Genetic Epidemiology 30: 277–289 (2006)

Centralizing the Non-Central Chi-Square: A New Method to Correct for Population Stratification in Genetic Case-Control Association Studies

Prakash Gorroochurn,[1]* Gary A. Heiman,[2] Susan E. Hodge,[1,3] and David A. Greenberg[1,3]

[1]*Division of Statistical Genetics, Department of Biostatistics, Mailman School of Public Health, Columbia University*
[2]*Department of Epidemiology, Mailman School of Public Health, Columbia University*
[3]*Clinical-Genetic Epidemiology Unit, New York State Psychiatric Institute, New York, NY*

We present a new method, the δ-centralization (DC) method, to correct for population stratification (PS) in case-control association studies. DC works well even when there is a lot of confounding due to PS. The latter causes overdispersion in the usual chi-square statistics which then have non-central chi-square distributions. Other methods approach the non-centrality indirectly, but we deal with it directly, by estimating the non-centrality parameter τ itself. Specifically: (1) We define a quantity δ, a function of the relevant subpopulation parameters. We show that, for relatively large samples, δ exactly predicts the elevation of the false positive rate due to PS, when there is no true association between marker genotype and disease. (This quantity δ is quite different from Wright's F_{ST} and can be large even when F_{ST} is small.) (2) We show how to estimate δ, using a panel of unlinked "neutral" loci. (3) We then show that δ^2 corresponds to τ the non-centrality parameter of the chi-square distribution. Thus, we can centralize the chi-square using our estimate of δ; this is the DC method. (4) We demonstrate, via computer simulations, that DC works well with as few as 25–30 unlinked markers, where the markers are chosen to have allele frequencies reasonably close (within $\pm.1$) to those at the test locus. (5) We compare DC with genomic control and show that where as the latter becomes overconservative when there is considerable confounding due to PS (i.e. when δ is large), DC performs well for all values of δ. *Genet. Epidemiol.* 30:277–289, 2006. © 2006 Wiley-Liss, Inc.

Key words: genomic control; structured association; non-centrality parameter; false positive rate; admixture

Contract grant sponsor: NIH; Contract grant numbers: R01 AA013654; DK-31813; NS27941; DK31775.
*Correspondence to: P. Gorroochurn, Department of Biostatistics, Columbia University, R620 (6th floor), 722 W 168th Street, New York, NY 10032. E-mail: pg2113@columbia.edu
Received 30 September 2005; Accepted 7 January 2006
Published online 24 February 2006 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/gepi.20143

INTRODUCTION

Recently, several authors have investigated the use of genomic information in an attempt to eliminate confounding due to population stratification (PS) in case-control association studies [Devlin and Roeder, 1999; Devlin et al., 2001; Bacanu et al., 2000; Pritchard et al., 2000a,b; Satten et al., 2001; Reich and Goldstein, 2001]. Two general methods have been proposed: structured association (SA) [Pritchard et al., 2000a,b; Satten et al., 2001] and genomic control (GC) [Devlin and Roeder, 1999; Devlin et al., 2001].

STRUCTURED ASSOCIATION

In SA [Pritchard et al., 2000a,b], the first part of the method consists of a Bayesian clustering program. This is used to infer the number of subpopulations (K) within the sample and the subpopulation membership for each individual, using unlinked or "neutral" markers (i.e. markers not associated with the disease) to identify subpopulations. This first part is performed by the program STRUCTURE. In the second part of the method, the inferred membership of each individual in the ith subpopulation is used to perform a test of association for that subpopulation. This is done by the program STRATA. A modification of the method of Pritchard et al. [2000a,b] has been implemented by Satten et al. [2001], who use latent class models, instead of Bayesian clustering, to infer subpopulation membership. Simulations have shown that SA performs well when the subpopulations are discrete.

GENOMIC CONTROL

Confounding due to PS causes both overdispersion (i.e. excess variance) and bias in the usual chi-square test statistic (X^2_{test}, see Eq. (18) below) for the null hypothesis of no (true) association between disease and marker. Devlin et al. [2001] have argued that the effect of overdispersion is more important than that of bias. Confounding due to PS leads to elevated false positive rates (FPRs) [see Heiman et al., 2004; Gorroochurn et al., 2004]. GC uses a panel of unlinked markers from the same individuals to scale X^2_{test} down. One advantage of GC is that it successfully deals with admixture.

At a more technical level, in the presence of PS, X^2_{test} has a non-central chi-square distribution, $\chi^2_1(\tau)$, with non-centrality parameter τ and one degree of freedom (d.f.). Devlin et al. [2001, p. 161] assume a hierarchical model in which the non-centrality parameter itself has a normal distribution across loci. By appealing to the Central Limit Theorem, they argue that the expectation of the non-centrality parameter is zero when the number of subpopulations is large. Based on that assumption, finally they argue that the marginal distribution of X^2_{test} is a multiple of the central chi-square distribution. Thus, GC attempts to find a λ such that, approximately,

$$\frac{X^2_{\text{test}}}{\lambda} \sim \chi^2_1(0) \tag{1}$$

where $\chi^2_1(0)$ is the central chi-square distribution with 1 d.f. The factor λ allows for the excess variance in X^2_{test} and is called the variance inflation factor. Devlin and Roeder [1999, Eq. (5)] give a theoretically derived expression for λ. The appeal of GC lies in the use of the above-mentioned unlinked markers to obtain estimates of λ. Devlin and Roeder [1999] propose the robust estimator

$$\hat{\lambda}_1 = \frac{\text{median}\left\{X^2_{(1)}, X^2_{(2)}, \ldots, X^2_{(L)}\right\}}{.456} \tag{2}$$

where $X^2_{(1)}, X^2_{(2)}, \ldots, X^2_{(L)}$ denote the calculated values of the chi-square statistic at the unlinked loci 1,2,…,L, and .456 is the median of the $X^2_1(0)$ distribution. On the other hand, Reich and Goldstein [2001] use the sample mean:

$$\hat{\lambda}_2 = \frac{1}{L}\sum_{i=1}^{L} X^2_{(i)}. \tag{3}$$

An important assumption made by GC is that λ is constant across the unlinked loci. Devlin and Roeder [1999] point out that this condition implies that Wright's F_{ST} (1951), which measures the degree of genetic differentiation between subpopulations, should be approximately constant across loci.

Before moving on any further, let us mention that the model we will use differs from that of Devlin et al. [2001] in that we will treat the non-centrality parameter τ solely as a parameter (which can be estimated from the neutral loci), instead of assigning a distribution to it. Neither model is superior, though they are based on different assumptions.

