Journal of Human Genetics (2013) 58, 657–666
© 2013 The Japan Society of Human Genetics All rights reserved 1434-5161/13
www.nature.com/jhg

ORIGINAL ARTICLE

Correcting for differential genotyping error in genetic association analysis

Min Yuan[1], Hongyan Fang[2] and Han Zhang[3]

Differential genotype error in case–control association studies occurs when cases and controls are genotyped under different conditions. Existence of differential errors can considerably bias the association test, resulting in inflation of type I error and spurious significance. With the availability of high-throughput genotyping technologies such as the SNPchip, null markers that are unlinked with the disease can be used to correct for the bias caused by differential errors. A similar method, known as the genomic control, had been used to correct for population stratification in association studies. In this paper, we show that the same idea can be used to correct for the bias caused by differential errors, under the assumption that the null markers and the candidate marker are subject to the same or similar genotyping error model. The variance inflation is shown to be minor and the bias in the association test is the major source of type I error inflation in the presence of differential errors. Our method centralizes the test statistic by deducting the bias estimated from null markers through a quadratic regression method, which adjusts for the variability of null marker allele frequencies. Simulation results show that the proposed method performs very well in correcting for the type I error inflations.
Journal of Human Genetics (2013) **58,** 657–666; doi:10.1038/jhg.2013.74; published online 18 July 2013

Keywords: Cochran–Armitage's trend test; differential error; genomic control; Pearson's chi-square test

INTRODUCTION

Case–control design is widely used in genetic association studies. In a well-designed case–control study, cases and controls have no differential structures except the obvious difference between the disease status and the genotypes (if linked to disease) under investigation. However, differential structure is a common issue in population-based case–control studies. The most well-known example is the population structure, correction of which has been studied extensively in the literature.[1–9] Differential genotype errors between cases and controls is another possible source of biases for association analysis, especially so in large-scale genetic studies for which cases and controls may be genotyped under different conditions.[10,11]

Genotype error occurs with inaccurate genotyping measurement or laboratory error. Genotype of single-nucleotide polymorphisms were reported to have misclassification rates in the range of 0.1–5%.[12,13] Most studies assumed that the genotype misclassification rates are the same (non-differential) for case and control groups, for which the association tests have correct type I error but may lose power. Different methods have been proposed to increase power in the presence of non-differential genotype error.[14–18]

However, the genotype measurement errors may have systematic variabilities between cases and controls when the two groups are subject to different experiment conditions.[19] Genotype error may depend on factors such as, among others, quality of blood samples, working conditions of genotyping instrument and expertise of laboratory researchers. For example, the DNA samples from cases and controls in a large-scale association study may be collected at different phases of a study or from different centers using different genotyping instruments. Another example is, if the person genotyping the samples were not blinded to the disease status, quality of genotype measurements may have systematic difference between cases and controls and genotype errors may be differential.

The importance of addressing the differential errors in association analysis was reported in Clayton *et al.*[10] In their study, the DNA sample were obtained from the same extraction protocol and of the same quality, but inflation in the single locus type I error was observed, much of which was caused by the systematic differential call rates between samples. Other heterogeneity factors such as population substructure contribute only minor proportion to the type I error inflation. Moskvina *et al.*[11] studied the effects of differential genotyping error rate on the type I error in case–control studies. Their results show that differential genotyping error can substantially inflate the type I error of association tests even at a low misclassification rate. In fact, any factors (observed or unobserved) that are differential for cases and controls lead to bias as well as variance inflation in association tests.

[1]Department of Probability and Statistics, School of Mathematics, University of Science and Technology of China, Anhui, China; [2]Department of Statistics and Finance, University of Science and Technology of China, Anhui, China and [3]Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
Correspondence: Dr M Yuan, Department of Probability and Statistics, School of Mathematics, University of Science and Technology of China, No. 96, JinZhai Road, Hefei, Anhui 230026, China.
E-mail: myuan@ustc.edu.cn
Received 24 January 2013; revised 6 May 2013; accepted 2 June 2013; published online 18 July 2013

Genomic control (GC) method[1] was proposed for correcting variance inflation caused by population substructure. It assumes the cases and controls have differential sub-populations with different cryptic relatedness structures. The GC method corrects for the effect of differential cryptic relatedness by using information from unlinked null (or null) loci. Devlin and Roeder[1] and Devlin *et al.*[5] argued that in the presence of many unknown sub-populations with cryptic relatedness, there is no bias on average and the variance inflation contribute most to the inflation of the type I error. On the other hand, Gorroochurn *et al.*[20] proposed a δ centralization method to correct for the bias of test statistic caused by population stratification.

In this paper, similar to the GC method, we assume that null markers (with similar allele frequencies) that are unlinked to the disease are available and these markers are subject to the same or similar misclassification models as the candidate marker. We use these null markers to estimate quantities related to the null distribution of the test statistic for the candidate marker. This study shows that the variance distortion is minor in the presence of differential errors and the bias caused by differential errors leads to the inflation of the type I error. We centralize the test statistics by using information estimated from the null markers. Note that with random misclassifications, perfect matching of allele frequencies is not feasible. Our strategy is to match allele frequencies as much as possible, and apply the regression-based procedure to correct for the biases. Simulation results show that our method is quite robust to the variabilities of the allele frequencies of the null markers and robust to mixing of a small proportion of linked loci in the selected 'null' markers.

MATERIALS AND METHODS

We focus on a bi-allelic candidate marker with alleles A and a. Let $p = P(A)$ be the population allele frequency. Denote the three genotypes aa, Aa and AA by $G' = 0, 1, 2$, respectively. In a case–control design, denote the disease status by $Y = 1$ for cases and $Y = 0$ for controls, and let the true genotype frequencies be $p'_{ij} = P(G' = j|Y = i)$, $j = 0, 1, 2$; $i = 0, 1$. The null hypothesis of no association is H_0: $p'_{0j} = p'_{1j} = p'_j$, $j = 0, 1, 2$. Let G be the observed genotype. The definitions of the observed genotype counts are given in Table 1.

Owing to the genotyping error, the true genotypes may be misclassified. Denote the probability of misclassifying true genotype k as j by

$$\pi_{kj}^{(i)} = P(G = j \mid G' = k,\ Y = i),\ k, j = 0, 1,\ 2;\ i = 0, 1,$$

where the superscript i denotes the group. Non-differential error means $\pi_{kj}^{(1)} = \pi_{kj}^{(0)}$, or $P(G = j|G' = k, Y = i) = P(G = j|G' = k)$. We assume that the misclassification rates are differential for the case and control groups, that is, $\pi_{kj}^{(1)} \neq \pi_{kj}^{(0)}$ for at least some k, $j = 0, 1, 2$. Then the frequencies of observed genotypes for group $i (i = 0, 1)$ are given as

$$\begin{aligned} p_{ij} &= P(G = j \mid Y = i) = \sum_{k=0}^{2} P(G = j \mid G' = k, Y = i) P(G' = k \mid Y = i) \\ &= \sum_{k=0}^{2} \pi_{kj}^{(i)} p'_{ik}. \end{aligned} \tag{1}$$

Note that when the error rates are non-differential, that is, $\pi_{kj}^{(1)} = \pi_{kj}^{(0)}$ for all k, j, we have $p_{0j} = p_{1j}$, $j = 0, 1, 2$ under the null H_0: $p'_{1j} = p'_{0j}$, $j = 0, 1, 2$. This means that association tests is valid with correct type I error. In this paper, we consider an error model with two parameters $\pi_0^{(i)}$ (the probability of misclassifying a homozygote as the heterozygote in group i and $\pi_1^{(i)}$ (the probability of misclassifying the heterozygote as a homozygote in group i for $i = 0, 1$. This model has been considered by other authors.[16,22] The error rates (or misclassification rates) with this model are presented in Table 2. We define the misclassification rate ratio (MRR) to be $\pi_1^{(1)}/\pi_1^{(0)} = \pi_0^{(1)}/\pi_0^{(0)}$ and use this as a measure of differential degree in genotype errors. Obviously, with the differential errors, the frequencies of observed genotypes for case (p_{1j}) and groups (p_{0j}) may not be identical under the null.

