

Simultaneously Correcting for Population Stratification and for Genotyping Error in Case-Control Association Studies

K. F. Cheng and W. J. Lin

In population-based case-control association studies, the regular χ^2 test is often used to investigate association between a candidate locus and disease. However, it is well known that this test may be biased in the presence of population stratification and/or genotyping error. Unlike some other biases, this bias will not go away with increasing sample size. On the contrary, the false-positive rate will be much larger when the sample size is increased. The usual family-based designs are robust against population stratification, but they are sensitive to genotype error. In this article, we propose a novel method of simultaneously correcting for the bias arising from population stratification and/or for the genotyping error in case-control studies. The appropriate corrections depend on sample odds ratios of the standard 2×3 tables of genotype by case and control from null loci. Therefore, the test is simple to apply. The corrected test is robust against misspecification of the genetic model. If the null hypothesis of no association is rejected, the corrections can be further used to estimate the effect of the genetic factor. We considered a simulation study to investigate the performance of the new method, using parameter values similar to those found in real-data examples. The results show that the corrected test approximately maintains the expected type I error rate under various simulation conditions. It also improves the power of the association test in the presence of population stratification and/or genotyping error. The discrepancy in power between the tests with correction and those without correction tends to be more extreme as the magnitude of the bias becomes larger. Therefore, the bias-correction method proposed in this article should be useful for the genetic analysis of complex traits.

Population-based case-control studies provide a powerful approach to identify the multiple variants of small effect that modulate susceptibility to common, complex diseases. However, the major shortcoming of these studies arises from the presence of population stratification (PS). When cases and controls have different allele frequencies attributable to diversity in background population, unrelated to the disease being studied, the study is said to have PS. PS is probably the most-often-cited reason for nonreplication of genetic association studies, since undetected PS can mimic the signal of association and lead to more false-positive findings or miss real effects.^{1,2} Besides these factors, often mentioned in the literature, an often-overlooked factor influencing the performance of case-control design is the presence of genotyping error (GE). Such error is important because without some method of correction, the power to detect association and thus to map genes may be significantly decreased.³⁻⁶

Family-based designs are robust against PS. However, under the assumption of no or small PS, case-control studies have been shown to be more powerful than family-based designs.^{7,8} Unfortunately, it is rarely clear when PS can be ignored. The existence of PS, in general, weights against the use of case-control designs. Using population-based data, Devlin and Roeder⁹ proposed an association method, termed "genomic control" (GC), to automatically

correct for the effects caused by PS and cryptic relatedness. Another computationally more extensive approach for correcting the effects of PS is the structured association (SA) method.¹⁰ Both GC and SA methods require genotyping at additional null loci to perform the tests. Bacanu et al.¹¹ claimed that the transmission/disequilibrium test (TDT) is more powerful when population substructure is substantial and that the GC is more powerful otherwise. However, a recent study by Campbell et al.¹² showed that both standard GC and SA methods failed to correct for the confounding effects of PS. The original TDT, GC, and SA methods are not intended to correct the bias due to GE. Recently, extensions of TDT methods to correct for nondifferential genotype error have been proposed.^{5,13-15} Clayton et al.¹⁶ also suggested that the idea of GC can be generalized to correct for the effects of differential errors in measurement of genotype. In their application, the variance inflation factor is not constant but depends on extra measures of genotyping accuracy, such as the half-call rate and the absolute difference in call rates between cases and controls.

The GC method is based on the assumption that variance inflation factor is approximately constant across the genome for all null loci. However, many results¹⁷⁻¹⁹ showed that the regular χ^2 statistic for testing independence follows a noncentral χ^2 distribution asymptotically

From the Biostatistics Center and Department of Public Health, China Medical University (K.F.C.), and Graduate Institute of Statistics, National Central University (W.J.L.), Taiwan

Received May 29, 2007; accepted for publication June 18, 2007; electronically published August 22, 2007.

Address for correspondence and reprints: Dr. K. F. Cheng, Biostatistics Center, China Medical University, Taichung, Taiwan. E-mail: kfcheng@mail.cmu.edu.tw

Am. J. Hum. Genet. 2007;81:726-743. © 2007 by The American Society of Human Genetics. All rights reserved. 0002-9297/2007/8104-0016\$15.00
DOI: 10.1086/520962

under stratified populations, even when there is no true association. They also showed that the noncentrality parameter can be large even when Wright's²⁰ F_{st} is small. Here, F_{st} measures a sort of inbreeding coefficient, or heterozygote deficit, that is due to population subdivision. A new correction for PS was recently suggested by Epstein et al.²¹ using substructure-informative loci, instead of the usual null loci. This method was shown to have improved performance but was not designed to protect against the confounding effects of GE.

It is well known that, for a single SNP locus, if the GE is random nondifferential with respect to affected status, then there is no effect on the expected type I error rate. However, its effect on the power is well recognized. Clayton et al.¹⁶ pointed out that there might exist different error rates in genotype scoring between case and control samples. Under this circumstance, Moskvina et al.⁶ used simulations showing that, even with very low error rates, differential error rates can result in false-positive rates much greater than 0.05. The effect was maximal for loci with small minor-allele frequency.^{4,22} The bias caused by PS and/or GE can be substantial, and it will not go away with increasing sample size. In fact, the false-positive rate will be much larger when the sample size is increased. The purpose of this article is to suggest a novel method that can automatically correct for the effects arising from PS and/or GE in case-control studies, only at the price of genotyping a panel of null loci. We remark that the usual approach for correcting the bias caused by GE is to assume an error model and often requires repeated genotyping or validation data. In contrast, the method proposed here does not depend on either an error model or validation data. To the best of our knowledge, this is the first article that gives a systematic study of the joint effect of PS and GE and provides a workable solution for correcting the related bias in case-control association studies.

In this article, we point out that, under the null hypothesis of no association, the confounding effect caused by PS depends on the sampling proportions and genotype frequencies of the subpopulations. In the special case of simple random sampling, it also depends on the disease risks in subpopulations. We show how to use information from null loci to estimate the PS effect efficiently and, on that basis, suggest a genotype-based χ^2 test (with 2 df) (hereafter called the "CS" test) for testing the existence of association. This method is very simple to apply and can be easily extended to provide a point (and interval) estimation of the genetic effect if the null hypothesis of no association is rejected.

When there are genotyping errors, the CS test also can correct the bias caused by GE. This is because the likelihood functions for testing the null hypothesis of no association under GE and PS have the same form. We will give reasons showing why this is true. In fact, even in situations where error rates are not constant within or between case and control samples, the CS test can still be applied to test no association. When there is no PS and

GE, the CS test automatically reduces to the regular χ^2 test if the sizes of the case and control samples are large. This means that the CS test is a natural extension of the regular χ^2 test for correcting the effects of PS and/or GE.

In this article, we also present simulation results, to illustrate the performance of the CS test. Under various simulation parameter values—which were very similar to those found in real-data examples—for PS and/or GE, the CS test was shown to approximately maintain the expected false-positive rate. In contrast, the regular χ^2 test tended to have inflated type I error rates. In most simulated instances, the CS test also showed improved power performance. Often, the increases in power were very significant. We report simulation results for the CS test on the basis of data from the candidate locus and 50 randomly selected null loci. Evidence from the simulation study also indicates that no advantage can be found by using a greater number of null loci in the analysis.

Material and Methods

CS Test for Correcting Bias Caused by Population Stratification

In case-control studies, the data for each locus are given in a standard 2×3 table of genotype by case and control. Let $D = 1$ denote that the individual has the disease and $D = 0$ otherwise. Let G (equal to 0, 1, or 2) denote the number of copies of the high-risk candidate allele carried by the individual. The primary interest is to test whether there exists association between the genetic risk factor G and disease D . We assume that the general population comprises K subpopulations, and covariable S is used to indicate the subpopulation to which a person belongs. We postulate the risk model²³

$$\log \left[\frac{\Pr(D = 1 | G = g, S = s)}{\Pr(D = 0 | G = g, S = s)} \right] = \mu + \mu_s + \beta_g. \quad (1)$$

For identifiability, we define μ_1 and β_0 to be zero, so that $s = 1$ and $g = 0$ represent the referent subpopulation and genotype, respectively. Model (1) assumes S to be a confounder, not an effect modifier.

In the presence of PS, one can show (appendix A) that even when there exists no association between G and D , the ratio of the case and control genotype frequencies can be expressed as

$$\frac{\Pr(G = g | D = 1)}{\Pr(G = g | D = 0)} = \exp(\alpha^* + \beta_g^*), \quad (2)$$

where $\beta_0^* = 0$ but parameters β_1^* and β_2^* depend on the genotype frequency $\Pr(G = g | S = s)$ and the sampling proportions $P^*(S = s | D = 1)$ and $P^*(S = s | D = 0)$ of the diseased and non-diseased individuals, respectively, from subpopulation S (see eqs. (A2) and (A3) in appendix A for the exact definition of β_g^*). Note that the values of $P^*(S = s | D = 1)$, $P^*(S = s | D = 0)$, and $\Pr(G = g | S = s)$ are assumed to be unknown, but they are not required to be estimated in the analysis. According to model (2), β_1^* and β_2^* are log odds ratios of the 2×3 table under no association. Thus, $\beta^* = (\beta_1^*, \beta_2^*)$ can be used to measure the level of PS. Model (2) implies that, even when there is no true association

between G and D , the case and control genotype frequencies cannot be identical if PS exists (i.e., β^* does not equal zero). This makes the regular χ^2 statistic for testing independence in a 2×3 table produce spurious association. In view of the definition of β_g^* in appendix A, if one can identify genetically distinct subpopulations and uses a design so that the sampling proportions are identical in the cases and controls (i.e., $P^*(S = s|D = 1) = P^*(S = s|D = 0)$), then the false-positive rate of the regular χ^2 test will not be elevated, since no PS effect ($\beta^* = 0$) exists in this case. Otherwise, the effect of PS might be severe because of the different sampling proportions used in the cases and controls. Note that, in the special case of simple random sampling, the level of PS also depends on the disease risks in subpopulations (see eq. (A4) in appendix A).