A major limitation of GC lies in the approximation made in Eq. (1), the accuracy of which depends heavily on the magnitude of the non-centrality parameter τ [Chen et al., 2003; see also Reich and Goldstein, 2001]. As Marchini et al. [2004b] have warned, dividing the inflated chi-square by a constant may not always produce a central chi-square. We can prove this solely by using distribution theory: recall that, by equating the first two cumulants of a $\chi^2_{v^*}(\tau)$ distribution to those of a $\chi^2_v(0)$ distribution, $\chi^2_{v^*}(\tau)/\rho$ has an approximate $\chi^2_v(0)$ distribution, where

$$v = v^* + \frac{\tau^2}{v^* + 2\tau} \tag{4}$$

$$\rho = 1 + \frac{\tau}{v^* + \tau} \tag{5}$$

[Stuart et al., 1999, pp. 243–244]. GC uses $v = v^* = 1$ [Devlin et al., 2001], and from Eq. (4), we see that the approximation is good as long as $\tau \approx 0$, but not otherwise [see also Chen et al., 2003]. Devlin et al. [2001, p. 161] try to circumvent the assumption $\tau \approx 0$ (the non-centrality parameter τ here not to be confused with Devlin et al.'s τ, which represents variance inflation due to correlated alleles) by assigning a normal distribution to τ with mean zero. However, for a reasonable number of subpopulations and especially in the presence of high levels of PS, it is doubtful that τ will be centered on zero.

Since dividing by a constant does not necessarily work well, we decided to estimate the non-centrality parameter τ itself, and develop a method that does not depend on the approximation in (1) and therefore does not require the assumption $\tau \approx 0$. We will also show in Section "Discussion" that when τ is large, GC becomes very conservative [see also Shmulewitz et al., 2004]. (As we shall see in Section "The Quantity δ" below, τ is a measure of the extent of confounding

due to PS, assuming no true association between marker genotype and disease.)

Another issue with GC is its dependence on the number of unlinked loci (L) that need to be genotyped in the cases and controls. In a genome-wide study, Marchini et al. [2004a] used very small values for the test size α ($\sim 10^{-4}$–10^{-8}) and tested the performance of GC. Their results suggest that using too few markers ($L < 100$) leads to undercorrection (anti-conservativeness) and using too many ($L \geq 500$) leads to overcorrection (conservativeness).

In this paper, we shall develop a method (which we call δ-centralization, DC) to correct for PS by centralizing X^2_{test}. The validity of our method does not depend on the approximation in Eq. (1), and is therefore independent of the magnitude of τ, and hence of the level of stratification. We shall see that it is possible to centralize any $\chi^2_1(\tau)$ distribution to a $\chi^2_1(0)$ by relying only on asymptotic arguments rather than on the magnitude of the non-centrality parameter τ as well. Our DC method makes use of a quantity δ, related to the non-centrality parameter, which is of prime importance in case-control association studies. Not only does δ completely determine the extent of confounding (as measured by the elevation in the FPR) due to stratification in case-control studies when there is no true association between disease and marker genotype, but δ is also critical to our method to centralize X^2_{test}.

THE QUANTITY δ

We use the same notation as in Gorroochurn et al. [2004]. Consider a sample of m cases and n controls from a population consisting of K subpopulations of relative sizes π_i (where K and π_i are unknown) such that

$$\sum_{i=1}^{K} \pi_i = 1. \tag{6}$$

We define the following events for $i = 1,2,\ldots,K$; C_i: a person belongs to subpopulation i; D: a person has the disease; M: a person has the marker genotype (i.e. a genotype containing one or two marker alleles). Within the ith subpopulation, we denote the (unknown) disease prevalence by d_i and the (unknown) marker genotype frequency by r_i. Then

$$\begin{aligned} &\pi_i \equiv \Pr\{C_i\},\ \ d_i \equiv \Pr\{D|C_i\},\ \ r_i \equiv \Pr\{M|C_i\}, \\ &i = 1, 2, \ldots, K. \end{aligned} \tag{7}$$

Let the probabilities of marker genotype in diseased and non-diseased individuals in the total population be s and t, respectively, i.e.

$$s \equiv \Pr\{M|D\}, \quad t \equiv \Pr\{M|\bar{D}\}. \tag{8}$$

If, in each subpopulation, we further assume that disease is truly independent of marker genotype, i.e. $\Pr\{D \cap M|C_i\} = \Pr\{D|C_i\}\Pr\{M|C_i\} = d_i r_i$, then, by using Bayes' Theorem,

$$s = \frac{\sum_{i=1}^{K} \pi_i d_i r_i}{\sum_{i=1}^{K} \pi_i d_i}, \quad t = \frac{\sum_{i=1}^{K} \pi_i (1 - d_i) r_i}{1 - \sum_{i=1}^{K} \pi_i d_i}. \tag{9}$$

Under this assumption of independence within the subpopulations, any perceived association between disease and marker genotype is spurious (i.e. a consequence of PS).

In general, a test of *any* association (whether true or spurious) between disease and marker genotype in the total population is really a test of the hypothesis:

$$H_0 : s = t \quad \text{versus} \quad H_a: s \neq t. \tag{10}$$

We now define the quantity δ as follows:

$$\delta \equiv \frac{s - t}{\sqrt{\frac{s(1-s)}{m} + \frac{t(1-t)}{n}}} \tag{11}$$

(where m and n are the number of cases and controls, respectively; we will see how δ naturally arises in Eq. (14) later) and show that, for reasonably large sample sizes, δ completely determines the inflation in the (FPR) when there is no true association, i.e. as a result of confounding due to PS. We have

$$\text{FPR} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} I\big(p \leq \alpha\big) \tag{12}$$

where p is the P-value of a test of size α for the null hypothesis in (10), N is the total number of simulations performed, and I is the indicator variable defined by

$$I(p \leq \alpha) \equiv \begin{cases} 1 & \text{if } p \leq \alpha \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

We use the usual Z statistic (Z_{test}) to test for the two proportions in (10). Then, in the presence of PS (i.e. under H_a in (10)), Z_{test} will be inflated just as X^2_{test} is. In particular, its asymptotic distribution is

$$Z_{\text{test}} = \frac{(a_0/m - c_0/n)}{\sqrt{\frac{a_0/m(1-a_0/m)}{m} + \frac{c_0/n(1-c_0/n)}{n}}} \sim N(\delta, 1) \tag{14}$$

where δ is defined in Eq. (11), and a_0 and c_0 are the number of cases and controls with

the marker genotype, respectively. Then Eq. (12) becomes

$$\begin{aligned}\text{FPR} &= \text{E}\, I(|Z_{\text{test}}| \geq z_{\alpha/2}) \\ &= \int_{z=-\infty}^{\infty} I(|z| \geq z_{\alpha/2}) \frac{1}{\sqrt{2\pi}} \text{e}^{-(z-\delta)^2/2}\, \text{d}z\end{aligned}$$

where the area under the standard normal distribution to the right of $z_{\alpha/2}$ is $\alpha/2$. Simplifying the above equation, we get

$$\text{FPR} = 1 - \Phi(z_{a/2} - \delta) + \Phi(-z_{a/2} - \delta) \qquad (15)$$

where

$$\Phi(u) \equiv \frac{1}{\sqrt{2\pi}} \int_{y=-\infty}^{u} \text{e}^{-y^2/2}\, \text{d}y$$

is the standard normal distribution function. Previously [Gorroochurn et al., 2004], we obtained an exact expression for FPR based on two binomial distributions. In that instance, FPR was shown to be a function of the various π_i,d_i,r_i, and of m and n, but here we have shown that, for reasonably large sample sizes, FPR actually depends on all those quantities only through δ. We have also used related forms of δ in Heiman et al. [2004].