Cochran–Armitage's trend test and Pearson's chi-square test

The Cochran–Armitage's trend (CAT) test is widely used in genetic association analysis.[23,24] The CAT test assigns scores 0, x, 1 to the three genotypes aa, Aa, AA, respectively. Different specifications of the score x for the heterozygote Aa rely on the underlying mode of inheritance. Specifically, scores $x = 0$, 0.5, 1 correspond, respectively, to the recessive, additive and dominant models.

Let $\Delta_x = (p_{12} - p_{02}) + x(p_{11} - p_{01})$, $\sigma_x^2 = (n_1^{-1} + n_0^{-1})[(p_2 + x^2 p_1) - (p_2 + x p_1)^2]$ and $\hat{\Delta}_x = (\hat{p}_{12} - \hat{p}_{02}) + x(\hat{p}_{11} - \hat{p}_{01})$, $\hat{\sigma}_x^2 = (n_1^{-1} + n_0^{-1})[(\hat{p}_2 + x^2 \hat{p}_1) - (\hat{p}_2 + x\hat{p}_1)^2]$ are their estimates, where $p_j = n_1 p_{1j}/n + n_0 p_{0j}/n$, $\hat{p}_j = n_1 \hat{p}_{1j}/n + n_0 \hat{p}_{0j}/n$ and $\hat{p}_{ij} = n_{ij}/n_i$. For a specific choice of the score x for genotype Aa, the CAT indexed by x is given as

$$Z_x = \frac{\hat{\Delta}_x}{\hat{\sigma}_x} = \frac{(\hat{p}_{12} - \hat{p}_{02}) + x(\hat{p}_{11} - \hat{p}_{01})}{\sqrt{(n_1^{-1} + n_0^{-1})[(\hat{p}_2 + x^2 \hat{p}_1) - (\hat{p}_2 + x\hat{p}_1)^2]}}. \tag{2}$$

This test is used for testing null hypothesis H_0: $p'_{1j} = p'_{0j} = p'_j$, $j = 0, 1, 2$ for the two groups with observed data. Let $\mu_x = E_{H_0}(\hat{\Delta}_x)$ and $\tau_x^2 = var_{H_0}(\hat{\Delta}_x)$. If there is no genotype error, then $\mu_x = 0$, $\tau_x^2 = \sigma_x^2$ and Z_x is asymptotically distributed as a standard normal distribution $N(0, 1)$.

If there exists differential error between cases and controls, then $\mu_x \neq 0$ and $\tau_x^2 \neq \sigma_x^2$. We call $\lambda_x = \tau_x^2/\sigma_x^2$ as the variance inflation factor (VIF). Figure 1 illustrates the bias μ_x and variance inflation factor λ_x as functions of allele frequency p or MRR, which measures the degree of differentiation of misclassification rates between cases and controls. For a given MRR, both of the μ_x and the λ_x decrease with allele p, indicating that the CAT is more severely biased when the allele A is rare. On the other hand, both of them are increasing with MRR (bottom panels of Figure 1), indicating the CAT is more biased when cases and controls are more differential in genotype misclassifications. The variance inflation factor is very close to 1 as shown in Figure 1. Extensive simulation and theoretical investigation also confirm this. We will thus focus on correcting the bias term in our study. In Appendix A, we show that

$$\mu_x = \beta_0 + \beta_1 p + \beta_2 p^2.$$

We therefore apply a quadratic regression method to correct for the biases caused by differential error.

Pearson's chi-square test (PCT) is given by $T_p = \sum_{i=0}^{1} \sum_{j=0}^{2} (n_{ij} - e_{ij})^2 / e_{ij}$ where $e_{ij} = n_i m_j / n$, which has several different representations. In our study, we choose to use the following representation,[25,26] which clearly links the PCT

Table 1 Genotype counts for case–control design

	G (genotype)			
Group	*0*	*1*	*2*	*Total*
Control ($Y = 0$)	n_{00}	n_{01}	n_{02}	n_0
Case ($Y = 1$)	n_{10}	n_{11}	n_{12}	n_1
Total	m_0	m_1	m_2	n

Table 2 Conditional probability of observed genotype given the true genotype in group *i* ($i = 0, 1$)

	Observed genotype (G)		
True genotype (G')	*0*	*1*	*2*
0	$1 - \pi_0^{(i)}$	$\pi_0^{(i)}/2$	$\pi_0^{(i)}/2$
1	$\pi_1^{(i)}/2$	$1 - \pi_1^{(i)}$	$\pi_1^{(i)}/2$
2	$\pi_0^{(i)}/2$	$\pi_0^{(i)}/2$	$1 - \pi_0^{(i)}$

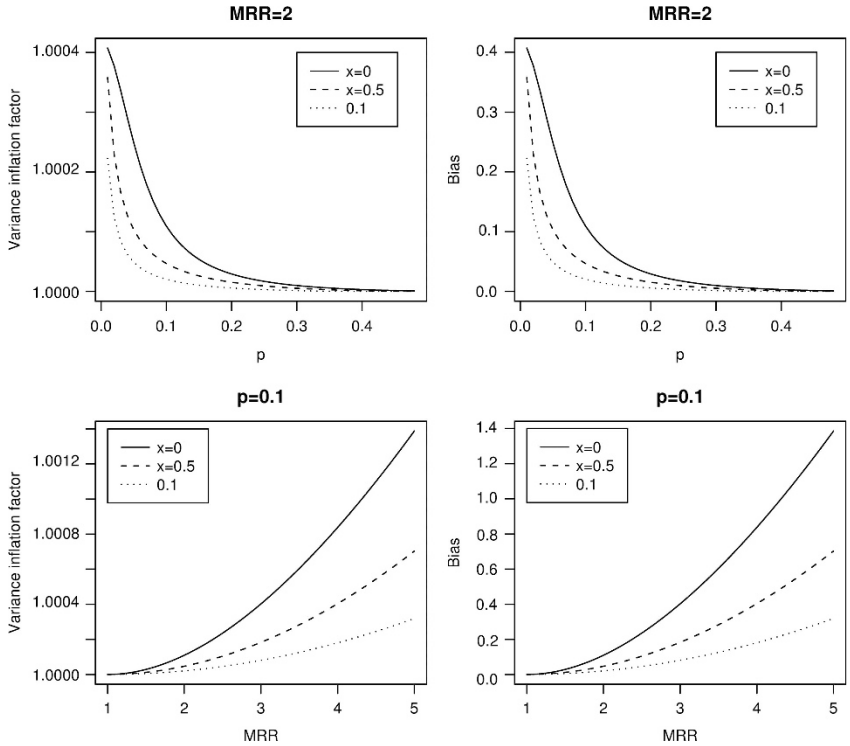

Figure 1 Variance inflation factor (VIF) and bias of Cochran–Armitage's trend (CAT) test with score x ($n_1 = n_0 = 500, \pi_1^{(0)} = \pi_0^{(0)} = 0.005$) and misclassification rate ratio (MRR).

with the CAT test for $x = 0$, 1:

$$T_p = \frac{Z_0^2 - 2\hat{\rho} Z_0 Z_1 + Z_1^2}{1 - \hat{\rho}^2} = (Z_0, Z_1) \begin{pmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{pmatrix}^{-1} \begin{pmatrix} Z_0 \\ Z_1 \end{pmatrix}, \quad (3)$$

Where $\hat{\rho} = \sqrt{m_0 m_2 / ((m_0 + m_1)(m_2 + m_1))}$. With this representation, we can correct the PCT through correcting the trend tests Z_0 and Z_1.[27]

Centralize the tests by using null markers

Centralizing the CAT test is possible if the error probabilities are known or estimable from additional information sources. Otherwise, the CAT has inflated type I error and cannot be used as a test for association when the misclassification rates are differential for the case and control groups. The GC method uses null markers that are unlinked to the disease and control to correct the bias and/or variance distortion caused by population structures. If the null markers satisfy the additional assumption that they are subject to the same (or roughly the same) misclassification distribution as the candidate marker, then these markers can also be used to correct the CAT similar to the GC method. As shown above, the variance distortion is not problematic if we consider small genotype errors. Thus, we consider only correction of bias using the information from the null markers.