The level of PS is locus dependent. We use $\beta^*(l)$ to denote the level of PS corresponding to the l th null locus, $l = 1, \dots, L$. The idea of the CS test is to first combine estimates $\hat{\beta}^*(l)$ of the PS levels at the null loci, to define a reasonable estimate $\tilde{\beta}^*$ of β^* , the PS level at the candidate locus. Next, using model (2), we define the CS test statistic (denoted by " $X^2(\tilde{\beta}^*)$ ") to be the regular likelihood-ratio test statistic for testing $H_0: \beta^* = \tilde{\beta}^*$ based on genotype data at the candidate locus. Define $N_0(g)$ and $N_1(g)$ to be the numbers of individuals in the control and case samples, respectively, with genotype $G = g$ at the candidate locus. Under model (2), the retrospective likelihood function²⁴ is

$$L = \prod_{g=0}^2 \left[\frac{1}{1 + \exp(\alpha_g^* + \beta_g^*)} \right]^{N_0(g) + N_1(g)} \times [\exp(\alpha_g^* + \beta_g^*)]^{N_1(g)}.$$

Let $\hat{L}(H_0)$ be the maximum likelihood under constraint H_0 and \hat{L} be the maximum likelihood under no constraint. The CS test statistic is defined as $X^2(\tilde{\beta}^*) = 2 \log[\hat{L}/\hat{L}(H_0)]$. By use of existing software packages, the CS test statistic can be computed easily. The corresponding P value of the test is $P = \Pr[\chi_2^2 > X^2(\tilde{\beta}^*)]$, where χ_2^2 has a χ^2 distribution with 2 df.

In this article, we define estimate $\hat{\beta}^*(l)$ to be the usual log of the sample odds ratio (maximum-likelihood estimate), using 2×3 genotype data at the l th null locus. Conceptually, if subpopulation genotype frequencies at the candidate locus approximately match those at the null loci, then the usual mean or median of $\hat{\beta}^*(l)$ can be a good estimate of β^* . However, it is difficult to verify this condition in real applications. Instead, we assume β^* to be unknown but a smooth function of the genotype frequencies in the controls (at least approximately) and suggest using a nonparametric regression technique²⁵ to estimate β^* . We let the sample genotype frequencies of the candidate locus and l th null locus in the controls be denoted by $P_0(g)$ and $P_l(g)$, respectively, and define the difference of the two frequencies as $d_l(g) = P_0(g) - P_l(g)$. A nonparametric regression estimate of β^* is defined as $\hat{\beta}^* = \sum_{l=1}^L \hat{\beta}^*(l) W_l$. This is a weighted average of $\hat{\beta}^*(l)$, with weights defined as

$$W_l = \frac{K[d_l(0)/b_n]K[d_l(1)/b_n]}{\sum_{r=1}^L K[d_r(0)/b_n]K[d_r(1)/b_n]}.$$

The weights are determined by "window size" $b_n > 0$ and the "quadratic kernel" $K(t) = 3(1 - t^2)I(|t| \leq 1)/4$.

It is well known that the performance of the nonparametric regression estimate is insensitive to the use of kernel function. However, it depends on the window size. We suggest that an optimal b_n be selected so that the proposed CS test applied to

each null locus can approximately maintain the correct type I error rate. To this end, for the l th null locus, we let $P_l = \Pr\{\chi_2^2 > X^2[\hat{\beta}^*(l)]\}$ denote the P value, where the nonparametric regression estimate $\tilde{\beta}^*(l)$ is computed from the genotype data at the remaining $L - 1$ null loci, with b_n fixed. Next, for a prespecified level of significance α , we propose choosing an optimal b_n (α dependent) from $(0, 1)$ so that $|L^{-1} \sum_{l=1}^L I(P_l < \alpha) - \alpha|$ is minimized. A free software (CS test software) for computing optimal window size b_n , an estimate of β^* , and the final P value of the CS test is available at Cheng's software Web site.

CS Test for Correcting Bias Caused by Genotyping Error

In this section, we show that the CS test also can be applied to correct the bias caused by GE in case-control studies. For simplicity, we assume that case and control samples have differential genotype error rates, but it is understood that our approach can be applied under more-general error modeling. For example, our approach still works even when there are differential error rates within the case (or control) sample. Let G_o (equal to 0, 1, 2) be the observed genotype, subject to genotyping error. We assume that the error rates are $\Pr(G_o = g_o | G = g, D = 1) = \phi_1(g_o; g)$ in the case sample and $\Pr(G_o = g_o | G = g, D = 0) = \phi_0(g_o; g)$ in the control sample. Thus, if one defines $W_1(g_o; g) = \phi_1(g_o; g) \Pr(G = g | D = 1)$ and $W_0(g_o; g) = \phi_0(g_o; g) \Pr(G = g | D = 0)$, then, under no true association, one can show that the ratio of the case and control genotype frequencies is

$$\frac{\Pr(G_o = g_o | D = 1)}{\Pr(G_o = g_o | D = 0)} = \log(\delta + \gamma_{g_o}), \quad (3)$$

where parameters γ_1 and γ_2 depend on $W_0(g_o; g)$ and $W_1(g_o; g)$ (appendix B), and their values may be nonzero if error rates $\phi_1(g_o; g)$ and $\phi_0(g_o; g)$ are not identical. Thus, even under the null case, there may exist nonzero log odds ratios in the 2×3 table. In this case, there exists bias because of GE. On the other hand, if error rates $\phi_1(g_o; g)$ and $\phi_0(g_o; g)$ are identical, then there is no effect on the expected type I error rate, since $\gamma_1 = \gamma_2 = 0$.

Suppose that, using the same genotyping technique, we also have genotype data from the null loci. For the l th null locus, let the corresponding bias be denoted by $\gamma(l) = [\gamma_1(l), \gamma_2(l)]$. This bias also can be estimated by use of the log of the sample odds-ratios from the l th null loci (denoted by $\hat{\gamma}(l)$). Next, using the same principle as above, we also define estimate $\hat{\gamma}$ of the bias (γ_1, γ_2) to be a weighted average of $\hat{\gamma}(l)$. Thus, on the basis of the observed 2×3 table at the candidate locus, the regular likelihood-ratio test for testing $H_0: (\gamma_1, \gamma_2) = (\hat{\gamma}_1, \hat{\gamma}_2)$ under model (3) is exactly identical to the CS test defined above. It is important to note that errors may not be distributed evenly across all loci—that is, the error rates are also locus dependent. Some loci may show error rates that are many times higher than those shown by other loci.²⁵ However, the validity of the CS test does not require error rates to be identical across candidate and null loci. We conclude that, in an association analysis, the CS test can be applied to correct for PS and GE simultaneously.

Simulations

We conducted several simulations to investigate the performance of the CS test and the regular χ^2 test without adjustment (hereafter called the "CS*" test) under PS and/or GE. We included the CS*

test in the study so that the empirical level of the bias caused by PS and/or GE could be measured.

There are three factors affecting the level of PS (appendix A): (i) the sampling proportions for each subpopulation among cases and controls, (ii) the allele frequency at the candidate locus in each subpopulation (under the assumption that the Hardy-Weinberg condition holds in each subpopulation), and (iii) the penetrances of the candidate locus in each subpopulation. In our simulation study, the general population was assumed to comprise two subpopulations, and the case data were sampled from the first and second subpopulations with probabilities $q = P^*(S = 1|D = 1)$ and $1 - q = P^*(S = 2|D = 1)$, respectively, and the control data were sampled from the first and second subpopulations with probabilities $1 - q = P^*(S = 1|D = 0)$ and $q = P^*(S = 2|D = 0)$, respectively. Three q values were used: 0.5, 0.7, and 1.0. $q = 0.5$ corresponds to the case of no PS effect, since the level of PS is zero. $q = 1.0$ corresponds to the case with the most severe PS effect. In this situation, case and control samples were drawn from two different subpopulations. Zheng et al.²⁷ also considered this extreme case in their simulation study.

The allele frequency at the candidate locus was chosen to be $p_1 = 0.30$ for the first subpopulation and $p_2 = 0.30 + t$ for the second subpopulation. A large difference, t , between the allele frequencies in the two subpopulations means that a large bias due to PS occurs in the study. In the simulations, $t = 0.03, 0.05,$ and 0.10 were considered, representing the range from weak PS to strong PS. Note that, on the basis of the International Project on Genetic Susceptibility to Environmental Carcinogens database, Garte et al.^{28,29} pointed out the differences in allele frequencies within white populations from different countries are much smaller (e.g., $t \leq 0.05$) but more significant among whites, Asians, and African Americans. For example, the allele frequency of the CYP3A4-V gene, which is thought to be related to prostate cancer, was highest among Nigerians (87%), lowest among European Americans (10%), and intermediate among African Americans (66%).³⁰ Therefore, our choices of frequency differences are consistent with real-data examples.

Finally, in the null and power simulations, the same penetrances were used for the two subpopulations. Under null simulations, identical penetrances $f_0 = f_1 = f_2 = 0.10$ were used. Under power simulations, penetrances $f_0 = 0.01$ and $f_1 = f_2 = 0.25$ were used for the dominant genetic model, $f_0 = f_1 = 0.10$ and $f_2 = 0.30$ were used for the recessive genetic model, and $f_0 = 0.10, f_1 = 0.20,$ and $f_2 = 0.30$ were used for the additive genetic model. Note that the penetrances are defined as $f_g = \Pr(D = 1|G = g)$, and similar values were also considered in the simulation study by Zheng et al.²⁷

Next, according to the definition of the bias caused by GE (eq. (3)), there are three factors affecting the bias level: (i) the genotype frequencies of the cases and controls, (ii) error models, and (iii) error rates. The genotype frequencies were defined above. Two error models were considered in the simulations. The first model is the symmetric allele-dropout error model,⁵ determined by one error rate, ε . This model assumes that one misclassifies homozygotes twice as frequently as heterozygotes. The second model is the allele-based error model,¹² determined by two error rates, ε_1 and ε_2 . In this model, the high-risk allele has constant probability ε_1 of being coded as a normal allele, and a normal allele has constant probability ε_2 of being coded as a high-risk allele. In the simulations, the same error model was used for the cases and controls to generate misclassified genotype data, but with different error rates. In the allele-based error model, the error rates

used for the cases were $\varepsilon_1 = 0$ and $\varepsilon_2 = 0.01$, but $\varepsilon_1 = 0.05$ and $\varepsilon_2 (= \varepsilon) = 0.01, 0.03, 0.05$ were used for the controls. On the other hand, in the symmetric allele-dropout error model, the error rate used for the cases was 0.01, but error rates $\varepsilon = 0.01, 0.03, 0.05$ were used for the controls. Only under the symmetric allele-dropout error model with $\varepsilon = 0.01$ does there exist no GE effect on type I error rate. Note that Tintle et al.³¹ reported that an 8% error rate is the maximum genotyping error rate when the missing genotype is included in the calculation of the genotyping error rate. On the other hand, Abecasis et al.³² considered error rates ≤ 0.05 to be moderate. Thus, the error rates used in our study are in a reasonable range.