Note that, when (10) is viewed as a test for any association, then Eq. (15) gives the Type I error of the test; on the other hand, when no true association is assumed and (10) is viewed as a test for no spurious association, then Eq. (15) gives the power of the test [see Gorroochurn et al., 2004]. Figure 1 shows a graph of FPR as a function of δ for $\alpha = .01$ and .01.

From Figure 1, we see that a δ of magnitude $\approx .6$ doubles FPR from .05 and a δ of magnitude ≈ 1 more than triples it. The graph for $\alpha = .05$ is also above that for $\alpha = .01$ because of the higher power of the test.

From Eq. (11), we see that, in fact, the extent of confounding due to PS (as determined by the value of δ and the elevation in FPR) increases not only with the amount of difference in genotype (hence allele) frequencies between the subpopulations (which measures the extent of PS), but also with the sample size of the study (m and n), the relative sizes of the subpopulations, and the amount of difference in disease prevalences among the subpopulations. Thus it is possible to have a reasonably low level of PS (i.e. a small F_{ST}), and yet a reasonably large amount of *confounding* due to PS (i.e. a large δ) (see Table 1 for some typical values). Table 1 also illustrates how δ increases with increasing sample size (m and n) and δ is maximized when the subpopulations are of the same size. In Section "*Centralizing Non-Central chi-square distributions for* 2×2 *Tables*" below, we show that δ is also the (signed) square root of the non-centrality parameter τ. Note that if we were doing an association study using marker

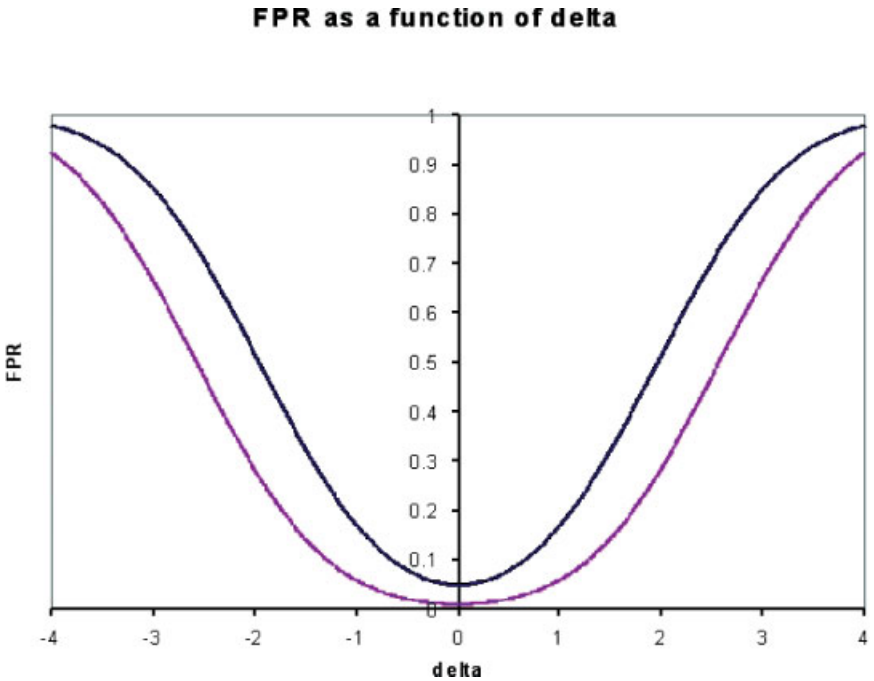

Fig. 1. Variation of uncorrected false positive rate (FPR) as a function of δ (see Eq. (15)) for $\alpha = .05$ (top curve) and $\alpha = .01$ (bottom curve).

TABLE 1. Comparison of some values of F_{ST} and δ for $K = 2$ subpopulations

(a) Values of δ for varying values of d and r when π = [.5, .5] and $m = n = 100$

	r = [.10, .15], F_{ST} = .003	r = [.10, .20], F_{ST} = .010
d = [.01, .05]	−.195	−.365
d = [.10, .10]	(0)	(0)
d = [.10, .15]	.121	.224
d = [.10, .20]	.207	.381
d = [.10, .50]	.501	.922
d = [.10, .75]	.706	1.305

(b) Values of δ for varying values of π, m, and n when d = [.10, .15] and r = [.10, .15]

	π = [.5, .5], F_{ST} = .003	π = [.25, .75], F_{ST} = .002
$m = n = 100$	.121	.104
$M = n = 150$	.148	.128
$M = n = 200$	.172	.148

To calculate F_{ST}, we calculated the Hardy-Weinberg (HW) allele frequency corresponding to each genotype frequency, then used $F_{ST} = 1\text{-}H_S/H_T$, where H_S = weighted average of heterozygosities expected under HW in the subpopulations, H_T = heterozygosity expected under HW in the total population.

alleles (instead of marker genotypes), then δ would be defined in the same way as in Eq. (11), except that r_i would then be the frequency of the marker allele in the ith subpopulation, s and t would be the probabilities of the marker allele in, respectively, diseased and non-diseased individuals in the population, and m and n would be twice the number of cases and controls, respectively.

Thus, if we knew the value of δ, Eq. (15) would enable us to know exactly how much the FPR is elevated due to PS, assuming no true association. This leads to the question: "Can we estimate δ from a case-control association study?" The answer is fortunately yes. We can estimate δ by using the unlinked marker loci. Ideally (i.e. for maximum efficiency), we would choose all unlinked marker loci whose genotype (hence allele) frequencies matched those at the test locus in the controls. However, this condition is not necessary for the method to work. Suppose that the 2×2 tables at each of L such unlinked loci are as shown in Table 2(a).