Suppose there are K null markers available. We can assume that all these K markers are in Hardy–Weinberg equilibrium (HWE). Because in a large-scale genome-wide association studies, there are sufficient null markers that it is easy to pick out K markers in Hardy–Weinberg equilibrium. Suppose numerators of the CAT test statistics for the null markers are $\hat{\Delta}_x^{(1)}, \ldots, \hat{\Delta}_x^{(K)}$, then we can estimate μ_x from these statistics. A basic requirement for the null marker to be able to correct for the bias of the test on the candidate marker is that the null markers should be subject to the same (or roughly the same) misclassifications as the candidate. We will consider two different situations: (i) the null markers are matched with the candidate marker in minor allele frequencies (MAFs). In this situation, the null markers have the same or closely the same MAFs as the candidate marker, therefore it is reasonable to assume that they have the same or closely the same misclassification rates. Thus, the homogeneity assumption on misclassification rates is unquestionable; (ii) the null markers are not matched with the candidate. In this situations, we do not require the null markers to have the same MAFs as the candidate and therefore allows much more flexibility in selecting null markers.

For situation (i), one can simply estimate the bias parameter by sample mean of the $\hat{\Delta}_x^{(k)}$'s. Namely, $\hat{\mu}_x = \sum_{k=1}^{K} \hat{\Delta}_x^{(k)} / K$. This case is essentially the same as what is done in the original GC method.[1] However, we would not explore this method further since matching MAFs for null markers is not reliable in presence of genotyping error and matching can also reduce the number of available null markers that can be used. For situation (ii), the null markers are not matched with the candidate marker in MAF, therefore we need to adjust for the variability of MAFs. Let the MAF of the k-th null marker be p_k, which can be estimated by counting alleles from the genotype in control group (denoted by $\hat{p}_k$). Appendix B shows that such estimates are asymptotically unbiased when misclassification rates are small (see Appendix B for details). Note that genotypes of null markers have errors, the MAF estimate may be biased. However, since presumably the misclassification rates are small (say, between 0.1% and 0.5%), $\hat{p}_k$ is not severely biased and can be used as linear regression. We propose estimating the bias parameter by the quadratic formula

$$\hat{\mu}_x = \hat{\beta}_0 + \hat{\beta}_1 \hat{p} + \hat{\beta}_2 \hat{p}^2,$$

where $\hat{\beta}_i$ are the regression parameters obtained from the linear regression model

$$E(\hat{\Delta}_x^{(k)}) = \beta_0 + \beta_1 \hat{p}_k + \beta_2 \hat{p}_k^2, k = 1, 2, \ldots, K,$$

where $\hat{p}$ is the estimated MAF for the candidate marker and $(\hat{\Delta}_x^{(k)}, \hat{p}_k)$, $k = 1, 2, \ldots, K$ are the computed CAT test statistics and estimated MAFs from the null markers. With either estimate of μ_x in situation (i) or (ii), one can centralize the CAT test statistic by

$$Z_x^* = \frac{\Delta_x - \hat{\mu}_x}{\hat{\sigma}_x},$$

and the corrected the PCT* by

$$T_p^* = \frac{Z_0^{*2} - 2\hat{\rho}Z_0^*Z_1^* + Z_1^{*2}}{1-\hat{\rho}^2} = (Z_0^*, Z_1^*)\begin{pmatrix} 1 & \hat{\rho} \\ \hat{\rho} & 1 \end{pmatrix}^{-1}\begin{pmatrix} Z_0^* \\ Z_1^* \end{pmatrix}.$$

Simulation

Let the risk allele be A, population frequency is $p = P(A)$, genotypes are AA, Aa, aa, use $G' = 2, 1, 0$ (number of allele A) to denote the three genotypes and G the observed genotype.

We assume the following penetrance model,

$$f_k = P(Y=1 \mid G'=k) = \frac{e^{\alpha+\beta z_k}}{1+e^{\alpha+\beta z_k}} \tag{4}$$

where Y is the indicator of disease, z_k is the score of genotype k. We take $z_0 \equiv 0$, $z_2 \equiv 1$ and use $z_1 = 1$ for dominant model, $z_1 = 0.5$ for additive model and $z_1 = 0$ for recessive model, respectively. Note that e^{β} is the genotype relative risk and α is baseline log odds. In our simulation studies, we will investigate the GC method under the recessive (REC), additive (ADD) and dominant models (DOM). The corresponding tests are CAT with $x = 0$, 0.5 and 1, respectively. Assuming Hardy–Weinberg equilibrium in the population, then the population genotype frequency $r_k = P(G'=k) = \binom{2}{k}p^k(2-p)^{2-k}$, where $k = 0$, 1, 2 and $p = P(A)$ is the MAF for allele A in population. The genotype frequencies for the two groups are

$$p_{ij} = P(G=j \mid Y=i) = \frac{\sum_{k=0}^{2} P(Y=i \mid G'=k)P(G=j \mid G'=k)P(G'=k)}{\sum_{j=0}^{2}\sum_{k=0}^{2} P(Y=i \mid G'=k)P(G=j \mid G'=k)P(G'=k)}$$

$$= \frac{\sum_{k=0}^{2} f_k^i(1-f_k)^{1-i}\pi_{kj}^{(i)} r_k}{\sum_{j=0}^{2}\sum_{k=0}^{2} f_k^i(1-f_k)^{1-i}\pi_{kj}^{(i)} r_k},$$

where $\pi_{kj}^{(i)}$ is specified by the given error model (for example, the (j, k) entry in Table 2).

Genotype data for null markers are generated in similar way except that these markers are not linked to the diseases (that is, $\beta = 0$ in equation 4). In our simulation study, we do not require the null markers match with the candidate marker in MAF, their MAFs are randomly generated from uniform

Table 3 Type I error of the corrected and uncorrected trend tests and the PCT ($\alpha = 0.05$, $\pi_1^{(0)} = 0.005$, $\pi_0^{(0)} = 0.003$, $\varepsilon = 0.002$, $n_1 = n_0 = 1000$, $K = 100$)

		P = *0.1*				P = *0.3*			
Model	*MRR*	*CAT*	*CAT**	*PCT*	*PCT**	*CAT*	*CAT**	*PCT*	*PCT**
		r = 0							
REC	1	0.057	0.056	0.054	0.057	0.053	0.058	0.044	0.048
	3	0.091	0.069	0.084	0.061	0.048	0.057	0.046	0.048
	5	0.194	0.061	0.155	0.046	0.070	0.049	0.053	0.048
	7	0.370	0.066	0.273	0.055	0.080	0.059	0.070	0.058
	9	0.535	0.054	0.442	0.058	0.103	0.054	0.081	0.052
ADD	1	0.052	0.053	0.059	0.064	0.048	0.051	0.056	0.057
	3	0.062	0.059	0.077	0.059	0.049	0.051	0.062	0.063
	5	0.119	0.047	0.158	0.058	0.053	0.048	0.059	0.060
	7	0.218	0.056	0.278	0.058	0.057	0.044	0.062	0.048
	9	0.333	0.054	0.440	0.057	0.080	0.049	0.097	0.056
DOM	1	0.052	0.056	0.051	0.056	0.047	0.051	0.044	0.046
	3	0.057	0.058	0.064	0.048	0.060	0.054	0.048	0.052
	5	0.105	0.058	0.152	0.059	0.052	0.046	0.054	0.053
	7	0.136	0.050	0.294	0.068	0.053	0.044	0.067	0.059
	9	0.204	0.053	0.446	0.060	0.059	0.054	0.084	0.057
					r = 0.05				
REC	1	0.059	0.084	0.048	0.051	0.046	0.061	0.047	0.055
	3	0.093	0.065	0.072	0.064	0.060	0.087	0.050	0.061
	5	0.202	0.064	0.141	0.054	0.070	0.072	0.060	0.056
	7	0.372	0.062	0.282	0.061	0.084	0.071	0.072	0.061
	9	0.520	0.061	0.450	0.058	0.100	0.060	0.079	0.048
ADD	1	0.054	0.084	0.059	0.066	0.050	0.068	0.052	0.060
	3	0.067	0.051	0.067	0.057	0.049	0.058	0.041	0.050
	5	0.125	0.052	0.154	0.053	0.050	0.052	0.062	0.054
	7	0.220	0.064	0.283	0.056	0.079	0.057	0.068	0.055
	9	0.351	0.053	0.450	0.061	0.077	0.047	0.086	0.066
DOM	1	0.049	0.051	0.053	0.065	0.061	0.059	0.051	0.058
	3	0.061	0.075	0.072	0.060	0.053	0.056	0.063	0.064
	5	0.090	0.054	0.154	0.050	0.053	0.058	0.060	0.065
	7	0.120	0.064	0.273	0.055	0.051	0.065	0.067	0.060
	9	0.204	0.052	0.462	0.064	0.063	0.081	0.085	0.057