The CS test also depends on the genotype data from the null loci. In our study, the observed genotype at the null loci were also generated from the same simulation model as for the candidate locus, but with different allele frequencies and genotype error rates. Specifically, under both the null and the power simulations, the penetrances used were $f_0 = f_1 = f_2 = 0.10$. The allele frequencies of the null loci in the i th subpopulation were randomly generated from a uniform random variable, $U(p_i - \nu, p_i + \nu)$, where p_1 and p_2 values were given above and values of ν were taken to be 0.00, 0.03, 0.05, 0.07, and 0.09. $\nu = 0.00$ corresponds to the scenario that the simulated candidate and null loci were perfectly matched. On the other hand, large ν values indicate that the candidate and null loci were poorly matched. We remark that the usual method for generating loci has been based on the beta-binomial distribution.^{9,27} The allele frequency was generated from beta distribution $\text{beta}[(1 - F_{st})p/F_{st}, (1 - F_{st})(1 - p)/F_{st}]$, where p is the minor allele frequency. If $p = p_1$ or p_2 and $F_{st} = 0.05$, then allele frequencies generated from $U(p_i - \nu, p_i + \nu), \nu \leq 0.05$, are between 35 and 70 percentile points of the beta distribution. Therefore, beta-binomial and uniform distribution-generating mechanisms essentially give similar results in the study. The error models for generating misclassified genotype data at the null loci were also identical to that for the candidate locus. However, the genotyping error rates at the null loci were randomly selected from a uniform random variable between $\max(\varepsilon - 0.02, 0.0)$ and $\min(\varepsilon + 0.02, 0.05)$, where ε is given above.

Under the given simulation conditions, we generated case and control genotype data biased by PS and/or GE. The numbers of cases and controls were both equal to 100 for the null and power simulations. The effect of PS and/or GE was corrected by use of L (equal to 50, 60, 70, 80, 90, or 100) null loci in the simulations. Estimates of type I error rates and powers were based on 2,000 replications. For a particular null or power simulation, each estimate is the proportion of the replicates for which the test statistic exceeds $\chi^2_2(0.05)$.

Results

Empirical Type I Error Rates

Results for the simulated type I error rates are presented in figures 1 and 2 under the symmetric allele-dropout error model and the allele-based error model, respectively. The results for the CS test are based on the use of $L = 50$ null loci. Later, we show that using other numbers of null loci produces similar conclusions. The bias level caused by PS and/or GE can be measured by the difference of the simulated type I error rate of the CS* test and 0.05. Note that $q = 0.5$ corresponds to the case of no PS and that $\varepsilon =$

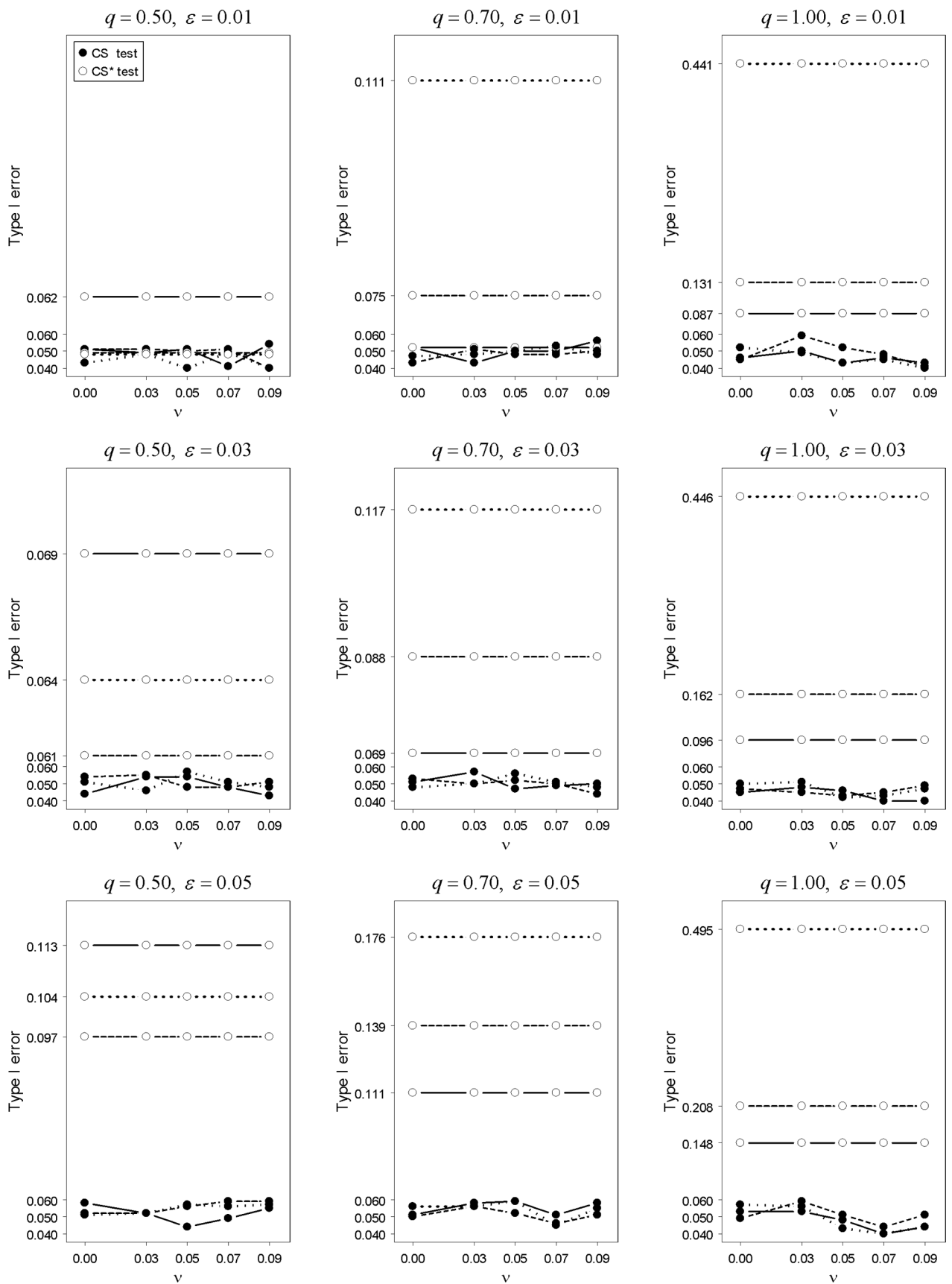


Figure 1. Curves of the empirical type I errors under the symmetric allele-dropout error model. The solid lines are for the cases where the second subpopulation has allele frequency $p_2 = 0.33$ at the candidate locus, dashed lines for the cases with $p_2 = 0.35$, and the dotted lines for the cases with $p_2 = 0.4$

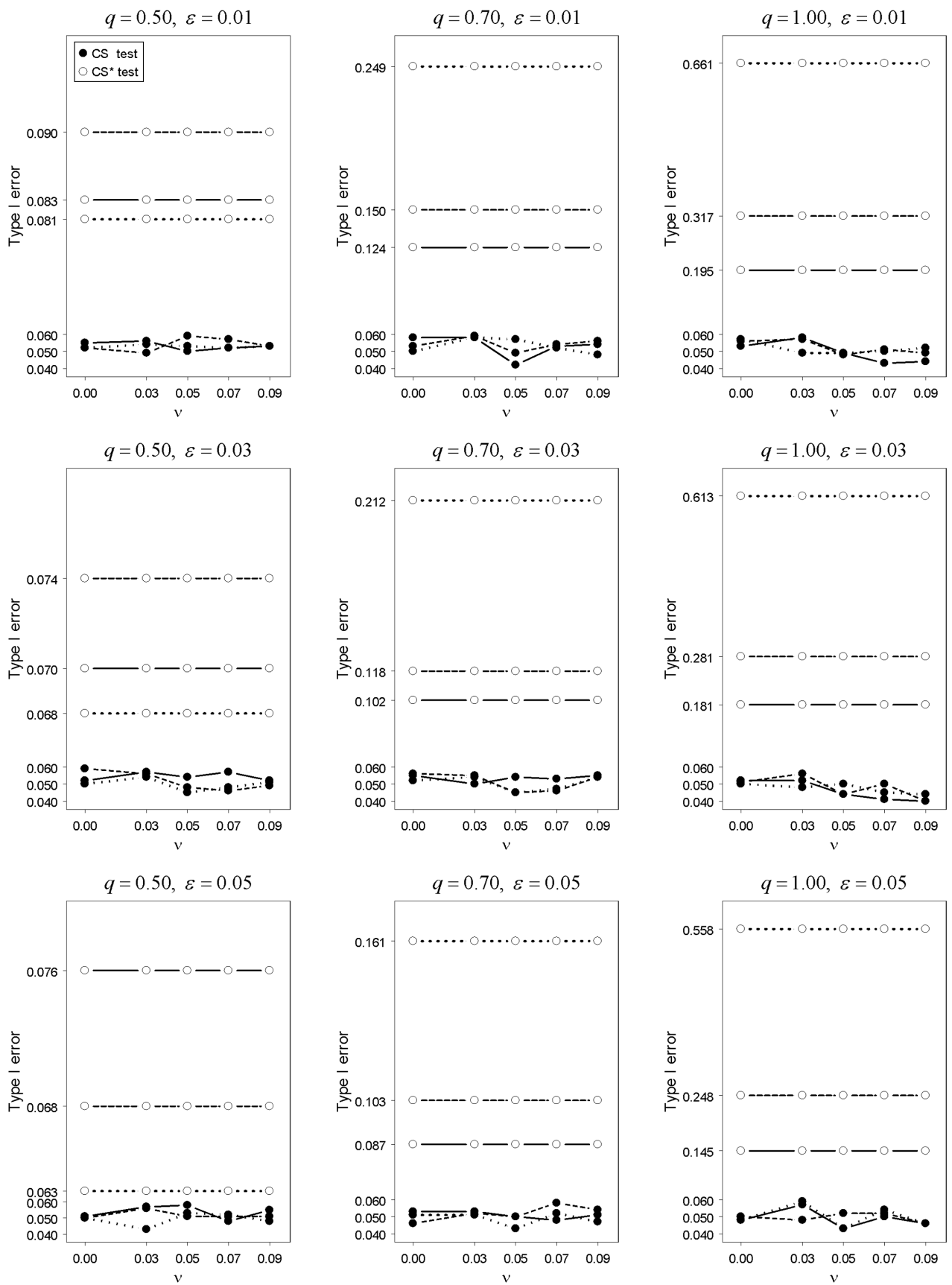


Figure 2. Curves of the empirical type I errors under the allele-based error model. The solid lines are for the cases where the second subpopulation has allele frequency $p_2 = 0.33$ at the candidate locus, dashed lines for the cases with $p_2 = 0.35$, and the dotted lines for the cases with $p_2 = 0.4$.