Then, natural estimators of s and t are

$$\hat{s} = \frac{1}{mL}\sum_{i=1}^{L} a_i, \quad \hat{t} = \frac{1}{nL}\sum_{i=1}^{L} c_i \tag{16}$$

TABLE 2. 2×2 tables

(a) Unlinked loci

locus 1				locus 2				…	locus L			
	M	$\overline{\text{M}}$			M	$\overline{\text{M}}$				M	$\overline{\text{M}}$	
D	a_1	b_1	m	D	a_2	b_2	m	…	D	a_L	b_L	m
$\overline{\text{D}}$	c_1	d_1	n	$\overline{\text{D}}$	c_2	d_2	n		$\overline{\text{D}}$	c_L	d_L	n

(b) Test locus

	M	$\overline{\text{M}}$	total
D	a_0	b_0	m
$\overline{\text{D}}$	c_0	d_0	n

where $m = a_i + b_i$ and $n = c_i + d_i$ for $i = 1,2,\ldots,L$. Hence, δ can be estimated by

$$\hat{\delta} = \frac{\hat{s} - \hat{t}}{\sqrt{\frac{\hat{s}(1-\hat{s})}{m} + \frac{\hat{t}(1-\hat{t})}{n}}}. \tag{17}$$

If one matches marker loci exactly to the test locus, one will require only a relatively small number of marker loci L in order to estimate δ accurately. However, that approach might require scanning many unlinked loci before the suitable ones were found (although use of SNPs might reduce this difficulty). On the other hand, since the estimation in Eq. (17) is based on averages, an exact matching is by no means necessary: we could choose any marker loci within a reasonable genotype frequency window (of, say, $\pm$.1) to that at the test locus (see Sections "*Simulation Studies*" and "*Discussion*" for further discussions on this).

CENTRALIZING NON-CENTRAL CHI-SQUARE DISTRIBUTIONS FOR 2×2 TABLES

We now show how DC can be used to correct for PS. Let us assume the model in Section "*The Quantity* δ" above and, in particular, that there is no true association between disease and marker genotype in each of the constituent subpopulations. As we saw in Section "*Introduction*", when there is PS, the usual chi-square test of association between disease and marker genotype at a test locus is inflated and has a non-central chi-square distribution with non-centrality parameter τ and 1 d.f., i.e. approximately (see Table 2b),

$$X^2_{\text{test}} = \frac{(m+n)(a_0 d_0 - b_0 c_0)^2}{mn(a_0 + c_0)(b_0 + d_0)} \sim \chi^2_1(\tau). \tag{18}$$

Now, we have $X^2_{\text{test}} = Z^2_{\text{test}}$ almost exactly, where $Z_{\text{test}} \sim N(\delta,1)$ is given in Eq. (14). We see that the non-centrality parameter τ is equal to δ^2. Then $\text{sign}(a_0d_0 - b_0c_0) \times \sqrt{X^2_{\text{test}}} \sim N(\delta, 1)$. We define our δ-centralized test statistic, T_{DC} as

$$T_{\text{DC}} \equiv \left\{ \text{sign}(a_0d_0 - b_0c_0) \times \sqrt{X^2_{\text{test}}} - \delta \right\}^2 \sim \chi^2_1(0) \tag{19a}$$

where

$$\text{sign}(x) \equiv \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

Asymptotically, T_{DC} is valid for all values of δ^2, and hence valid whatever the extent of stratification. Of course, δ is unknown and should be estimated from the panel of unlinked markers as in Eq. (17). (Again, to estimate δ quite accurately, we do not need to exactly match the marker genotype frequencies; if the marker loci are chosen so that their genotype frequencies are within, say, a $\pm$.1 window of that at the test locus, then δ can be estimated quite accurately). For any 2×2 table in which there is potentially a component of true association between marker genotype and disease, and potentially a component of false association due to PS, we therefore propose the following statistic to test for the true association while correcting for the false one:

$$T^*_{\text{DC}} = \left\{ \text{sign}(a_0d_0 - b_0c_0) \times \sqrt{X^2_{\text{test}}} - \hat{\delta} \right\}^2 \sim \chi^2_1(0) \quad \text{under } H_0: \text{no true association} \tag{19b}$$

where X^2_{test} is calculated as in Eq. (18). T^*_{DC} is thus the statistic that does the δ-centralization, based on the estimated (signed) square root of the non-centrality parameter. T^*_{DC} adjusts X^2_{test} so that the latter has both the proper mean and variance, i.e. it gets rid of both the bias and overdispersion caused by PS in X^2_{test}.

CENTRALIZING NON-CENTRAL CHI-SQUARE DISTRIBUTIONS FOR TREND TESTS IN 2×3 TABLES

Sasieni [1997] recommends the use of 2×2 tables (the so-called "serological tables") of disease status versus marker genotype when the marker allele is dominant, and 2×3 tables of disease status versus actual genotype when the marker allele is codominant. In the latter case, the table for the test locus looks as in Table 3.

In this case, the usual chi-square test for no association between disease and genotype is Armitage's trend test [1955], calculated under an additive genetic model:

$$X^2_{\text{trend}} \equiv \frac{N\{N(r_1 + 2r_2) - R(n_1 + 2n_2)\}^2}{R(N-R)\{N(n_1 + 4n_2) - (n_1 + 2n_2)^2\}}. \tag{20}$$

Again, in the presence of stratification, X^2_{trend} is inflated and has a non-central $X^2_1(\tau)$ distribution. To correct for PS, we need to obtain the non-centrality parameter $\tau = \delta^2_{\text{trend}}$. To calculate this, note that Eq. (20) can be written as

$$X^2_{\text{trend}} = \frac{NR\left\{\left(\frac{r_1+2r_2}{R}\right) - \left(\frac{n_1+2n_2}{N}\right)\right\}^2}{(N-R)\left\{\left(\frac{n_1+4n_2}{N}\right) - \left(\frac{n_1+2n_2}{N}\right)^2\right\}}.$$

This shows that the non-centrality parameter is

$$\delta^2_{\text{trend}} \equiv \frac{NR\{(P_1 + 2P_2) - (Q_1 + 2Q_2)\}^2}{(N-R)\{(Q_1 + 4Q_2) - (Q_1 + 2Q_2)^2\}} \tag{21}$$

where P_1,P_2 are the proportions of *Aa*,*AA* genotypes within diseased individuals, respectively, and Q_1,Q_2 are the proportions of these genotypes in the total population, respectively. We again use the unlinked markers 1,2,…,L to estimate δ_{trend}, because the $P_i's$ and $Q_i's$ can be estimated from the 2×3 table at each locus. In particular, if the values of P_i and Q_i ($i = 1,2$) for the jth table are written as $\hat{P}_i^{(j)}$ and $\hat{Q}_i^{(j)}$, then

$$\hat{P}_i = \frac{1}{L}\sum_{j=1}^{L} \hat{P}_i^{(j)}, \quad \widehat{Q}_i = \frac{1}{L}\sum_{j=1}^{L} \hat{Q}_i^{(j)} \quad i = 1,2. \tag{22}$$

Thus,

$$\hat{\delta}_{\text{trend}} = \frac{(\hat{P}_1 + 2\hat{P}_2) - (\hat{Q}_1 + 2\hat{Q}_2)}{\sqrt{\left(\frac{1}{R} - \frac{1}{N}\right)\{(\hat{Q}_1 + 4\hat{Q}_2) - (\hat{Q}_1 + 2\hat{Q}_2)\}^2}}. \tag{23}$$

TABLE 3. The 2×3 table for the trend test

	AA	Aa	aa	Total
D	r_2	r_1	$R-r_1-r_2$	R
$\overline{\text{D}}$	n_2-r_2	n_1-r_1	$N-R-n_1-n_2+r_1+r_2$	$N-R$
Total	n_2	n_1	$N-n_1-n_2$	N

Therefore, the statistic that tests for the true association while correcting for the false one is

$$T^*_{\mathrm{DC,trend}} \equiv \Big[\mathrm{sign}\{N(r_1+2r_2)-R(n_1+2n_2)\} \times\sqrt{X^2_{\mathrm{trend}}-\hat{\delta}_{\mathrm{trend}}}\Big]^2 \sim \chi^2_1(0) \quad \text{under}\, H_0 \tag{24}$$

since the denominator in Eq. (20) is always positive.