Abbreviations: ADD, additive; CAT, Cochran–Armitage's trend test; DOM, dominant; MRR, misclassification rate ratio; PCT, Pearson's chi-square test; REC, recessive. CAT* and PCT* are respectively the corrected version of CAT and PCT.

distribution $U(0, 0.5)$. The genotype misclassification for the null markers are taken to be close to those for the candidate marker. Specifically, for each null marker we generate the error probabilities for control from $U(\pi_i^{(0)} - \varepsilon, \pi_i^{(0)} + \varepsilon), i = 0, 1$ and the error probabilities of case group are then determined by these values and the given value of MRR.

RESULTS

For the candidate marker, MAF p is assumed to be 0.1 and 0.3, respectively. The misclassification rates are $\pi_1^{(0)} = 0.005$ and $\pi_0^{(0)} = 0.003$ for the control group in our simulation. The differential index, MRR, is taken to be 1, 3, 5, 7, 9. Note that when MRR = 1, the errors are not differential. We use the error model shown in Table 2. Then the misclassification rates for the case group can be calculated by $\pi_i^{(0)} (i = 0,1)$ and MRRs. Null markers are generated independently with the candidate marker. Their MAFs are generated from uniform distribution $U(0, 0.5)$. The null markers are in principle unlinked to the disease, but we allow a proportion, r, of them to be linked with disease. When $r = 0$, the null markers are unlinked to the disease. The misclassification rates for the null markers are taken to be close to the candidate marker allowing slight deviations ($\varepsilon = 0.002$), specifically, the error rates for the null markers in control groups are generated from $\pi_1^{(0)} \sim U(0.003, 0.007)$, $\pi_0^{(0)} \sim U(0.001, 0.005)$. The number of null markers is $K = 100$. The sample size is $n = 2000$ (1000 cases and 1000 controls). We investigate the three common modes of inheritance of recessive, additive and dominant models indexed respectively by $x = 0$, 0.5 and 1. All simulations are run with 2000 replicates.

Tables 3 and 4 show the simulation results for nominal significance levels $\alpha = 0.05$ and 0.01, respectively. From Tables 3 and 4, the type I errors of CAT are considerably inflated for large MRRs. For example, when the disease model is recessive, MAF = 0.1 and MRR = 5, the type I error is 0.194 for CAT and 0.155 for PCT if the nominal significance level is 0.05. Generally, the type I error of CAT or PCT is increasingly inflated when MRR increases or MAF decreases. The inflation is most evident for rare allele cases and is negligible when MAF is close to 0.5. In addition, the inflation is largest for recessive model and smallest for dominant model. The same phenomenon is observed in all the scenarios we investigated with different error models. Tables 5 and 6 give the type I errors for $K = 50$ and $K = 200$.

Table 4 Type I error of the corrected and uncorrected trend tests and the PCT ($\alpha = 0.01, \pi_1^{(0)} = 0.005, \pi_0^{(0)} = 0.003, \varepsilon = 0.002, n_1 = n_0 = 1000, K = 100$)

		P = *0.1*				P = *0.3*			
Model	*MRR*	*CAT*	*CAT**	*PCT*	*PCT**	*CAT*	*CAT**	*PCT*	*PCT**
					$r = 0$				
REC	1	0.009	0.015	0.007	0.011	0.011	0.011	0.009	0.013
	3	0.021	0.017	0.021	0.012	0.009	0.010	0.009	0.008
	5	0.063	0.017	0.043	0.013	0.014	0.010	0.012	0.009
	7	0.147	0.015	0.101	0.010	0.025	0.013	0.020	0.014
	9	0.266	0.011	0.203	0.016	0.029	0.007	0.020	0.011
ADD	1	0.004	0.005	0.013	0.014	0.009	0.011	0.008	0.011
	3	0.015	0.011	0.021	0.014	0.013	0.011	0.013	0.015
	5	0.030	0.013	0.044	0.009	0.009	0.009	0.012	0.010
	7	0.084	0.011	0.113	0.013	0.013	0.008	0.011	0.007
	9	0.134	0.009	0.216	0.017	0.021	0.009	0.021	0.013
DOM	1	0.011	0.014	0.005	0.009	0.008	0.007	0.007	0.007
	3	0.015	0.013	0.012	0.005	0.011	0.011	0.011	0.013
	5	0.023	0.014	0.050	0.012	0.011	0.011	0.014	0.013
	7	0.044	0.013	0.128	0.015	0.011	0.011	0.018	0.010
	9	0.067	0.011	0.223	0.011	0.013	0.011	0.025	0.013
					$r = 0.05$				
REC	1	0.007	0.015	0.007	0.012	0.005	0.008	0.010	0.011
	3	0.021	0.011	0.017	0.013	0.013	0.022	0.013	0.013
	5	0.054	0.011	0.038	0.010	0.017	0.018	0.015	0.012
	7	0.139	0.018	0.104	0.013	0.021	0.017	0.019	0.012
	9	0.257	0.026	0.227	0.017	0.030	0.015	0.019	0.013
ADD	1	0.009	0.011	0.010	0.016	0.009	0.018	0.009	0.012
	3	0.013	0.007	0.018	0.011	0.008	0.009	0.007	0.010
	5	0.032	0.014	0.051	0.014	0.010	0.010	0.013	0.014
	7	0.085	0.014	0.108	0.010	0.021	0.010	0.018	0.013
	9	0.149	0.013	0.223	0.011	0.018	0.012	0.021	0.011
DOM	1	0.011	0.014	0.011	0.014	0.011	0.012	0.011	0.015
	3	0.012	0.027	0.019	0.013	0.013	0.013	0.015	0.015
	5	0.027	0.011	0.042	0.013	0.007	0.009	0.010	0.015
	7	0.037	0.010	0.102	0.015	0.011	0.016	0.015	0.014
	9	0.064	0.011	0.224	0.013	0.011	0.025	0.016	0.012

Abbreviations: ADD, additive; CAT, Cochran–Armitage's trend test; DOM, dominant; MRR, misclassification rate ratio; PCT, Pearson's chi-square test; REC, recessive. CAT* and PCT* are respectively the corrected version of CAT and PCT.

Table 5 Type I error of the corrected and uncorrected trend tests and the Pearson's chi-square test ($\alpha = 0.05, \pi_1^{(0)} = 0.005, \pi_0^{(0)} = 0.003, \varepsilon = 0.002, n_1 = n_0 = 1000, K = 50, 200$)

