$$T_x^2 / \chi_2^2 = W_x(\hat{\Sigma}) / W_2(\hat{\Sigma}) \sim \text{Beta}(1/2, 1/2)$$

which is independent of Pearson's chi-square test  $\chi_2^2$ .

## COMPUTING PROGRAMS

The SAS programs for simulations and applications are available upon request.

## ACKNOWLEDGEMENTS

The work of Mark Meyer is completed when he is a summer intern at Office of Biostatistics Research, National Heart, Lung and Blood Institute. The research of Yaning Yang was supported by China NSF Grant 10671189 and Chinese Academy of Science Grant No. KJCX3-SYW-S02. Wentian Li acknowledges the support from The Robert S Boas Center for Genomics and Human Genetics at the Feinstein Institute for Medical Research. The views of this article only present those of the authors, not representing any of the National Heart, Lung and Blood Institute, National Institutes of Health. The authors would like to thank three referees for their valuable comments and suggestions which improved our presentations.

## REFERENCES

1. Cochran WG. Some methods for strengthening the common chi-square tests. *Biometrics* 1954; **10**:417–451.
2. Armitage P. Tests for linear trends in proportions and frequencies. *Biometrics* 1955; **11**:375–386.
3. Sasieni PD. From genotypes to genes: doubling the sample size. *Biometrics* 1997; **53**:1253–1261.
4. Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend tests for case–control studies of genetic markers: power, sample size and robustness. *Human Heredity* 2002; **53**:146–152.
5. Zheng G, Freidlin B, Gastwirth JL. Comparison of robust tests for genetic association using case–control studies. *Second Lehmann Symposium—Optimality*. IMS Lecture Notes—Monograph Series, vol. 49. IMS: Beachwood, OH, 2006; 253–265.
6. Wang K, Sheffield VC. A constrained-likelihood approach to marker-trait association studies. *American Journal of Human Genetics* 2005; **77**:768–780.
7. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature Genetics* 2006; **38**:209–213.
8. Elston RC, Lin DY, Zheng G. Multistage sampling for genetic studies. *Annual Review of Genomics and Human Genetics* 2007; **8**:327–342.
9. Zheng G, Song K, Elston RC. Adaptive two-stage analysis of genetic association in case–control designs. *Human Heredity* 2007; **63**:175–186.
10. Zheng G, Ng HKT. Genetic model selection in two-stage analysis for case–control association studies. *Biostatistics* 2008; DOI: 10.1093/biostatistics/kxm039.
11. Song K, Elston RC. A powerful method of combining measures of association and Hardy–Weinberg disequilibrium for fine-mapping in case–control studies. *Statistics in Medicine* 2006; **25**:105–126.
12. Slager SL, Schaid DJ. Case–control studies of genetic markers: power and sample size approximations for Armitage’s test for trend. *Human Heredity* 2001; **52**:149–153.
13. Zheng G, Freidlin B, Li Z, Gastwirth JL. Choice of scores in trend tests for case–control studies of candidate–gene associations. *Biometrical Journal* 2003; **45**:335–348.
14. Zheng G, Gastwirth JL. On estimation of the variance in Cochran–Armitage trend tests for genetic association using case–control studies. *Statistics in Medicine* 2006; **25**:3150–3159.
15. Agresti A. *Categorical Data Analysis*. Wiley: New York, 1990.
16. Balding DJ. A tutorial on statistical methods for population association studies. *Nature Review Genetics* 2006; **7**:781–791.
17. 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, Pshzhetsky 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.
18. Gastwirth JL. The use of maximin efficiency robust tests in combining contingency tables and survival analysis. *Journal of the American Statistical Association* 1985; **80**:380–384.
19. Freidlin B, Podgor MJ, Gastwirth JL. Efficiency robust tests for survival or ordered categorical data. *Biometrics* 1999; **55**:883–886.
20. Zaykin DV, Nielsen DM. Hardy–Weinberg disequilibrium (HWD) fine mapping for case–control samples. *American Journal of Human Genetics* 2000; **67**(Suppl.):1238.
21. Wittke-Thompson JK, Pluzhnikov A, Cox NJ. Rational inferences about departure from Hardy–Weinberg equilibrium. *American Journal of Human Genetics* 2005; **76**:967–986.
22. Suh YJ, Li W. Genotype-based case–control analysis, violation of Hardy–Weinberg equilibrium, and phase diagram. *Proceedings of the 5th Asia–Pacific Bioinformatics Conference*. Imperial College Press: U.K., 2007; 185–194.
23. 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.
24. AREDS Research Group. Risk factors associated with age-related macular degeneration: a case–control study in the age-related eye disease study: age related eye disease study report number 3. *Ophthalmology* 2000; **107**:2224–2232.
25. Pooley KA, Healey CS, Smith PL, Pharoah PD, Thompson D, Tee L, West J, Jordan C, Easton DF, Ponder BA, Dunning AM. Association of the progesterone receptor gene with breast cancer risk: a single-nucleotide polymorphism tagging approach. *Cancer Epidemiology, Biomarkers and Prevention* 2006; **15**:675–782.

26. Van Steen K, McQueen MB, Herbert A, Raby B, Lyon H, DeMeo DL, Murphy A, Su J, Datta S, Rosenow C, Christman M, Silverman EK, Laird NM, Weiss ST, Lange C. Genomic screening and replication using the same data set in family-based association testing. *Nature Genetics* 2005; **37**:683–691.
27. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**:661–683.
28. Zheng G, Joo J, Lin JP, Stylianou M, Waclawiw MA, Geller NL. Robust ranks of true associations in genome-wide case–control association studies. *BMC Proceedings* 2007; **1**(Suppl. 1):S165.
29. Zaykin DV, Zhivotovsky LA. Ranks of genuine associations in whole-genome scans. *Genetics* 2005; **171**:813–823.
30. Gail MH, Pfeiffer RM, Wheeler W, Pee D. Probability of detecting disease-associated single nucleotide polymorphisms in case–control genome-wide association studies. *Biostatistics* 2008; **9**:201–215.