SIMULATION STUDIES

IMPLEMENTATION OF DC

The usual method for generating loci has been based on the beta-binomial distribution [see for example, Devlin and Roeder, 1999; Marchini et al., 2004]. Since the DC method depends on frequency matching, instead of using the beta-binomial for the loci and then selecting those loci that match, we shall simply use the uniform distribution for the distribution of those loci that match the candidate locus. By using the uniform distribution, therefore we are not assuming that this is the distribution of loci, but rather this is the approximate distribution of those loci that closely match the candidate locus. (Our initial simulations using the beta-binomial and then matching gives us essentially the same FPRs as when using a uniform generating mechanism for the matching loci. We also initially tried a beta-binomial generating mechanism, then selected matching loci for DC but used all loci for GC; this also resulted in essentially the same FPRs (results not shown)).

Table 4 shows five sets of simulations for two subpopulations. For each set, values of $\pi^*_1, \pi^*_2, r_1, r_2, d_1$ and d_2 were chosen randomly from a $U(0,1)$ distribution (we assumed no true association between marker and disease within each subpopulation). We then standardized π^*_1, π^*_2 to π_1, π_2 so that $\pi_1 + \pi_2 = 1$. Using the value of δ obtained, we calculated the inflated FPR as given in Eq. (15). To estimate the corrected FPR (cFPR (δ)) for T_{DC} in Eq. (19a), we note that if X and Y denote the number of individuals with the marker genotype in the cases and controls, respectively, then $X \sim$Bin (m,s) and $Y \sim$Bin (n,t), where we used $m = n = 100$. By using these distributions, we generated 5,000 2×2 tables. We then applied the test statistic T_{DC} in Eq. (19a) to each of these tables and were thus able to estimate cFPR (δ), using a test of size $\alpha = .05$. To estimate the corrected FPR (cFPR ($\hat{\delta}$)) for T^*_{DC} in Eq. (19b), we simulated 5,000 sets of 25 2×2 tables (one for the test locus, one for each of the 24 unlinked loci used). For each one of the 5,000 simulations, a different $\hat{\delta}$ was estimated from the unlinked loci. (The same binomial distribution was used to generate the 2×2 tables at each of the unlinked loci as that at the test locus, i.e. exact matching of marker genotype frequencies at the unlinked loci was assumed.) Table 5 shows the same, but this time with three subpopulations, $\pi^*_i \sim U(0,1)$, $r_i \sim U(.4,.7)$ and $d_i \sim U(.0,.1)$ for $i = 1,2,3$.

As these two tables show, the DC-corrected FPRs based on the true value of δ (i.e. cFPR(δ)) were quite close to the nominal $\alpha = .05$, and the ones based on the estimated δ (i.e. cFPR($\hat{\delta}$)) also performed quite well.

HOW MANY UNLINKED MARKERS DOES DC NEED?

It is especially important for practical and cost-related reasons to have an idea of the number of unlinked markers required for our method to work reasonably well. We expect good performance as long as $\hat{\delta}$ estimates δ with little bias. As long as the frequencies of the markers at the

TABLE 4. Uncorrected false postive rates (FPR), δ-corrected false positive rates, δ-corrected false positive rates for five values of δ

π	*d*	*r*	δ	FPR[a]	cFPR(δ)[b]	cFPR ($\hat{\delta}$)[c] ($\pm$1.96SE)
[.571, .429]	[.344, .474]	[.558, .747]	.370	.066	.052	.054 $\pm$ .006 = (.048, .006)
[.507, .493]	[.544, .169]	[.500, .355]	.847	.135	.051	.052 $\pm$ .006 = (.046, .058)
[.414, .586]	[.160, .182]	[.893, .624]	−.165	.053	.043	.052 $\pm$ .006 = (.046, .058)
[.212, .788]	[.444, .311]	[.111, .731]	−.887	.144	.048	.050 $\pm$ .006 = (.044, .056)
[.633, .367]	[.266, .147]	[.937, .398]	1.451	.306	.050	.054 $\pm$ .006 = (.048, .006)

Values of π_i, d_i, r_i were all chosen from a U (0, 1) distribution. δ was estimated from a panel of 24 unlinked markers. ($\alpha = .05$, $m = n = 100$, $K = 2$ subpopulations).

[a]FPR = uncorrected false positive rate; see Eq. (15).

[b]cFPR(δ) = δ-corrected false positive rate using computer simulations; see Eq. (19a).

[c]cFPR($\hat{\delta}$) δ-corrected false positive rate using computer simulations; see Eq. (19b).

unlinked loci are reasonably similar, this bias should be small for a reasonably small number L of unlinked loci.

To verify the above claim, we chose 10 combinations of values π_i^*, r_i and d_i $(i = 1,2)$ at random from U (0,1), U (.1,.6) and U (.02,.2) distributions, respectively. Each combination represents parameter values at the test locus. For a given combination, the marker genotype frequency was chosen from a U $(r_i-.1,\ r_i+.1)$ distribution for each marker locus (i.e. the genotype frequency of each marker locus was matched to within $\pm.1$ of that at the test locus). We then calculated cFPR $(\hat{\delta})$ by estimating δ from Eq. (17) by using $L = 10$, 20, and 30 unlinked markers (thus, in all, $10 \times 3 = 30$ combinations were used). The results are displayed in Fig. 2 (using SPSS 12.0 for Windows), and in Table 6. These show that as

TABLE 5. Uncorrected false positive rates (FPR), δ-corrected false positive rates, δ-corrected false positive rates for five values of δ

π	d	r	δ	FPR[a]	cFPR(δ)[b]	cFPR (δ)[c] ($\pm$1.96SE)
[.375, .436, .189]	[.071, .061, .094]	[.542, .652, .454]	−.182	.054	.051	.060 $\pm$ .006 = (.054,.066)
[.123, .405, .472]	[.019, .036, .068]	[.455, .497, .694]	.595	.091	.049	.055 $\pm$ .006 = (.049,.061)
[.262, .313, .424]	[.041, .056, .074]	[.659, .663, .545]	−.180	.054	.049	.056 $\pm$ .006 = (.050,.062)
[.066, .043, .891]	[.068, .048, .098]	[.434, .623, .537]	.001	.050	.054	.054 $\pm$ .006 = (.048,.060)
[.241, .325, .434]	[.010, .006, .061]	[.689, .629, .538]	−.718	.111	.049	.051 $\pm$ .006 = (.045,.057)

Values of π_i, d_i, r_i were chosen from a $U(0, 1)$, $U(0, .1)$ and $U(.4, .7)$ distribution, respectively. δ was estimated from a panel of 24 unlinked markers. ($\alpha = .05$, $m = n = 100$, $K = 3$ subpopulations).
[a]FPR = uncorrected false positive rate; see Eq. (15).
[b]cFPR(δ) = δ-corrected false positive rate using computer simulations; see Eq. (19a).
[c]cFPR(δ) = δ-corrected false positive rate using computer simulations; see Eq. (19b).