			P = *0.1*		P = *0.2*		P = *0.3*		P = *0.4*		P = *0.5*	
K	*Model*	*MRR*	*CAT*	*CAT**	*CAT*	*CAT**	*CAT*	*CAT**	*CAT*	*CAT**	*CAT*	*CAT**
50	REC	1	0.050	0.058	0.055	0.064	0.051	0.059	0.050	0.058	0.052	0.048
		3	0.086	0.095	0.058	0.061	0.053	0.058	0.056	0.056	0.048	0.048
		5	0.203	0.071	0.093	0.054	0.067	0.059	0.050	0.050	0.055	0.051
		7	0.346	0.061	0.133	0.053	0.081	0.065	0.060	0.051	0.054	0.049
		9	0.511	0.071	0.214	0.060	0.120	0.061	0.069	0.052	0.057	0.058
	ADD	1	0.057	0.066	0.045	0.049	0.050	0.054	0.045	0.050	0.045	0.053
		3	0.072	0.060	0.052	0.051	0.051	0.051	0.050	0.051	0.050	0.051
		5	0.127	0.062	0.067	0.058	0.051	0.045	0.053	0.055	0.044	0.050
		7	0.220	0.057	0.105	0.045	0.057	0.046	0.049	0.050	0.051	0.047
		9	0.345	0.054	0.128	0.043	0.067	0.046	0.056	0.057	0.053	0.056
	DOM	1	0.044	0.060	0.045	0.046	0.060	0.060	0.055	0.059	0.048	0.051
		3	0.050	0.048	0.048	0.046	0.045	0.049	0.048	0.054	0.048	0.049
		5	0.081	0.051	0.054	0.048	0.053	0.054	0.041	0.043	0.060	0.061
		7	0.148	0.046	0.061	0.050	0.051	0.051	0.043	0.044	0.056	0.052
		9	0.209	0.050	0.086	0.058	0.054	0.054	0.053	0.054	0.061	0.060
200	REC	1	0.049	0.052	0.049	0.051	0.053	0.056	0.051	0.053	0.044	0.045
		3	0.092	0.059	0.059	0.047	0.050	0.045	0.053	0.055	0.047	0.048
		5	0.193	0.055	0.085	0.049	0.068	0.051	0.057	0.048	0.054	0.055
		7	0.345	0.053	0.129	0.052	0.083	0.051	0.064	0.053	0.051	0.051
		9	0.518	0.054	0.194	0.051	0.112	0.058	0.067	0.052	0.052	0.047
	ADD	1	0.047	0.051	0.052	0.052	0.050	0.052	0.049	0.048	0.051	0.052
		3	0.065	0.047	0.055	0.051	0.049	0.050	0.051	0.050	0.046	0.046
		5	0.123	0.049	0.073	0.051	0.061	0.050	0.055	0.051	0.049	0.050
		7	0.217	0.053	0.097	0.046	0.060	0.051	0.050	0.043	0.054	0.050
		9	0.336	0.044	0.127	0.047	0.077	0.044	0.056	0.045	0.048	0.045
	DOM	1	0.047	0.048	0.051	0.052	0.053	0.053	0.053	0.054	0.052	0.053
		3	0.054	0.051	0.052	0.049	0.051	0.048	0.056	0.053	0.049	0.049
		5	0.089	0.047	0.054	0.044	0.050	0.046	0.055	0.055	0.048	0.050
		7	0.130	0.046	0.066	0.046	0.057	0.048	0.050	0.047	0.052	0.045
		9	0.201	0.050	0.084	0.044	0.054	0.046	0.053	0.052	0.050	0.043

Abbreviations: ADD, additive; CAT, Cochran–Armitage's trend test; DOM, dominant; MRR, misclassification rate ratio; REC, recessive. CAT* and PCT* are respectively the corrected version of CAT and PCT.

Table 6 Type I error of the corrected and uncorrected trend tests and the Pearson's chi-square test ($\alpha = 0.01, \pi_1^{(0)} = 0.005, \pi_0^{(0)} = 0.003, \varepsilon = 0.002, n_1 = n_0 = 1000, K = 50, 200$)

			P = *0.1*		P = *0.2*		P = *0.3*		P = *0.4*		P = *0.5*	
K	*Model*	*MRR*	*CAT*	*CAT**	*CAT*	*CAT**	*CAT*	*CAT**	*CAT*	*CAT**	*CAT*	*CAT**
50	REC	1	0.005	0.010	0.011	0.014	0.011	0.013	0.008	0.013	0.005	0.011
		3	0.025	0.012	0.014	0.011	0.013	0.013	0.010	0.011	0.009	0.009
		5	0.071	0.015	0.028	0.014	0.018	0.015	0.011	0.007	0.009	0.011
		7	0.135	0.013	0.044	0.015	0.024	0.013	0.017	0.011	0.008	0.011
		9	0.267	0.015	0.088	0.014	0.043	0.013	0.014	0.011	0.008	0.011
	ADD	1	0.013	0.020	0.014	0.017	0.009	0.009	0.007	0.009	0.010	0.013
		3	0.011	0.015	0.011	0.012	0.008	0.007	0.011	0.009	0.008	0.009
		5	0.048	0.015	0.018	0.013	0.013	0.009	0.008	0.006	0.010	0.010
		7	0.077	0.011	0.029	0.009	0.010	0.007	0.009	0.010	0.012	0.013
		9	0.159	0.009	0.038	0.011	0.017	0.010	0.009	0.008	0.008	0.009
	DOM	1	0.007	0.011	0.011	0.013	0.016	0.014	0.018	0.018	0.007	0.008
		3	0.012	0.010	0.008	0.010	0.013	0.012	0.011	0.011	0.008	0.010
		5	0.018	0.011	0.009	0.009	0.010	0.011	0.007	0.008	0.014	0.015
		7	0.038	0.011	0.018	0.011	0.015	0.011	0.010	0.009	0.013	0.013
		9	0.072	0.010	0.025	0.015	0.011	0.011	0.011	0.013	0.013	0.012
200	REC	1	0.008	0.009	0.009	0.010	0.011	0.011	0.009	0.009	0.009	0.010
		3	0.025	0.012	0.011	0.010	0.011	0.008	0.011	0.009	0.008	0.007
		5	0.061	0.010	0.022	0.009	0.016	0.008	0.011	0.010	0.012	0.013
		7	0.140	0.011	0.039	0.012	0.020	0.010	0.013	0.011	0.010	0.011
		9	0.264	0.012	0.070	0.010	0.034	0.012	0.016	0.010	0.011	0.010
	ADD	1	0.010	0.011	0.010	0.011	0.011	0.011	0.009	0.010	0.012	0.012
		3	0.017	0.012	0.012	0.012	0.009	0.009	0.009	0.008	0.009	0.009
		5	0.034	0.009	0.018	0.010	0.013	0.012	0.010	0.011	0.009	0.009
		7	0.077	0.010	0.026	0.007	0.015	0.010	0.009	0.006	0.009	0.008
		9	0.151	0.010	0.036	0.009	0.018	0.009	0.012	0.010	0.010	0.008
	DOM	1	0.008	0.008	0.010	0.010	0.008	0.007	0.011	0.011	0.010	0.012
		3	0.011	0.010	0.010	0.011	0.010	0.008	0.010	0.011	0.009	0.008
		5	0.024	0.007	0.011	0.010	0.011	0.011	0.012	0.012	0.010	0.010
		7	0.043	0.011	0.016	0.010	0.012	0.012	0.009	0.008	0.012	0.009
		9	0.073	0.012	0.021	0.009	0.013	0.009	0.010	0.009	0.010	0.008

Abbreviations: ADD, additive; CAT, Cochran–Armitage's trend test; DOM, dominant; MRR, misclassification rate ratio; REC, recessive. CAT* and PCT* are respectively the corrected version of CAT and PCT.

Table 7 Type I error of the corrected and uncorrected trend tests and the PCT ($\alpha = 0.05, \pi_1^{(0)} = 0.02, \pi_0^{(0)} = 0.01, \varepsilon = 0.01$, $n_1 = n_0 = 1000, K = 100$)