0.01 corresponds to the case of no GE, if the underlying error model is the symmetric allele-dropout error model. Under the former condition, we have $\beta_1^* = \beta_2^* = 0$, and, if the latter condition holds, we have $\gamma_1 = \gamma_2 = 0$. Therefore, under the symmetric allele-dropout error model with $q = 0.5$ and $\varepsilon = 0.01$, the CS* test should approximately achieve the expected type I error rate in the simulations. According to our results in the upper left panel of figure 1, the corresponding empirical type I error rates of the CS* test range from 0.048 to 0.062. This shows that our simulation study has very reasonable quality. In general, the CS* test tends to have elevated type I errors when PS and/or GE exists. For example, in the case of PS but no GE (see fig. 1 under the cases of $\varepsilon = 0.01$), the largest type I error rate is 0.441, which occurs in the case of $p_2 = 0.40$. On the other hand, in the case of GE but no PS (see figs. 1 and 2 under the cases of $q = 0.5$), the largest empirical type I error rate of the CS* test is 0.113, which occurs in the case of $\varepsilon = 0.05$ under the symmetric allele-dropout error model. However, under the allele-based error model, the largest type I error rate is only 0.076, showing mild inflation in the false-positive rate. Finally, if both PS and GE exist in the association study, the largest empirical type I error rate of the CS* test was increased to 0.661, which occurs in the case of $q = 1.0$, $p_2 = 0.40$, and $\varepsilon = 0.01$ under the allele-based error model (upper right panel of fig. 2). From these results, it is also seen that the existence of PS causes more severe bias in an association study than does the existence of GE. The level of PS increases as the difference of the sampling proportions (for subpopulations) in cases and controls or the difference of the allele frequencies in subpopulations increases. Similarly, the bias level of GE increases as the difference of the error rates (measured by $\varepsilon - 0.01$ in the symmetric allele-dropout error model and by $0.05 - \varepsilon$ in the allele-based error model) increases.

Next, inspecting the curves of the type I error rates for the proposed CS test in figures 1 and 2, we find that the performance of the CS test is very stable and that the type I errors are very close to the expected value (0.05) under all simulation conditions. For example, in the case of PS but no GE (fig. 1), the type I error rates of the CS test range from 0.40 to 0.56; in the case of GE but no PS, they range from 0.43 to 0.59 (figs. 1 and 2). If PS and GE exist simultaneously, the corresponding range is 0.40–0.59. These results are still very satisfactory. We note that the CS test shows very reasonable performance for type I error rates, even when the maximum deviation of the allele frequencies among the candidate and selected null loci is as large as 0.09. This suggests that, when the CS test is applied, the allele frequencies of the candidate and null loci are not required to be matched. Our optimal choice of window size b_n automatically excludes unnecessary null loci from analysis.

Empirical Powers

The powers of the two tests depend on the genetic model and the level of PS and/or GE. In general, the CS* test tends to have smaller powers under a larger level of PS. For example, in the case of no GE (under the symmetric allele-dropout error model) the smallest power of the CS* test under a recessive genetic model (fig. 3) is 0.816 if there is no PS, but it becomes 0.728 under mild PS ($q = 0.70$) and 0.498 under more-severe PS ($q = 1.0$). Using the same genetic and error models, we find that the smallest power is only 0.204 when there is a joint effect of PS and GE. If one uses the same symmetric allele-dropout error model, but the genetic models are additive or dominant, the smallest powers of the CS* test are equal to 0.303 and 0.668, respectively (see figs. 4 and 5 under case $q = 1.00$ and $\varepsilon = 0.05$).

Under the allele-based error model, the power performance of the CS* test is similar. First, the power also tends to decrease as the level of PS increases. However, the smallest power of the CS* test is 0.469 under the recessive genetic model, 0.091 under the additive model, and 0.377 under the dominant model (figs. 6–8). In contrast, under the same simulation conditions but with no PS ($q = 0.50$), the corresponding smallest power increases to 0.689, 0.631, and 0.742 under recessive, additive, and dominant models, respectively. It is of interest that the percentage decrease in power (from no PS, $q = 0.50$, to more-severe PS, $q = 1.00$) of the CS* test under the additive model is ~86%. However, under the same level of PS, the largest percentage change in power of the CS* test is only ~6% for the recessive model, 44% for the additive model, and 14% for the dominant model, because of the different genotyping error rates. This shows that, under the allele-based error model (and under the symmetric allele-dropout error model; see figs. 3–5), PS has a more serious effect on the power performance of CS* than does GE.

Regarding the performance of the CS test, it is important to note that, if there exist PS and/or GE, the new test tends to have much larger power than the unadjusted CS* test. For example, the largest power difference of the two tests is 0.69, and the relative increase in power is >525%. This occurs in the case of an additive genetic model under $q = 1.00$ and $\varepsilon = 0.05$ (fig. 7). It is also of interest that the power performance of the new test is very robust against the underlying genetic model. Its powers are approximately independent of the PS level, genotyping error rate, and frequency difference between different loci. For example, except in the case of a dominant genetic model and the symmetric allele-dropout error model, the range of the powers of the CS test is only 0.70–0.88. In the former case, the powers are in the range 0.814–0.930. The smallest power (0.70) of the CS test occurs in the case of a recessive genetic model under more-severe PS and the largest error rate (see the case of $q = 1.00$ and $\varepsilon = 0.05$ in fig. 3). In comparison, under the same conditions, the

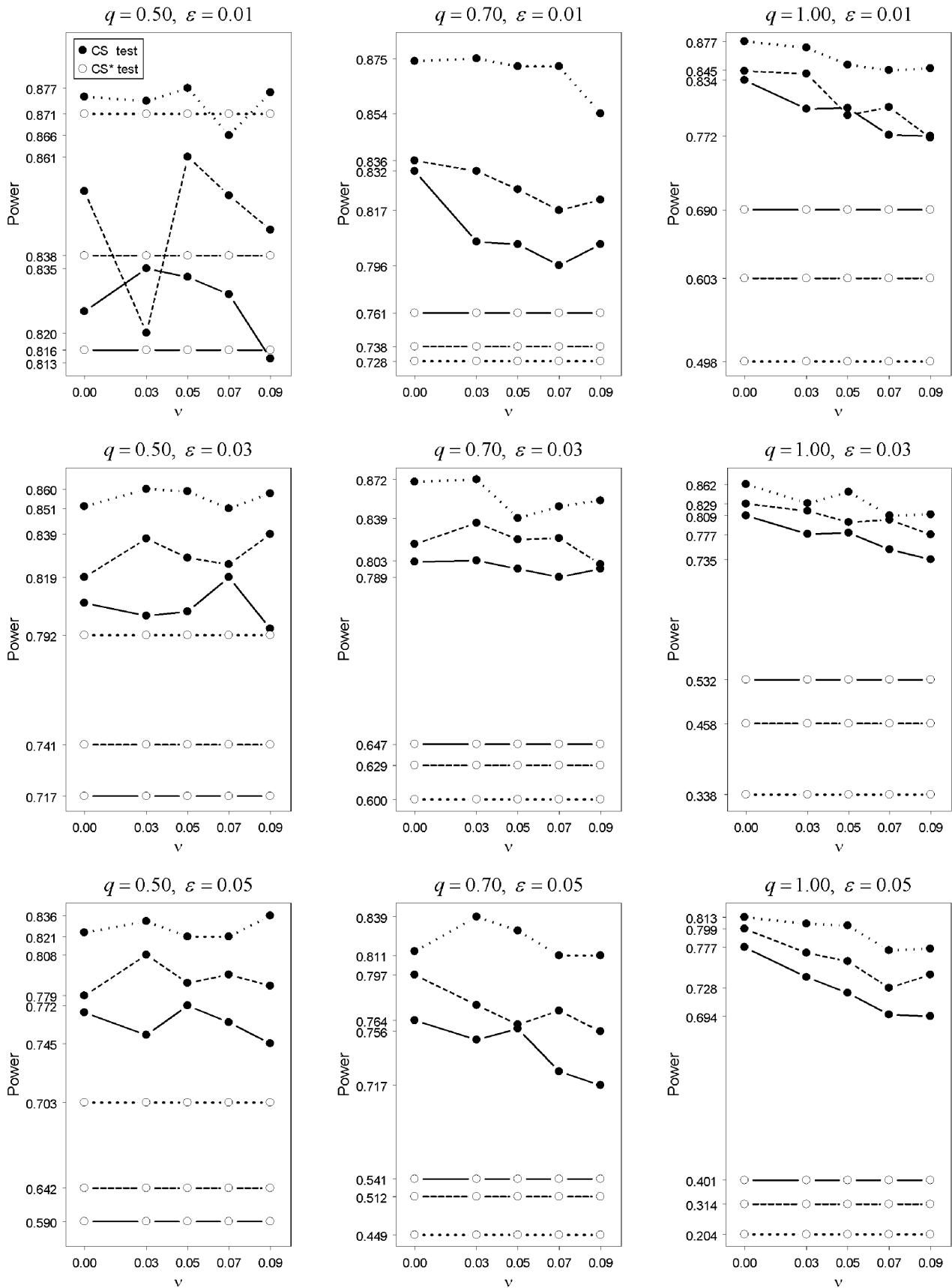


Figure 3. Curves of the powers under the recessive genetic model and symmetric allele-dropout error model. The solid lines are for the cases where the second subpopulation has allele frequency $p_2 = 0.33$, dashed lines for the cases with $p_2 = 0.35$, and the dotted lines for the cases with $p_2 = 0.4$.