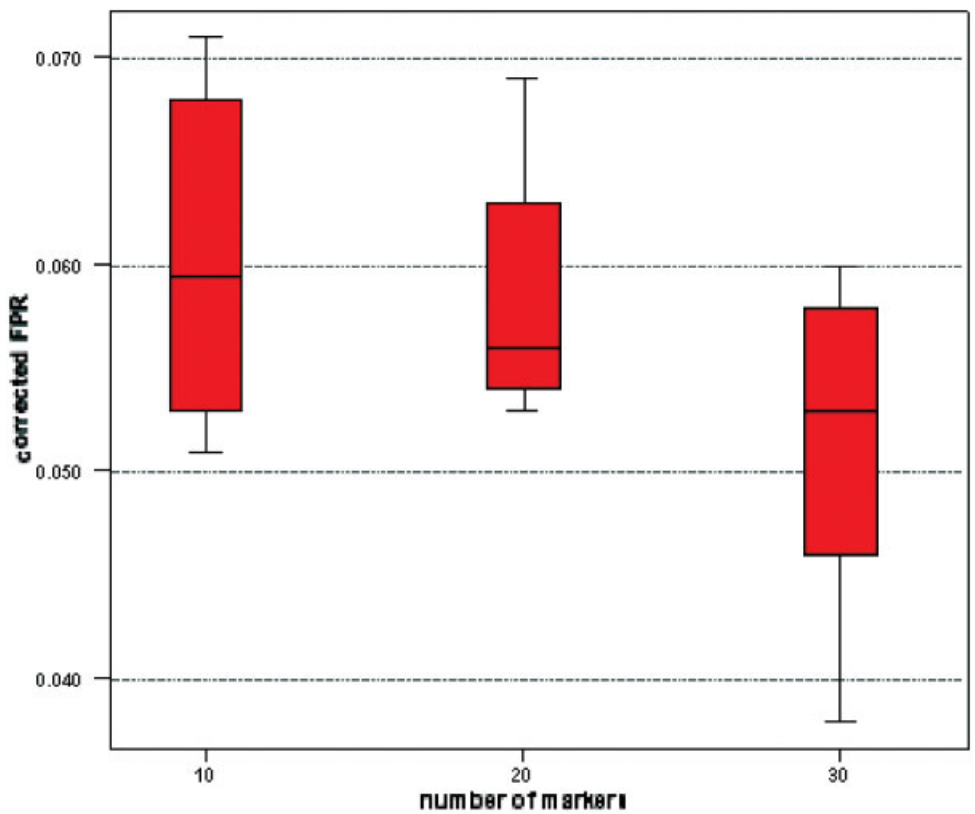

Fig. 2. Ten combination of values π_I, r_i, and d_i were chosen from a $U(0,1)$ distribution. For each combination, cFPR $(\hat{\delta})$ was calculated as described in Section "Simulation Studies" from simulations by successively using L = 10, 20 and 30 unlinked markers to estimate δ (These were markers with allele frequencies within $\pm$.1 of the allele frequencies at the test locus.) The ten values of cFPR$(\hat{\delta})$ were then plotted for each value of L ($\alpha = .05$, $m = n = 100$, $K = 2$ subpopulations).

TABLE 6. Summary statistics for simulations in Figure 2

No. of markers(L)	Mean cFPR($\hat{\delta}$)	5% trimmed mean	SE of mean
10	.060	.060	.008
20	.058	.057	.006
30	.052	.052	.007

few as about 30 unlinked markers are sufficient for cFPR ($\hat{\delta}$) to be very close to the nominal value of $\alpha = .05$, when the marker genotype frequencies are within $\pm .1$ of that at the test locus. Of course, the greater the number of unlinked markers, the closer will $\hat{\delta}$ be to δ and the better the correction.

COMPARISON WITH GENOMIC CONTROL

In Table 7, we look at two matching scenarios in which the matching of marker genotype frequencies at the unlinked loci to that at the test locus is, in turns, exact and to within $\pm .1$ (we also matched to within $\pm .15$ and the results were very similar to $\pm .1$ case). We show simulations for different values of δ ($\delta = .1, .5, 1, 2$ and 4 approximately) and $L = 25$. We compare the four correction methods: GC using the median (see Eq. (2)), GC using the mean (see Eq. (3)), DC using T^*_{DC} (see Eq. (19b)), and DC using T^*_{DC} where, however, $\hat{s}$ and $\hat{t}$ in Eq. (16) are the 5% trimmed means instead of the regular means (i.e. 5% of the lowest and 5% of the highest values of the $a_i's$ and $c_i's$ are discarded; the corresponding estimated δ is denoted by $\hat{\delta}_{tr}$).

Table 8 is similar to Table 7 but looks at exact matching for four subpopulations with different values of δ ($\delta = .1, .5, 1, 2$ and 4 approximately) and different values of L ($L = 25, 50, 100, 300$). We do not expect the results to be very different if a matching to within $\pm .10$ was done, as for Table 7 above. Both tables show that GC using the median is quite anti-conservative for δ less than .5, but becomes very conservative for δ close to 1 and overly conservative for δ greater than about 2. GC using the mean performs better than GC using the median for δ less than 1.0, but is also excessively conservative for δ greater than about 2. (These results are in accordance with those of Shmulewitz et al., 2004). On the other hand, DC performs quite well in all these situations, and even *very well* when there are very high levels of confounding due to PS (δ close to 4). Moreover, there does not seem to be any advantage in using the 5% trimmed means over the regular means.

TABLE 7. For each value of δ, the numbers are the various simulated FPRs under exact matching and matching to within $\pm .1$

Approx. δ		False positive rates under exact matching	False positive rates under matching to within $\pm .1$[a]
.1[b]	Uncorrected FPR	.051	.051
	FPR using GC with median	.073	.071
	FPR using GC with mean	.063	.063
	FPR using T^*_{DC} with $\hat{\delta}$	.058	.053
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.059	.053
.5[c]	Uncorrected FPR	.079	.079
	FPR using GC with median	.065	.061
	FPR using GC with mean	.057	.057
	FPR using T^*_{DC} with $\hat{\delta}$	.060	.053
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.060	.054
1.0[d]	Uncorrected FPR	.172	.172
	FPR using GC with median	.039	.038
	FPR using GC with mean	.052	.048
	FPR using T^*_{DC} with $\hat{\delta}$	.059	.051
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.059	.052
2.0[e]	Uncorrected FPR	.513	.513
	FPR using GC with median	.001	.001
	FPR using GC with mean	.010	.010
	FPR using T^*_{DC} with $\hat{\delta}$	.051	.051
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.054	.054
4.0[f]	Uncorrected FPR	.980	.980
	FPR using GC with median	.000	.000
	FPR using GC with mean	.000	.000
	FPR using T^*_{DC} with $\hat{\delta}$	.048	.044
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.048	.045

To obtain the values of π, d, r for a value of δ, we fixed all but one of the values of π_i, d_i, r_i, and then solved Eq. (11) for the unknown. $L = 25$, $\alpha = .05$, $m = n = 100$, $K = 2$ subpopulations.