		P = *0.1*					P = *0.3*			
Model	*MRR*	*CAT*	*CAT**	*PCT*	*PCT**		*CAT*	*CAT**	*PCT*	*PCT**
						$r = 0$				
REC	1	0.051	0.063	0.049	0.056		0.052	0.058	0.051	0.049
	3	0.057	0.053	0.046	0.051		0.053	0.051	0.052	0.058
	5	0.069	0.060	0.063	0.057		0.054	0.056	0.054	0.053
	7	0.085	0.057	0.079	0.054		0.052	0.049	0.047	0.049
	9	0.128	0.060	0.102	0.060		0.053	0.051	0.061	0.068
ADD	1	0.054	0.052	0.050	0.056		0.041	0.044	0.043	0.051
	3	0.060	0.059	0.046	0.048		0.043	0.046	0.061	0.062
	5	0.052	0.051	0.058	0.051		0.051	0.046	0.045	0.049
	7	0.068	0.053	0.076	0.060		0.051	0.049	0.051	0.055
	9	0.091	0.052	0.108	0.054		0.053	0.050	0.048	0.047
DOM	1	0.041	0.046	0.043	0.049		0.048	0.047	0.049	0.053
	3	0.051	0.055	0.049	0.058		0.049	0.049	0.058	0.060
	5	0.052	0.052	0.063	0.059		0.054	0.051	0.051	0.052
	7	0.060	0.049	0.084	0.051		0.057	0.056	0.050	0.052
	9	0.071	0.053	0.103	0.053		0.051	0.046	0.054	0.059
						$r = 0.05$				
REC	1	0.051	0.123	0.049	0.052		0.050	0.057	0.045	0.047
	3	0.059	0.066	0.054	0.062		0.051	0.053	0.048	0.053
	5	0.071	0.059	0.060	0.059		0.045	0.070	0.056	0.060
	7	0.083	0.048	0.078	0.050		0.051	0.053	0.055	0.058
	9	0.132	0.060	0.107	0.063		0.065	0.059	0.050	0.051
ADD	1	0.051	0.058	0.050	0.056		0.054	0.055	0.048	0.050
	3	0.051	0.065	0.047	0.050		0.049	0.059	0.050	0.052
	5	0.054	0.058	0.062	0.051		0.040	0.043	0.049	0.051
	7	0.071	0.053	0.085	0.057		0.053	0.048	0.051	0.056
	9	0.098	0.056	0.099	0.059		0.056	0.056	0.053	0.054
DOM	1	0.050	0.056	0.043	0.051		0.056	0.075	0.049	0.052
	3	0.062	0.099	0.051	0.059		0.062	0.060	0.042	0.044
	5	0.050	0.047	0.069	0.053		0.047	0.044	0.051	0.051
	7	0.068	0.060	0.068	0.053		0.051	0.059	0.052	0.052
	9	0.064	0.055	0.098	0.056		0.052	0.057	0.049	0.049

Abbreviations: ADD, additive; CAT, Cochran–Armitage's trend test; DOM, dominant; MRR, misclassification rate ratio; PCT, Pearson's chi-square test; REC, recessive. CAT* and PCT* are respectively the corrected version of CAT and PCT.

Table 8 Type I error of the corrected and uncorrected trend tests and the PCT ($\alpha = 0.01, \pi_1^{(0)} = 0.02, \pi_0^{(0)} = 0.01, \varepsilon = 0.01$, $n_1 = n_0 = 1000, K = 100$)

		P = *0.1*					P = *0.3*			
Model	*MRR*	*CAT*	*CAT**	*PCT*	*PCT**		*CAT*	*CAT**	*PCT*	*PCT**
						$r = 0$				
REC	1	0.011	0.018	0.007	0.009		0.011	0.012	0.013	0.011
	3	0.009	0.011	0.011	0.011		0.012	0.011	0.011	0.013
	5	0.013	0.014	0.015	0.011		0.011	0.009	0.011	0.013
	7	0.021	0.011	0.019	0.009		0.011	0.010	0.011	0.011
	9	0.029	0.011	0.024	0.015		0.011	0.011	0.011	0.017
ADD	1	0.009	0.011	0.009	0.010		0.006	0.006	0.009	0.011
	3	0.011	0.012	0.013	0.010		0.010	0.009	0.010	0.011
	5	0.016	0.009	0.011	0.009		0.005	0.006	0.009	0.009
	7	0.011	0.013	0.020	0.012		0.014	0.013	0.013	0.013
	9	0.028	0.010	0.027	0.010		0.013	0.011	0.011	0.007
DOM	1	0.006	0.006	0.010	0.011		0.011	0.010	0.009	0.009
	3	0.010	0.014	0.006	0.009		0.009	0.009	0.011	0.013
	5	0.015	0.010	0.017	0.013		0.010	0.009	0.006	0.006
	7	0.011	0.011	0.021	0.013		0.012	0.011	0.013	0.015
	9	0.018	0.013	0.023	0.009		0.011	0.011	0.010	0.010
						$r = 0.05$				
REC	1	0.008	0.042	0.005	0.010		0.011	0.013	0.011	0.009
	3	0.009	0.018	0.009	0.011		0.009	0.011	0.011	0.015
	5	0.011	0.011	0.015	0.011		0.009	0.018	0.012	0.011
	7	0.021	0.009	0.018	0.013		0.007	0.010	0.011	0.013
	9	0.034	0.014	0.031	0.012		0.014	0.015	0.013	0.013
ADD	1	0.010	0.013	0.009	0.013		0.010	0.011	0.011	0.009
	3	0.013	0.015	0.010	0.009		0.011	0.012	0.007	0.007
	5	0.011	0.014	0.016	0.011		0.009	0.007	0.009	0.011
	7	0.015	0.011	0.017	0.013		0.010	0.009	0.012	0.011
	9	0.024	0.010	0.029	0.016		0.012	0.013	0.012	0.011
DOM	1	0.009	0.009	0.010	0.011		0.015	0.018	0.004	0.013
	3	0.018	0.034	0.010	0.013		0.011	0.010	0.009	0.011
	5	0.011	0.009	0.013	0.011		0.013	0.013	0.011	0.010
	7	0.017	0.011	0.016	0.009		0.011	0.012	0.009	0.012
	9	0.016	0.010	0.022	0.009		0.009	0.010	0.009	0.007

Abbreviations: ADD, additive; CAT, Cochran–Armitage's trend test; DOM, dominant; MRR, misclassification rate ratio; PCT, Pearson's chi-square test; REC, recessive. CAT* and PCT* are respectively the corrected version of CAT and PCT.

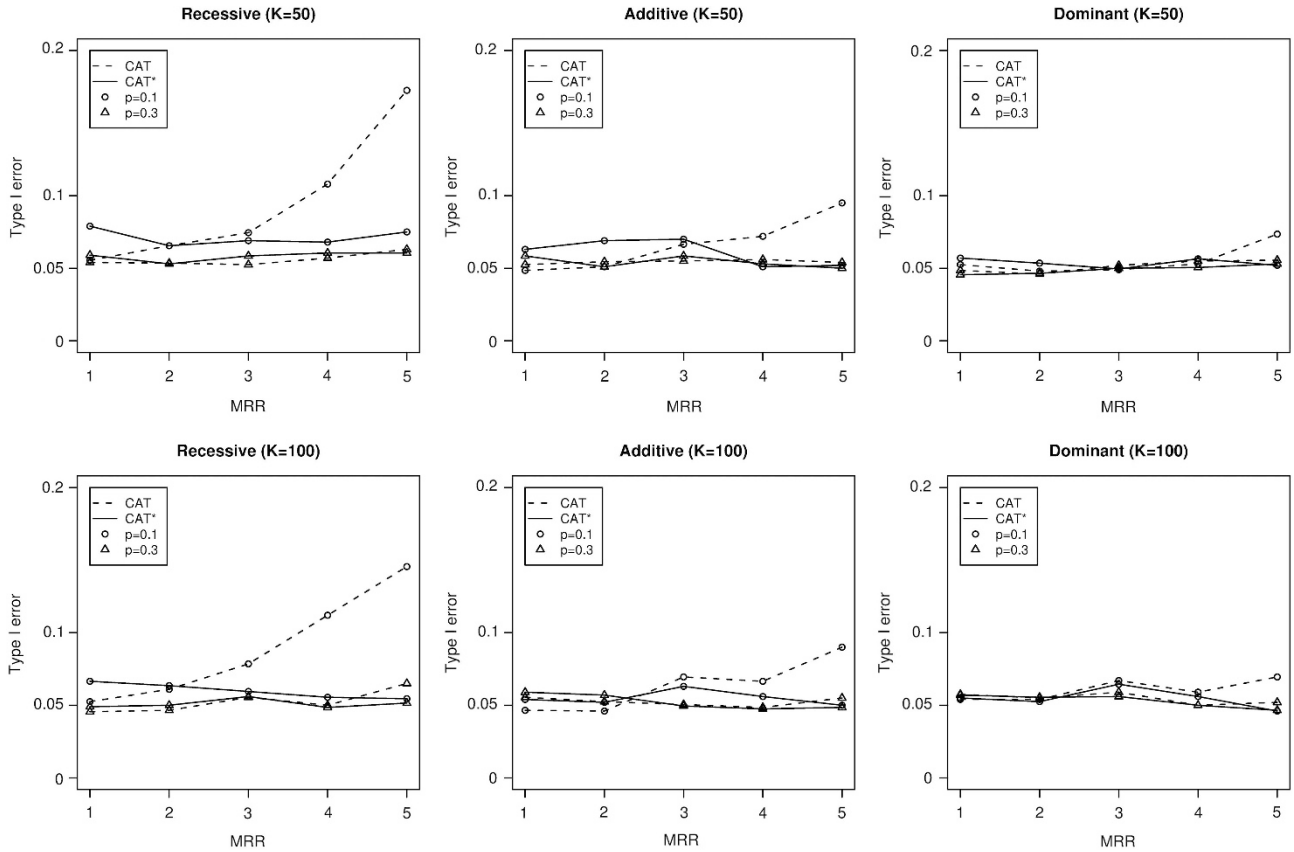

Figure 2 Type I error of Cochran–Armitage's trend (CAT) tests and the corrected trend tests (CAT*); ($n_1 = n_0 = 1000$, $\alpha = 0.05$, $\pi_1^{(0)} = 0.0025$, $\pi_0^{(0)} = 0.0025$, $\varepsilon = 0.0015$). MRR, misclassification rate ratio.