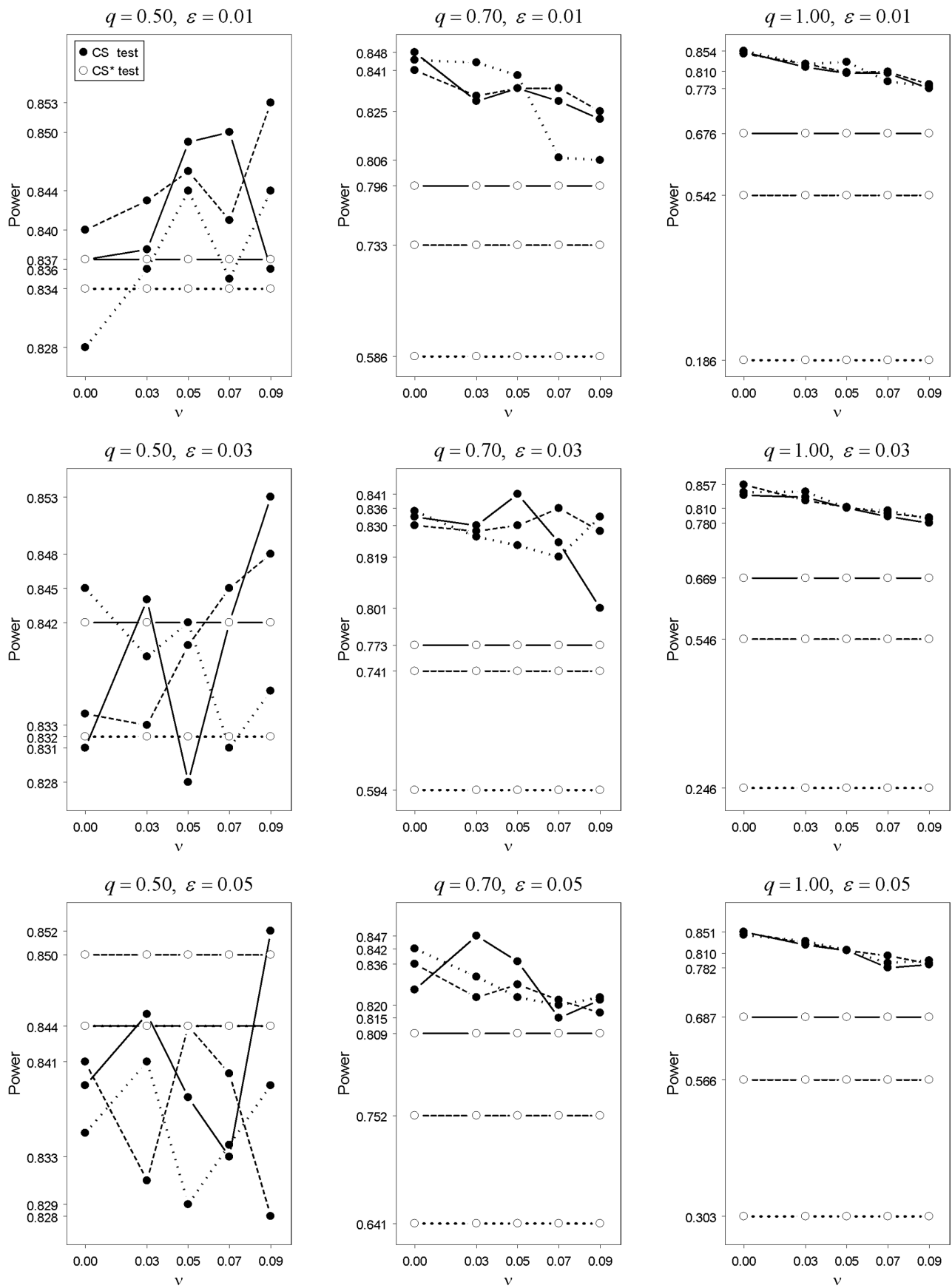


Figure 4. Curves of the powers under the additive genetic model and symmetric allele-dropout error model. The solid lines are for the cases where the second subpopulation has allele frequency $p_2 = 0.33$, the dashed lines for the cases with $p_2 = 0.35$, and the dotted lines for the cases with $p_2 = 0.4$.

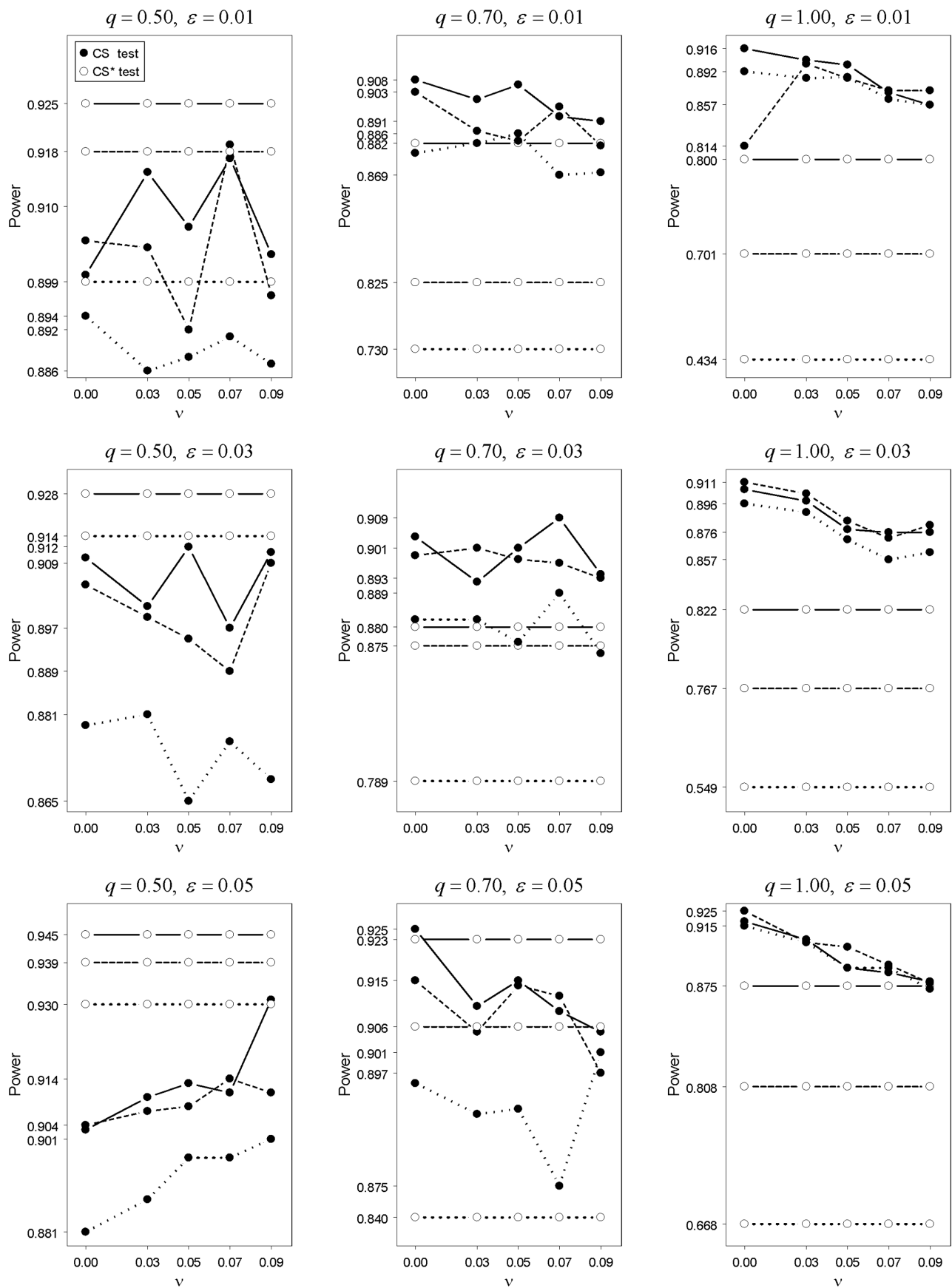


Figure 5. Curves of the powers under the dominant genetic model and symmetric allele-dropout error model. The solid lines are for the cases where the second subpopulation has allele frequency $p_2 = 0.33$, the dashed lines for the cases with $p_2 = 0.35$, and the dotted lines for the cases with $p_2 = 0.4$.

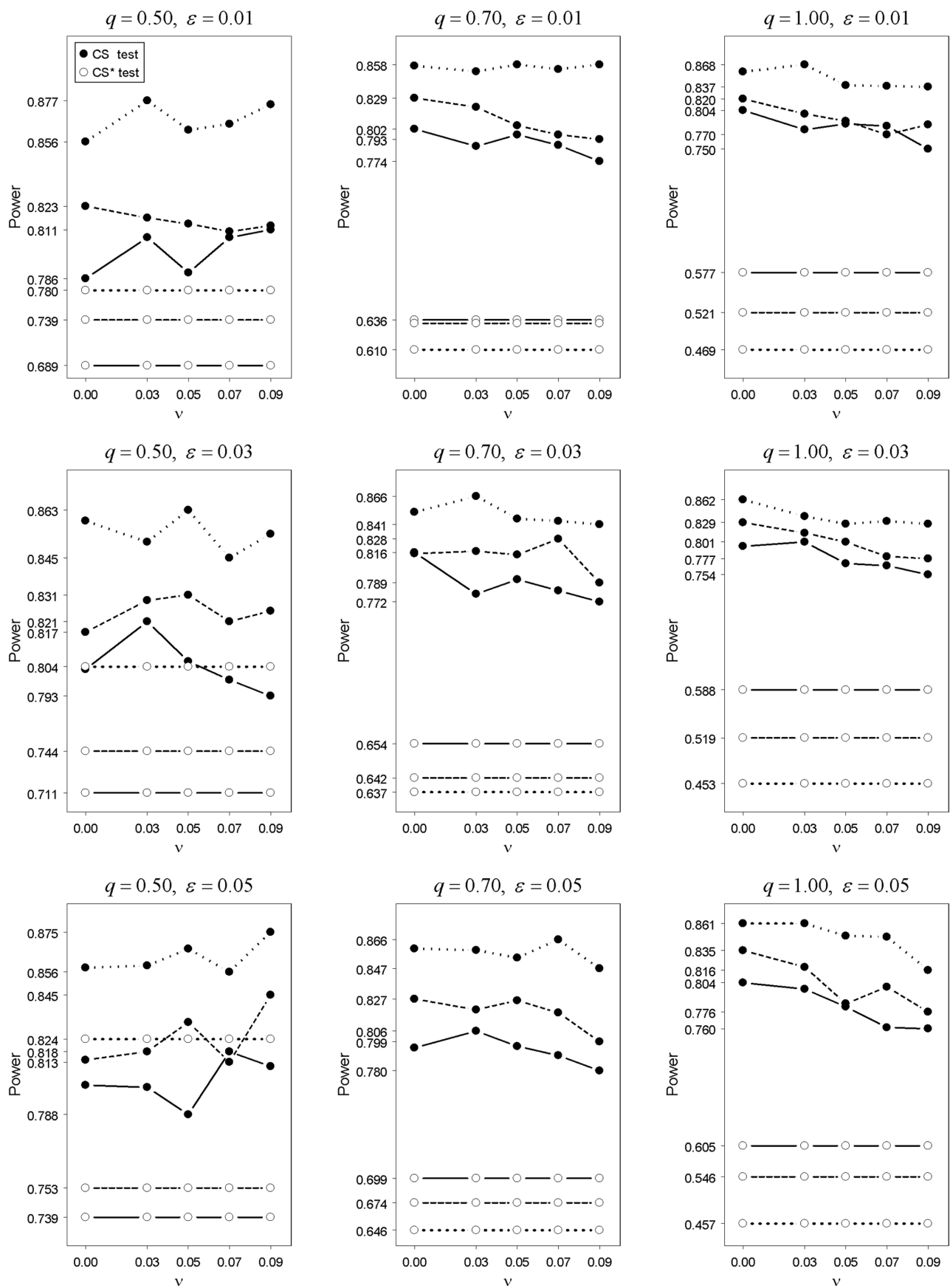


Figure 6. Curves of the powers under the recessive genetic model and allele-based error model. The solid lines are for the cases where the second subpopulation has allele frequency $p_2 = 0.33$, the dashed lines for the cases with $p_2 = 0.35$, and the dotted lines for the cases with $p_2 = 0.4$.