[a]Except for $\delta = 4.0$,where we used $\pm .081$.

[b]$\delta = .112$, $\pi = [.453, .547]$, $d = [.219, .204]$, $r = [.679, .322]$.

[c]$\delta = .449$, $\pi = [.450, .550]$, $d = [.325, .176]$, $r = [.496, .323]$.

[d]$\delta = 1.006$, $\pi = [.541, .459]$, $d = [.116, .206]$, $r = [.226, .644]$.

[e]$\delta = 1.992$, $\pi = [.447, .553]$, $d = [.371, .149]$, $r = [.618, .154]$.

[f]$\delta = 4.005$, $\pi = [.609, .391]$, $d = [.501, .180]$, $r = [.919, .115]$.

DISCUSSION

In the presence of PS, case-control association studies between marker and disease can reveal evidence of association which might be an artifact of PS rather than a consequence of any true association. Mathematically, this is because PS causes the usual chi-square test statistic to become inflated. More precisely, the chi-square statistic no longer has a central distribution, but instead is non-centrally distributed.

GC attempts to correct the confounding due to PS by dividing the inflated chi–square by a scale factor so that it approximately has a central

distribution. However, the scaling works reasonably well only when the non-centrality parameter is small. Real studies may have moderate to high values of this parameter.

In this paper, we have developed a quantity (δ) (see Eq. (11)) which, in the presence of no true association between disease and marker genotype, completely determines the elevation in FPR due to confounding by PS in reasonably large samples (see Eq. (15)). Note that $|\delta|$ and FPR both increase when (a) the sample size of the case-control study increases, or (b) the subpopulations are of similar size, or (c) the differences between the genotype frequencies in the subpopulations increase (i.e. there is more PS), or (d) the differences between the disease prevalences in the subpopulations increase. Thus, it is possible to have low PS (small F_{ST}) and at the same time a relatively large amount of confounding due to PS (large δ). Our formula indicates that, with a test size of $\alpha = .05$, a value of $|\delta| = 1$ leads to an elevation of FPR to about 17%; a value of $|\delta| = 2$ leads to an elevation of FPR to about 50%. Moreover, δ can be estimated from a panel of unlinked markers (Eq. (17)) whose genotype (and hence allele) frequencies are within a reasonable window to that at the test locus. Finally, we have shown (see "*Centralizing Non-Central chi-square distributions for 2 × 2 Tables*") that δ is also the (signed) square root of the non-centrality parameter of the inflated chi-square statistic. Thus, rather than deal with the non-centrality issue indirectly, as GC does, we directly estimate the non-centrality parameter itself.

Other authors have used various "deltas" as measures of confounding due to PS [e.g. Pritchard and Rosenberg, 1999; Devlin et al., 2001; Heiman et al., 2004; Gorroochurn et al., 2004]. For example, Devlin et al. [2001] used $\delta = s - t$. In earlier work [Heiman et al., 2004], we used $\delta = |s - t| / \times \sqrt{2n/\{(s+t)(2-s-t)\}}$, where the number of cases and controls are both equal to n. However, our current δ in Eq. (11), because of its statistical properties, exactly predicts the extent of confounding in relatively large samples and is thus ideally suited to the task.

We have shown that (see "*Centralizing chi-square distributions for 2 × 2 Tables)* $\hat{\delta}$ can be used to correct for PS by centralizing the inflated chi-quare statistic through a procedure whose validity, unlike GC, does not depend on the magnitude of the non-centrality parameter (and therefore on δ). Thus, $\hat{\delta}$ partitions the true effects of a gene from the effects due to PS. Our DC method is very much in the spirit of GC, because it too uses a panel of unlinked markers to correct for PS. However, since it is based only on large-sample theory, it is asymptotically exact (and its validity does not depend on the magnitude of the non-centrality parameter). We therefore expect our method to be superior to GC especially with moderate to high values of δ. This is substantiated by our simulation results (see Tables 7 and 8). When δ is close to 1, GC using the median correction in Eq. (2) is quite conservative, while GC using the mean correction in Eq. (3) is only slightly conservative. With δ close to 2, both corrections become very conservative, and with δ close to 4, both almost never reject. In all these situations, DC performs quite well.

A second difference has to do with the issue of genotype (and allele) frequency matching. Although GC assumes strict matching, Reich and Goldstein [2001] report that, unless the degree of stratification is very large or the allele frequency at the test locus is very small, GC requires little or no matching. (However, as we demonstrated in Section "*The Quantity* δ", even when the amount of stratification is small, δ can be reasonably large and GC will become conservative). On the other hand, if DC is applied with no matching, then, whatever the underlying distribution of the genotypes at the unlinked marker loci in each subpopulation, it is very unlikely that $\hat{\delta}$ will be close to δ. For example, assuming a U (0,1) distribution for genotypes at the unlinked marker loci in each subpopulation, $\hat{\delta}$ will converge to zero as L increases, and our method will not correct for PS at all. Thus, GC has the advantage of robustness over DC with respect to genotype (or allele) frequency matching. DC will correct for any level of PS as long as the genotype frequencies are matched, if not exactly, at least approximately to within a reasonable window. On the other hand, GC remains inaccurate even if matched exactly, when δ is moderate to large (see Tables 7 and 8). For DC, we recommend the choice of marker loci whose genotype frequencies are within $a \pm .1$ window to that of the test locus. The narrower the window, the more accurate the estimator of δ will be; however, the more loci will have to be genotyped before the appropriate ones are found. We are currently testing various window sizes to determine the greatest window size while still controlling the Type I error and maintaining adequate power.