Type I error can be well controlled even with small number of null markers; for example, $K = 50$. When the number of null markers is becoming larger, the type I errors are better controlled, but there is not too much improvement when the number is greater than $K = 100$. Thus, we suggest that $K = 100$ is much enough in practice. Tables 7 and 8 report the type I errors when the misclassification rates are as large as 2%. From these two tables, we can conclude that the type I errors of the corrected CAT and PCT are much closer to the nominal level than the non-corrected ones with large misclassification rates.

Note that when MRR = 1, the genotype errors are not differential for cases and controls. The type I errors of CAT and PCT are generally close to the nominal values except when the mode of inheritance is recessive. When sample size becomes larger, the type I errors of CAT or PCT are all close to the nominal values when MRR = 1 (results not shown here). We observed that the corrected CAT (denoted by CAT* in the table) has slightly inflated type I error compared with the original CAT. This is intuitively true since use of null markers introduced more variabilities. When MRR is greater than 1, that is, differential errors are present, the original CAT has substantially inflated type I error and the corrected CAT has type I error that are reasonably close to the nominal values for all the situations we considered.

Figure 2 illustrates the type I errors and the corrected type I errors under another setting of parameters ($\pi_1^{(0)} = \pi_0^{(0)} = 0.0025$, $\varepsilon = 0.0015$). Figure 3 illustrates the power of the CAT and the quadratic regression GC-corrected CAT. Upper row is for MRR = 1 and lower row is for MRR = 4. We can see from the upper row that, when MRR = 1, use of null markers does not affect the power performance of the trend test under the three genetic models. From the lower row, because of inflation of type I error, power of CAT test is also inflated. Our method can correctly control type I error of CAT and consequently report correct power and avoid spurious association.

DISCUSSION

In this study, we proposed to correct for the bias introduced by differential errors in association analysis by a regression method using null markers. The bias of the association tests can be effectively captured by null markers that are unlinked to the disease. We have not required that the null markers match with the candidate marker in MAFs and we estimate the bias by using a quadratic regression method to adjust for the variabilities of MAFs of the null markers. The regression method fit the bias by a quadratic function of the MAFs. We have shown that the approach works well although the MAF estimates have biases with the presence of genotyping errors. This is because the bias of MAF estimates are small and does not have a significant role in the quadratic regression.

Misclassification rates for null markers (allele frequencies that are close to that of the candidate marker) are assumed to be the same with the candidates. However, it might be true that different loci may have different misclassification rates, which might depend on the specific allele frequency or genotype frequencies on each locus. We have recognized this and allowed the misclassification rates to be deviated slightly from that for the candidate marker. Simulation

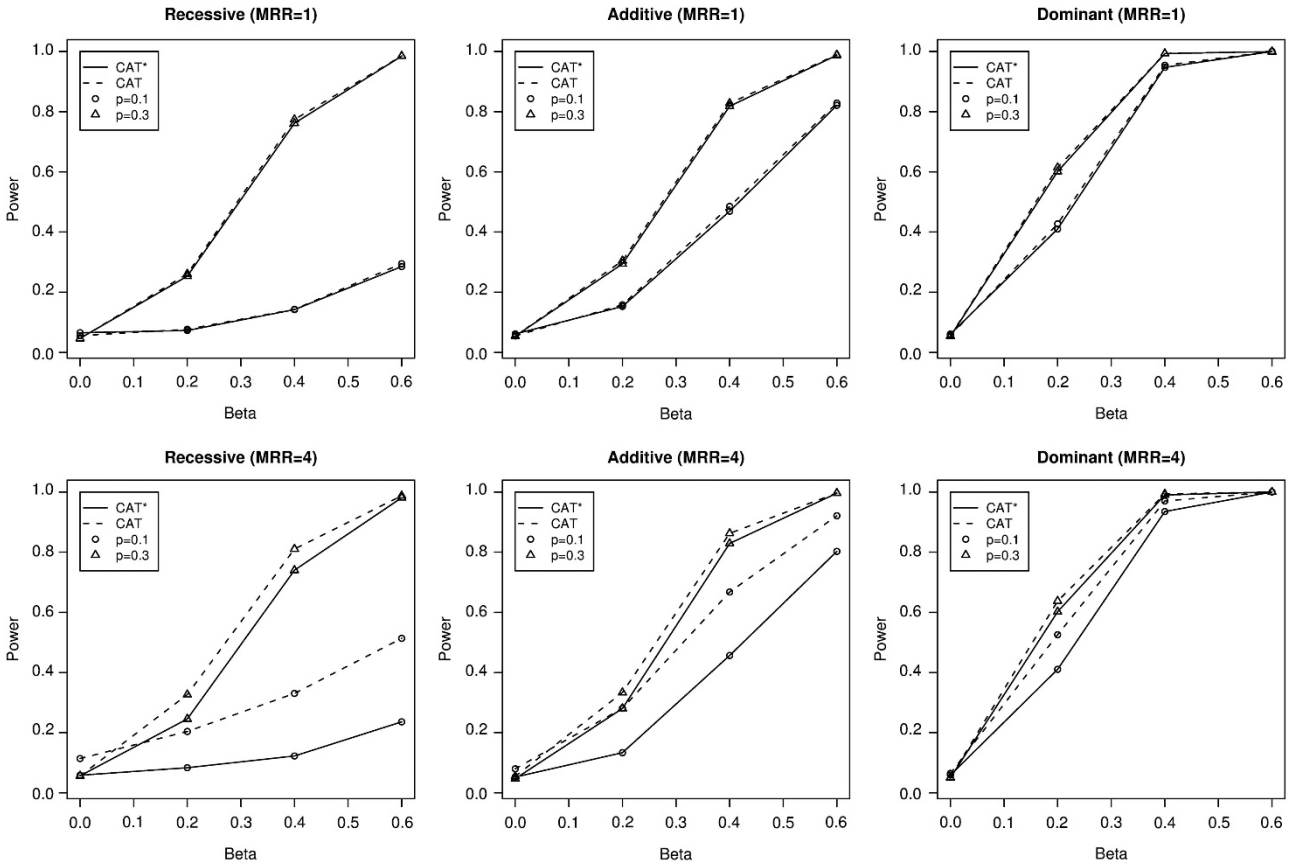

Figure 3 Power of the Cochran–Armitage's trend (CAT) tests and the corrected trend tests (CAT*); ($n_1 = n_0 = 1000$, nominal level $\alpha = 0.05$, number of null markers $K = 100$; $\pi_1^{(0)} = 0.0025, \pi_0^{(0)} = 0.0025$, $\varepsilon = 0.0015$). MRR, misclassification rate ratio.

studies show that with slight fluctuations in misclassification rates of the null markers, the proposed method performs reasonably well.

ACKNOWLEDGEMENTS

This work is partially supported by the National Science Foundation of China (NSFC), grant no. 11201452 and no. 11271346.