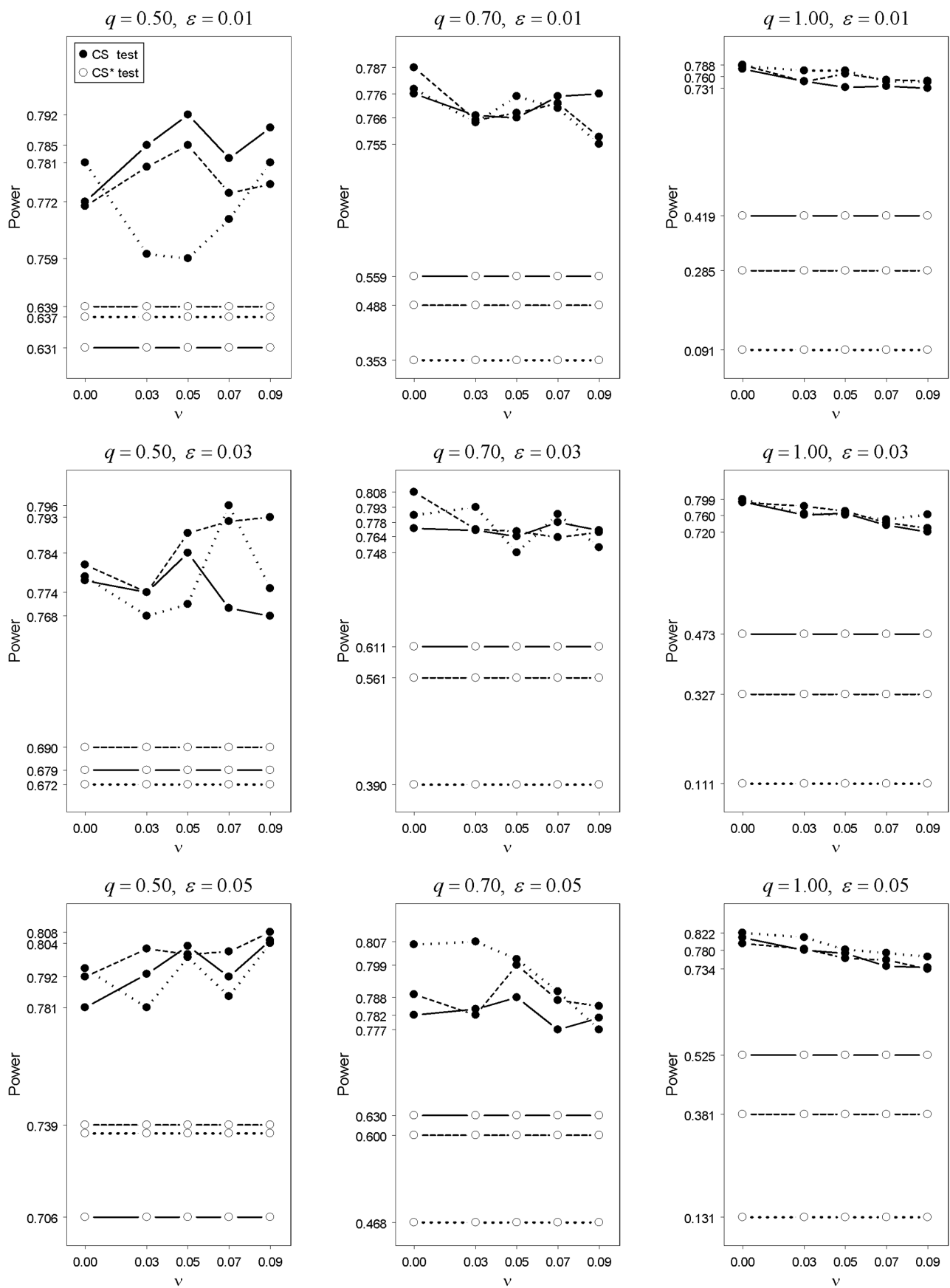


Figure 7. Curves of the powers under the additive genetic model and allele-based error model. The solid lines are for the cases where the second subpopulation has allele frequency $p_2 = 0.33$, the dashed lines for the cases with $p_2 = 0.35$, and the dotted lines for the cases with $p_2 = 0.4$.

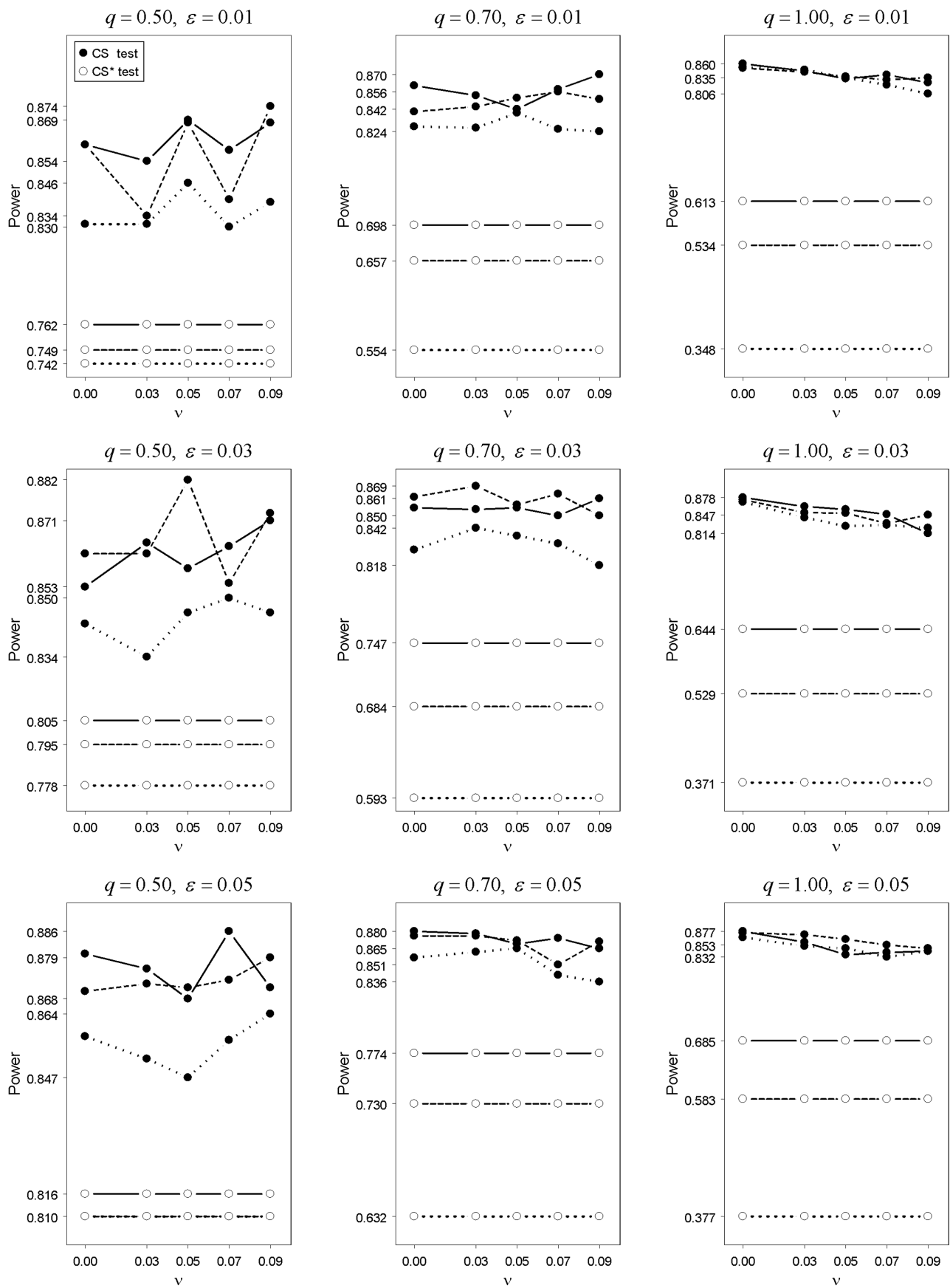


Figure 8. Curves of the powers under the dominant genetic model and allele-based error model. The solid lines are for the cases where the second subpopulation has allele frequency $p_2 = 0.33$, the dashed lines for the cases with $p_2 = 0.35$, and the dotted lines for the cases with $p_2 = 0.4$.

power of the CS* test is only 0.40. Under no PS and no GE ($q = 0.50$ and $\varepsilon = 0.01$ in fig. 3) and if the allele frequencies of the candidate and null loci are perfectly matched, the new test has a minimum power of 0.825, which occurs in the case of a recessive genetic model. In contrast, the corresponding power of the CS* test is 0.816. That is, even when no systematic bias exists, the new test is still slightly better than the regular test.

Null Loci

Our previous reports about the empirical type I errors and powers were based on the use of 50 null loci for computing the CS statistic. In table 1, we report empirical type I error rates of the CS test on the basis of the use of different numbers of null loci ($L = 50, 60, \dots, 100$). Recall that, in principle, if the allele frequencies and genotyping error rates of the candidate and null loci are approximately identical in each subpopulation, then one needs only a few null loci to correct for the bias caused by PS and GE. However, if the allele frequencies or error rates differ too much among the candidate and null loci, then the use of too many unnecessary null loci in the analysis might lead to poor performance of the corrected association test. In this article, we suggest using an optimal window size b_n to determine useful null loci for analysis. Thus, some genotype data from the null loci might be excluded from estimations of the bias. The results from table 1 show that the CS test with the use of the optimal window size has the desired performance, since the CS test shows very stable type I error rates under different numbers of null loci. Inspecting table 1, we find that the largest difference in the type I error rates between the CS test with 50 and 100 null loci is <1% in the case of the symmetric allele-dropout error model and 1.5% in the case of the allele-based error model. In fact, all empirical type I error rates presented in table 1 are in the range 0.040–0.058. This conclusion shows that 50 null loci are sufficient for correcting the bias caused by PS and/or GE if the CS test is applied in the association study.

Discussion

A recent article by Clayton et al.¹⁶ showed, in an analysis of a case-control study of type I diabetes in Great Britain, that population structure explained part of the significant 11.2% inflation of test statistics, and differential bias in genotyping scoring between case and control DNA samples explained the remainder of the inflation. It is well known that the regular χ^2 test in a case-control study is sensitive to PS and GE. In contrast, the usual TDT is not sensitive to PS, but its false-positive rate may be elevated because of GE.^{13,26} In this article, we have proposed a novel method to correct simultaneously for the biases caused by PS and GE in case-control studies. The bias can be estimated using a weighted average of the log of the sample odds ratios, which are computed from 2×3 tables of the

Table 1. Empirical Type I Errors of the CS Test with Different Numbers of Null Loci

| Error Model, p_2^a , q , ^b and v^c | Type I Error for $L^d =$ | | | | | |
|--|--------------------------|------|------|------|------|------|
| | 50 | 60 | 70 | 80 | 90 | 100 |
| Symmetric allele-dropout: | | | | | | |
| .33: | | | | | | |
| .70: | | | | | | |
| .03 | .057 | .047 | .043 | .048 | .043 | .057 |
| .05 | .047 | .046 | .043 | .047 | .048 | .049 |
| 1.00: | | | | | | |
| .03 | .048 | .053 | .045 | .041 | .044 | .045 |
| .05 | .046 | .041 | .042 | .042 | .041 | .041 |
| .35: | | | | | | |
| .70: | | | | | | |
| .03 | .050 | .047 | .051 | .048 | .054 | .045 |
| .05 | .052 | .040 | .051 | .043 | .041 | .052 |
| 1.00: | | | | | | |
| .03 | .047 | .049 | .053 | .042 | .041 | .046 |
| .05 | .045 | .052 | .057 | .044 | .042 | .043 |
| .40: | | | | | | |
| .70: | | | | | | |
| .03 | .050 | .050 | .052 | .051 | .043 | .042 |
| .05 | .056 | .051 | .058 | .052 | .051 | .051 |
| 1.00: | | | | | | |
| .03 | .510 | .044 | .043 | .043 | .041 | .042 |
| .05 | .420 | .047 | .051 | .048 | .048 | .051 |
| Allele-based: | | | | | | |
| .33: | | | | | | |
| .70: | | | | | | |
| .03 | .050 | .055 | .050 | .052 | .055 | .059 |
| .05 | .054 | .047 | .054 | .056 | .053 | .040 |
| 1.00: | | | | | | |
| .03 | .058 | .052 | .053 | .052 | .050 | .043 |
| .05 | .049 | .041 | .056 | .042 | .042 | .045 |
| .35: | | | | | | |
| .70: | | | | | | |
| .03 | .055 | .052 | .052 | .059 | .049 | .045 |
| .05 | .045 | .049 | .056 | .047 | .048 | .050 |
| 1.00: | | | | | | |
| .03 | .056 | .051 | .046 | .053 | .050 | .040 |
| .05 | .044 | .044 | .044 | .041 | .043 | .044 |
| .40: | | | | | | |
| .70: | | | | | | |
| .03 | .054 | .053 | .053 | .049 | .056 | .047 |
| .05 | .045 | .046 | .046 | .056 | .043 | .053 |
| 1.00: | | | | | | |
| .03 | .048 | .056 | .048 | .043 | .046 | .041 |
| .05 | .050 | .040 | .045 | .043 | .045 | .041 |

^a Allele frequency of the second subpopulation.

^b Proportion of the diseased (nondiseased) individuals sampled from the first (second) subpopulation.

^c Maximum deviation of the allele frequencies among the candidate and null loci in each subpopulation.

^d L = number of null loci used for computing the CS test.

null loci. By use of this estimate, the CS test is defined as a likelihood-ratio test. If the null hypothesis of no association is rejected, the effect of the genetic factor can also be estimated, through application of equation (A2) in ap-

pendix A. The computation of our test statistic is simple, and the availability of an enormous number of null loci can provide many opportunities to apply our method.

Khlat et al.³³ argued that, under realistic scenarios—in which subpopulations account for ≤ 0.10 of the study population and allelic frequency differences are ≤ 0.20 —the inflation of the type I error is of limited concern. However, we show that, under general sampling of cases and controls, genotype frequencies and sampling proportions $P^*(S = s|D = 1)$ and $P^*(S = s|D = 0)$ of the subpopulation determine the level of PS (appendix A). In the case of simple random sampling, one can show that the level of PS also depends on the disease risks of the subpopulations. Our result implies that, if one can identify genetically distinct subpopulations and select identical sampling proportions in cases and controls ($P^*(S = s|D = 1) = P^*(S = s|D = 0)$), then the false-positive rate of the regular χ^2 test will not be inflated. Otherwise, even under the scenario considered by Khlat et al.,³³ the effect of PS might be severe because of different sampling proportions being used in the cases and controls.

In the presence of PS, one popular approach to preserving the nominal type I error rate is to apply the GC method. This method attempts to adjust the variance of the Cochran-Armitage (CA) trend test by calculating the statistic with data from the null loci. However, many published results indicated that, in some situations (e.g., when the PS level is large), this approach may not be satisfactory.¹⁹ One important reason is that, in the presence of PS, the regular χ^2 statistic has a noncentral χ^2 distribution, and dividing the noncentral χ^2 by a constant does not always produce a central χ^2 . The GC method is also sensitive to GE, since the test statistic in GC depends on the CS* statistic.

Epstein et al.²¹ used a stratification-score approach for controlling the stratification. Their stratification scores depend on the use of generalized least squares and data from substructure-informative loci. On the other hand, Gorroochurn et al.¹⁹ proposed an approach, called the “ δ -centralization” (DC) method, to correct for PS by using data from null loci, similar to our method. They suggested estimating the square root of the noncentrality parameter directly and the adjusted statistic to produce a central χ^2 . The success of the DC method depends crucially on whether the noncentrality parameter can be estimated accurately. If the candidate and null loci are well matched, in the sense that they have similar genotype frequencies in subpopulations, then this approach can successfully eliminate the effect of PS. However, on the basis of the available data, it is not easy to verify this condition. The DC method suggests choosing null loci so that their genotype frequencies are within a window of size 0.10 to that at the candidate locus. However, under the same simulation conditions considered in this article, our unreported results show that sometimes the DC test is conservative, although its general performance is better than that of the GC method. One drawback of the DC method is that it depends on the CA trend test. However, the CA

trend test is not robust against misspecification of the genetic model. For example, the CA trend test, which is optimal under the dominant genetic model, may perform poorly under the recessive genetic model. Our unreported simulation results show that, under the dominant and additive genetic models, the smallest empirical powers of the DC test, which is efficient in power under the dominant genetic model, are 0.935 and 0.841, respectively, in the case of no GE (under the symmetric allele-dropout error model with $\varepsilon = 0.01$). However, under the recessive genetic model, the corresponding powers range from 0.121 to 0.255. This shows that the DC test has poor performance in power when the underlying genetic model is misspecified. In general, the DC-type tests based on δ -centralizing any CA trend test have similar drawbacks. The DC-type tests are not robust against GE.

Theoretical results indicate that the performance of the CS test is robust to the error model. In our simulations, we have considered two error models for alleles. In fact, a random-genotype error model³⁴ was also investigated in the simulations but is not reported. Under this model, each genotype was randomly replaced with another genotype, in a manner proportional to genotype frequencies. We assumed that each genotype has a constant probability, ε , of being misclassified, and we selected $\varepsilon = 0.01$ for the case sample and $\varepsilon = 0.01, 0.03,$ and 0.05 for the control sample. The rest of the parameter values were defined as in the “Simulations” section. According to the simulated results, the type I error rates of the CS test are also in the range 0.04–0.059, the same as that reported in the “Empirical Type I Error Rates” section. Under the dominant genetic model, the range of the powers of the CS test is 0.873–0.929. If the genetic model is recessive or additive, the range becomes 0.791–0.909. Note that the corresponding ranges reported in the “Empirical Powers” section are 0.814–0.930 and 0.70–0.88. This shows that the CS test is indeed not sensitive to the choice of error model.

The CS test can be applied to admixture populations. However, unlike for the method suggested by Pritchard et al.,¹⁰ one does not need to infer details of population structure and to estimate the ancestry of sampled individuals before applying the CS test. The test also holds under the general risk model of Epstein et al.²¹ It is of interest that, if necessary, the CS test can be modified further to incorporate stratification variables, such as ethnicity. Stratified analysis often can reduce the level of PS and makes the bias caused by PS smaller and more uniform among the candidate and null loci. Under this scenario, the CS test should be more efficient. Stratified analysis can be done by first classifying the sample into more-homogeneous groups and by then applying the suggested method separately for each group. The final test statistic is a combination of the CS statistics for different groups. For example, suppose the case-control sample is stratified into R strata. Let X_r^2 be the CS test statistic for the r th stratum. These statistics are independent; hence, under the null hypothesis, $X^2 = \sum_{r=1}^R X_r^2$ is asymptotically distributed as a central χ^2 with $2R$

df. The modified CS test suggests using X^2 to test the null hypothesis of no association. Our initial simulation results (not reported here) show that the CS test with stratification sometimes outperforms the CS test without stratification, but the difference in their powers is not very significant.

Acknowledgments

This research was supported in part by the National Science Council, Taiwan, under contract NSC95-2118-M-039-002-MY2. We thank editors and reviewers for their constructive comments, which improved the presentation of this article.