1 Devlin, B. & Roeder, K Genomic control for association studies. *Biometrics* **55,** 997–1004 (1999).
2 Pritchard, J. K. & Rosenberg, N. A Use of unlinked genetic markers to detect population stratification in association studies. *Am. J. Hum. Genet* **65,** 220–228 (1999).
3 Pritchard, J. K., Stephens, M., Rosenberg, N. A. & Donnelly, P Association mapping in structured populations. *Am. J. Hum. Genet.* **67,** 170–181 (2000).
4 Bacanu, S. A., Devlin, B. & Roeder, K The power of genomic control. *Am. J. Hum. Genet.* **66,** 1933–1944 (2000).
5 Devlin, B., Roeder, K. & Wasserman, L Genomic control, a new approach to genetic-based association studies. *Theoret. Popul. Biol* **60,** 155–166 (2001).
6 Satten, G. A., Flanders, W. D. & Yang, Q Accounting for unmeasured population structure in case-control studies of genetic association using a novel latent-class model. *Am. J. Hum. Genet* **68,** 466–477 (2001).
7 Patterson, N., Price, A. L. & Reich, D Population structure and eigenanalysis. *PLoS Genet.* **2,** e190 (2006).
8 Yu, J., Pressoir, G., Briggs, W. H., Vroh, B. i. I., Yamasaki, M., Doebley, J. F. *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet* **38,** 203–208 (2006).
9 Kimmel, G., Jordan, M. I., Halperin, E., Shamir, R. & Karp, R. M. A randomization test for controlling population stratification in whole-genome association studies. *Am. J. Hum. Genet* **81,** 895–905 (2007).
10 Clayton, D. G., Walker, N. M., Smyth, D. J., Pask, R., Cooper, J. D., Maier, L. M. *et al.* Population structure, differential bias, and genomic control in a large scale, case-control association study. *Nat. Genet* **37,** 1243–1246 (2005).
11 Moskvina, V., Craddock, N., Holmans, P., Owen, M. J. & O'Donovan, M. C Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum. Hered* **61,** 55–64 (2006).
12 Hao, K., Li, C., Rosenow, C. & Hung Wong, W Estimation of genotype error rate using samples with pedigree information–an application on the GeneChip Mapping 10K array. *Genomics* **84,** 623–630 (2004).
13 Tintle, N. L., Ahn, K., Mendell, N. R., Gordon, D. & Finch, S. J Characteristics of replicated single-nucleotide polymorphism genotypes from COGA: affymetrix and center for inherited disease research. *BMC. Genet.* **6** (suppl 1), S154 (2005).
14 Gordon, D., Yang, Y., Haynes, C., Finch, S. J., Mendell, N. R., Brown, A. M. *et al.* Increasing power for tests of genetic association in the presence of phenotype and/or genotype error by use of double-sampling. *Stat. Appl. Genet. Mol. Biol* **3,** Article 26 (2004).
15 Gordon, D., Haynes, C., Yang, Y., Kramer, P. L. & Finch, S. J Linear trend tests for case-control genetic association that incorporate random phenotype and genotype misclassification error. *Genet. Epidemiol* **31,** 853–870 (2007).
16 Ji, F., Yang, Y., Haynes, C., Finch, S. J. & Gordon, D Computing asymptotic power and sample size for case-control genetic association studies in the presence of phenotype and/or genotype misclassification errors. *Stat. Appl. Genet. Mol. Biol* **4,** Article 37 (2005).
17 Levenstien, M. A., Ott, J. & Gordon, D Are molecular haplotypes worth the time and expense? A cost-effective method for applying molecular haplotypes. *PLoS Genet.* **2,** e127 (2006).
18 Barral, S., Haynes, C., Stone, M. & Gordon, D LRTae. Improving statistical power for genetic association with case/control data when phenotype and/or genotype misclassification errors are present. *BMC Genet.* **7,** article 24 (2006).
19 Pearce, C. L., Van Den Berg, D. J., Makridakis, N., Reichardt, J. K., Ross, R. K., Pike, M. C. *et al.* No association between the SRD5A2 gene A49T missense variant and prostate cancer risk: lessons learned. *Hum. Mol. Genet* **17,** 2456–2461 (2008).
20 Gorroochurn, P., Heiman, G. A., Hodge, S. E. & Greenberg, D. A Centralizing the noncentral chi-square: a new method to correct for population stratification

in genetic case-control association studies. *Genet. Epidemiol* **30,** 277–289 (2006).
21 Moorhead, M., Hardenbol, P., Siddiqui, F., Falkowski, M., Bruckner, C., Ireland, J. *et al.* Optimal genotype determination in highly multiplexed SNP data. *Eur. J. Hum. Genet.* **14,** 207–215 (2006).
22 Douglas, J. A., Skol, A. D. & Boehnke, M Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *Am. J. Hum. Genet* **70,** 487–495 (2002).
23 Cochran, W. G Some methods for strengthening the common chi-squared tests. *Biometrics* **10,** 417–451 (1954).
24 Armitage, P Tests for linear trends in proportions and frequencies. *Biometrics* **11,** 375–386 (1955).
25 Zheng, G., Freidlin, B. & Gastwirth, J. L Robust genomic control for association studies. *Am. J. Hum. Genet.* **78,** 350–356 (2006).
26 Ahn, K., Gordon, D. & Finch, S. J Increase of rejection rate in case-control studies with the differential genotyping error rates. *Stat. Appl. Genet. Mol. Biol* **8,** Article 25 (2009).
27 Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A. & Reich, D Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet* **38,** 904–909 (2006).

APPENDIX A

Under the null hypothesis

$$\mu_x = E_{H_0}(\hat{\Delta}_x) = p_{12} - p_{02} + x(p_{11} - p_{01}) \mid H_0$$
$$= \sum_{k=0}^{2}\left\{(\pi_{k2}^{(1)} - \pi_{k2}^{(0)} + x(\pi_{k1}^{(1)} - \pi_{k1}^{(0)})\right\}p'_k,$$

which is generally not 0 if the misclassification rates are differential (that is, $\pi_{ij}^{(1)} \neq \pi_{ij}^{(0)}$ for some i, j), and

$$\tau_x^2 = \mathrm{var}_{H_0}(\hat{\Delta}_x) = \frac{1}{n_1}\left\{\sum_{k=0}^{2}\left(\pi_{k2}^{(1)} + x^2\pi_{k1}^{(1)}\right)p'_k - \left[\sum_{k=0}^{2}\left(\pi_{k2}^{(1)} + x\pi_{k1}^{(1)}\right)p'_k\right]^2\right\}$$
$$+ \frac{1}{n_0}\left\{\sum_{k=0}^{2}\left(\pi_{k2}^{(0)} + x^2\pi_{k1}^{(0)}\right)p'_k - \left[\sum_{k=0}^{2}\left(\pi_{k2}^{(0)} + x\pi_{k1}^{(0)}\right)p'_k\right]^2\right\}$$

Then $Z_x \sim N(\delta_x, \lambda_x)$, $Z_x^2 \sim \lambda_x\chi_1^2(\delta_x^2)$, where $\lambda_x = \tau_x^2/\sigma_x^2$ is the variance inflation factor and $\delta_x^2 = \mu_x^2/\sigma_x^2$ is the noncentrality parameter.

With the Hardy–Weinberg equilibrium under the null, the bias term μ_x is a quadratic function of MAF p, and τ^2 is a quartic function of p.

APPENDIX B

We estimate allele frequency by using controls only, namely $\hat{p} = \hat{p}_{02} + \hat{p}_{01}/2$, then we can see this estimate is approximately unbiased

$$E(\hat{p}) = \left\{\pi_{02}^{(0)} + \pi_{01}^{(0)}/2\right\} + \left\{-2\pi_{02}^{(0)} - \pi_{01}^{(0)} + 2\pi_{12}^{(0)} + \pi_{11}^{(0)}\right\}p$$
$$+ \left\{\pi_{02}^{(0)} + \pi_{01}^{(0)}/2 + \pi_{22}^{(0)} + \pi_{21}^{(0)}/2 - 2\pi_{12}^{(0)} - \pi_{11}^{(0)}\right\}p^2$$
$$\approx p^2$$

as long as the misclassification rates $\pi_{jk}^{(0)}$, $j \neq k$, are small and $\pi_{jj}^{(0)} \approx 1$.