Appendix A

Using model (1) and Bayes theorem, the ratio of the case and control genotype frequencies can be written as

$$\log \left[\frac{\Pr(G = g | D = 1, S = s)}{\Pr(G = g | D = 0, S = s)} \right] = \alpha + \alpha_s + \beta_g, \tag{A1}$$

where

$$\alpha_s = \mu_s + \log \left[\frac{\Pr(D = 0 | S = s) \Pr(D = 1 | S = 1)}{\Pr(D = 1 | S = s) \Pr(D = 0 | S = 1)} \right] \text{ and } \alpha = \mu + \log \left[\frac{\Pr(D = 0 | S = 1)}{\Pr(D = 1 | S = 1)} \right].$$

Next, with application of equation (A1), the case genotype frequency can be written as

$$\Pr(G = g | D = 1) = \exp(\beta_g) \sum_{s=1}^K [\Pr(G = g | D = 0, S = s) P^*(S = s | D = 1) \exp(\alpha + \alpha_s)]$$

and the control genotype frequency as

$$\Pr(G = g | D = 0) = \sum_{s=1}^K [\Pr(G = g | D = 0, S = s) P^*(S = s | D = 0)].$$

The ratio of these two frequencies leads to model

$$\frac{\Pr(G = g | D = 1)}{\Pr(G = g | D = 0)} = \exp(\beta_g) H(g) = \exp(\alpha^* + \beta_g^* + \beta_g), \tag{A2}$$

with $\alpha^* = \log H(0)$, $\beta_g^* = \log [H(g)/H(0)]$, and

$$H(g) = \frac{\sum_{s=1}^K [\Pr(G = g | D = 0, S = s) \exp(\alpha + \alpha_s) P^*(S = s | D = 1)]}{\sum_{s=1}^K [\Pr(G = g | D = 0, S = s) P^*(S = s | D = 0)]}.$$

If there is no true association, then $\beta_1 = \beta_2 = 0$ and $\exp(\alpha + \alpha_s) = 1$; hence, $H(g)$ can be written as

$$H(g) = \frac{\sum_{s=1}^K [\Pr(G = g | S = s) P^*(S = s | D = 1)]}{\sum_{s=1}^K [\Pr(G = g | S = s) P^*(S = s | D = 0)]}. \tag{A3}$$

If observations in the case and control samples were collected under simple random sampling, then $H(g)$ can be further simplified as

$$H(g) = \frac{\Pr(D = 0) \sum_{s=1}^K [\Pr(G = g | S = s) \Pr(D = 1 | S = s) \Pr(S = s)]}{\Pr(D = 1) \sum_{s=1}^K [\Pr(G = g | S = s) \Pr(D = 0 | S = s) \Pr(S = s)]}. \tag{A4}$$

Appendix B

We assume that the error rates are $\Pr(G_o = g_o | G = g, D = 1) = \phi_1(g_o; g)$ in the case sample and $\Pr(G_o = g_o | G = g, D = 0) = \phi_0(g_o; g)$ in the control sample. Thus, if one defines $W_1(g_o; g) = \phi_1(g_o; g) \Pr(G = g | D = 1)$ and $W_0(g_o; g) = \phi_0(g_o; g) \Pr(G = g | D = 0)$, then, under no true association, the ratio of the case and control genotype frequencies can be expressed as

$$\frac{\Pr(G_o = g_o | D = 1)}{\Pr(G_o = g_o | D = 0)} = \exp(\delta + \gamma_{g_o}),$$

where

$$\gamma_{g_o} = \log \left[\left(\sum_{g=0}^2 W_1(g_o, g) \times \sum_{g=0}^2 W_0(0, g) \right) / \left(\sum_{g=0}^2 W_1(0, g) \times \sum_{g=0}^2 W_0(g_o, g) \right) \right]$$

and

$$\delta = \log \left[\sum_{g=0}^2 W_1(0, g) / \sum_{g=0}^2 W_0(0, g) \right].$$

If one applies equation (A2) to replace $\Pr(G = g | D = 1)$ by $\exp(\alpha^* + \beta_g^* + \beta_g) \times \Pr(G = g | D = 0)$ in the definition of γ_{g_o} , then one can express

$$\gamma_{g_o} = \beta_{g_o} + \delta_{g_o},$$

where $\delta_{g_o} = \gamma_{g_o}^* + \beta_{g_o}^*$ and $\gamma_{g_o}^*$ is defined as γ_{g_o} , but with $W_1(g_o, g)$ replaced with $\phi_1(g_o, g) \Pr(G = g_o | D = 0)$. Here, β_{g_o} are the true log odds ratios between disease and genetic factor, and $\beta_{g_o}^*$ and $\gamma_{g_o}^*$ are the effects caused by PS and GE, respectively.

Web Resource

The URL for data presented herein is as follows:

Cheng's software, <http://www2.cmu.edu.tw/~biostat/downloads/KFCpro/KFCprogram.htm>

References

- Knowler WC, Williams RC, Pettitt DJ, Steinberg AG (1988) *Gm*^{3,5,13,14} and type 2 diabetes mellitus: an association in American Indians with genetic mixture. *Am J Hum Genet* 43: 520–526
- Lander ES, Schrok NJ (1994) Genetic dissection of complex traits. *Science* 265:2037–2048
- Gordon D, Matisse TC, Heath SC, Ott J (1999) Power loss for multiallelic transmission/disequilibrium test when errors introduced: GAW11 simulated data. *Genet Epidemiol Suppl* 1 17:S587–S592
- Gordon D, Finch SJ, Nothnagel M, Ott J (2002) Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Hum Hered* 54:22–33
- Morris RW, Kaplan NL (2004) Testing for association with a case-parents design in the presence of genotyping errors. *Genet Epidemiol* 26:142–154
- Moskvina V, Graddock N, Hlmans P, Owen MJ, O'Donovan MC (2006) Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum Hered* 61:55–64
- Morton NE, Collins A (1998) Tests and estimates of allelic association in complex inheritance. *Proc Natl Acad Sci USA* 95:11389–11393
- Risch N, Teng J (1998) The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human disease. I. DNA pooling. *Genome Res* 8: 1273–1288
- Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55:997–1004
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. *Am J Hum Genet* 67:170–181
- Bacanu SA, Devlin B, Roeder K (2000) The power of genomic control. *Am J Hum Genet* 66:1933–1944
- Campbell CD, Ogburn EL, Lunetta KL, Lyon HN, Freedman ML, Groop LC, Altshuler D, Ardlie KG, Hirschhorn JN (2005) Demonstrating stratification in a European American population. *Nat Genet* 37:868–872
- Gordon D, Heath SC, Liu X, Ott J (2001) A transmission/disequilibrium test that allows for genotyping errors in the analysis of single-nucleotide polymorphism data. *Am J Hum Genet* 69:371–380
- Bernardinelli L, Berzuini C, Seaman S, Holmans P (2004) Bayesian trio models for association in the presence of genotyping errors. *Genet Epidemiol* 26:70–80
- Gordon D, Haynes C, Johnnidis C, Patel SB, Bowcock AM, Ott J (2004) A transmission disequilibrium test for general pedigrees that is robust to the presence of random genotyping errors and any number of untyped parents. *Eur J Hum Genet* 12:752–761
- Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier

- LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, et al (2005) Population structure, differential bias and genomic control in a large scale, case-control association study. *Nat Genet* 37: 1243–1246
17. Chen HS, Zhu X, Zhao H, Zhang S (2003) Qualitative semi-parametric test for genetic associations in case-control designs under structured populations. *Ann Hum Genet* 67:250–264
 18. Shmulewitz D, Zhang J, Greenberg DA (2004) Case-control association studies in mixed populations: correcting using genomic control. *Hum Hered* 58:145–153
 19. Gorroochurn D, Heiman GA, Hodge SE, Greenberg DA (2006) Centralizing the noncentral chi-square: a new method to correct for population stratification in genetic case-control association studies. *Genet Epidemiol* 30:277–289
 20. Wright S (1951) The genetic structure of populations. *Ann Eugen* 15:323–354
 21. Epstein MP, Allen AS, Satten GA (2007) A simple and improved correction for population stratification in case-control studies. *Am J Hum Genet* 80:921–930
 22. Mote VL, Anderson RL (1965) An investigation of the effect of misclassification on the properties of chi-square-tests in the analysis of categorical data. *Biometrika* 52:95–109
 23. Satten GA, Flanders WD, Yang Q (2001) Accounting for unmeasured population substructure in case-control studies of genetic association using novel latent-class model. *Am J Hum Genet* 68:466–477
 24. Prentice RL, Pyke R (1979) Logistic disease incidence models and case-control studies. *Biometrika* 66:403–441
 25. Simonoff J (1996) Smoothing methods in statistics. Springer Verlag, New York
 26. Mitchell AA, Cutler DJ, Chakravarti A (2003) Undetected genotyping errors cause apparent overtransmission of common alleles in the transmission/disequilibrium test. *Am J Hum Genet* 72:598–610
 27. Zheng G, Freidlin B, Gastwirth JL (2006) Robust genomic control for association studies. *Am J Hum Genet* 78:350–356
 28. Garte S (1998) The role of ethnicity in cancer susceptibility gene polymorphisms: the example of *CYP1A1*. *Carcinogenesis* 19:1329–1332
 29. Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, Autrup JL, Baranova H, Bathum L, Benhamou S, Boffetta P, et al (2001) Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol Biomarkers Prev* 10: 1239–1248
 30. Kittles RA, Chen W, Panguluri RK, Ahaghotu C, Jackson A, Adebamowo CA, Griffin R, Williams T, Ukoli F, Adams-Campbell U, et al (2002) CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification? *Hum Genet* 110:553–560
 31. Tintle NL, Ahn K, Mendell NR, Gordon D, Finch SJ (2005) Characteristics of replicated single-nucleotide polymorphism genotypes from COGA: Affymetrix and Center for Inherited Disease Research. *BMC Genet Suppl* 1 6:S154
 32. Abecasis GR, Cherny SS, Cardon LR (2001) The impact of genotyping error on family-based analysis of quantitative traits. *Eur J Hum Genet* 9:130–134
 33. Khlal M, Cazes MH, Genin E, Guiguet M (2004) Robustness of case-control studies of genetic factors to population stratification: magnitude of bias and type I error. *Cancer Epidemiol Biomarkers Prev* 13:1660–1664
 34. Douglas JA, Skol AD, Boehnke M (2002) Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *Am J Hum Genet* 70: 487–495