A third difference between GC and DC lies in the model used. For GC, Devlin et al. [2001]

TABLE 8. For each *L*-δ combination, the numbers are the simulated FPRs under exact matching

Approx. δ		$L = 25$	$L = 50$	$L = 100$	$L = 300$
.1[a]	Uncorrected FPR	.051	.051	.051	.051
	FPR using GC with median	.071	.059	.058	.055
	FPR using GC with mean	.059	.050	.053	.054
	FPR using T^*_{DC} with $\hat{\delta}$	.054	.052	.050	.054
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.053	.053	.049	.054
.5[b]	Uncorrected FPR	.080	.080	.080	.080
	FPR using GC with median	.065	.053	.050	.046
	FPR using GC with mean	.058	.050	.052	.044
	FPR using T^*_{DC} with $\hat{\delta}$	.055	.051	.046	.045
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.056	.051	.046	.045
1.0[c]	Uncorrected FPR	.168	.168	.168	.168
	FPR using GC with median	.034	.032	.026	.021
	FPR using GC with mean	.041	.048	.045	.034
	FPR using T^*_{DC} with $\hat{\delta}$	.051	.057	.053	.044
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.053	.057	.052	.044
2.0[d]	Uncorrected FPR	.514	.514	.514	.514
	FPR using GC with median	.001	.000	.000	.000
	FPR using GC with mean	.015	.010	.010	.000
	FPR using T^*_{DC} with $\hat{\delta}$	.049	.055	.055	.044
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.050	.056	.056	.044
4.0[e]	Uncorrected FPR	.979	.979	.979	.979
	FPR using GC with median	.000	.000	.000	.000
	FPR using GC with mean	.000	.000	.000	.000
	FPR using T^*_{DC} with $\hat{\delta}$	.046	.044	.041	.040
	FPR using T^*_{DC} with $\hat{\delta}_{tr}$	.046	.043	.044	.040

To obtain the values of π, d, r for a value of δ, we fixed all but one of the values of π_i, d_i, r_i, and then solved Eq. (11) for the unknown. $\alpha = .05$, $m = n = 100$, $K = 4$ subpopulations.
[a]$\delta = .102$, $\pi = [.135, .305, .215, .345]$, $d = [.212, .197, .260, .171]$, $r = [.333, .578, .314, .169]$.
[b]$\delta = .507$, $\pi = [.305, .260, .186, .249]$, $d = [.194, .205, .123, .158]$, $r = [.610, .493, .170, .186]$.
[c]$\delta = .991$, $\pi = [.336, .334, .175, .155]$, $d = [.391, .389, .199, .187]$, $r = [.632, .669, .215, .385]$.
[d]$\delta = 1.996$, $\pi = [.116, .082, .461, .341]$, $d = [.269, .271, .118, .367]$, $r = [.323, .254, .169, .686]$.
[e]$\delta = 3.990$, $\pi = [.100, .021, .021, .858]$, $d = [.412, .195, .146, .011]$, $r = [.530, .452, .683, .181]$.

assume that the non-centrality parameter τ itself has a distribution across loci. For DC, we treat τ solely as a parameter. This is why τ can be estimated in DC, whereas it makes no sense to estimate τ in GC.

Recently, Devlin et al. [2004] have proposed a modification of their GC method, which they call GCF, to handle very small test sizes α and uncertainty in the values of λ. Devlin et al.'s proposal was motivated by the anti-conservativeness of GC when α is small. The only difference between GC and GCF is that, whereas the former uses a cut-off based on a χ^2_1 distribution, the latter uses a slightly larger value, based on an $F_{1,L}$ distribution, where L is the number of neutral loci. This makes GCF a more conservative procedure than GC, i.e. GCF will *always* result is smaller FPRs than GC. Since GC is already very conservative for large value of δ (see Tables 7 and 8), applying GCF in those cases will yield even smaller rejection rates. Thus GCF cannot correct the deficiencies of GC when δ is large.

Regarding the lack of robustness of GC when the value of δ (and therefore the amount of confounding due to PS) is moderate to high, we should mention that Devlin and Roeder [1999, 2004] do point out the need for careful matching of cases and controls for ethnicity and other potential confounders before applying GC. They recommend matching not so as to make the approximation in Eq. (1) better, but rather to reduce heterogeneity in F_{ST} across the unlinked loci, so that the inflation factor λ is constant for each unlinked locus. This has pay-offs in terms of power. However, the extent to which such matching can decrease the value of δ remains to be seen (see comments above). With DC, there is no need for matching of cases and controls, as the method

will work, whatever the value of the non-centrality parameter (and the level of stratification). Moreover, the very fact that DC requires matching the genotype frequencies of the null loci within a suitable window to that of the candidate locus guarantees that F_{ST} remains (almost) constant.

Few authors have found the FPRs under GC to be off-target because none of them have investigated the performance of GC with respect to the magnitude of the non-centrality parameter (δ^2), although many have speculated that GC might not perform well if the magnitude of δ^2 is large. For example, Chen et al. [2003] point out (p. 250): "However, if δ (Note: Chen et al. use δ, not τ, to represent the non-centrality parameter) is large, using a χ^2 adjusted by a constant as an approximation of a non-central chi-square $\chi^2(\delta)$ will either lead to false positives or lose power". We have proved this on pp. 5–6 of our paper. Shmulewitz et al. [2004], using a different simulation system, did find GC to be off-target in certain situations. Although they were not aware of δ, they did find GC to be conservative "with increasing subpopulation prevalence and marker differences" (this corresponds to large values of delta).

As far as the number of unlinked markers required is concerned, DC requires a moderate number ($\sim$25–30) to perform reasonably well, because the correction is based on averages. As we mentioned above, if the markers are chosen such that their genotype frequencies are within a considerably narrow window to that at the test locus, we would expect a better performance of our method, at the expense of having to genotype a larger number of marker loci before the appropriate ones are found.

Will DC be necessary when K is large (say 10 or more)? For whatever value of K, if δ turns out to be large, GC will have the same problem of extreme conservativeness whilst DC will fare well. However, as K becomes large, it becomes less likely that δ will be large (unless there is a possible correlation between subpopulation size and marker prevalence due to a founder effect in which case δ could remain large: see Gorroochurn et al. [2004] and Heiman et al. [2004]).

In this paper, we have looked only at the problem of pointwise significance, i.e. testing association at a single particular test locus. Our method can also be used in a genome-wide scan, where several candidate loci are tested, and for each test locus, several marker loci are then used. Because of multiple testing, the genome-wide test size [Strachan and Read, 2004] is then very small (typically $\alpha \sim 10^{-4}$–10^{-8}). In such cases, we recommend using DC together with Benjamini and Hochberg's FDR procedure [1995].

The results of Marchini et al. [2004a] suggest that GC is very anti-conservative when α is small. Devlin et al. [2004] argue that this is because of significant variability in λ (see Eq. (1)). Since our T^*_{DC} statistic in Eq. (19b) is really a $\chi^2_1(0)$ statistic under the null hypothesis of no true association, elementary statistical theory would point to a decrease in power with decreasing α, and a more conservative test. However, we will investigate more rigorously how DC behaves in such situations in the future, and also more generally investigate the issue of power. Other aspects of DC that we plan to investigate in the future are its behavior under admixture and comparison with SA.

ACKNOWLEDGMENTS

We thank Dr. Bruce Levin for his invaluable advice and to Junying Zhang for computational help at the start of this work.

REFERENCES

Armitage P. 1955. Tests for linear trends in proportions and frequencies. Biometrics 11:375–386.

Bacanu SA, Devlin B, Roeder K. 2000. The power of genetic control. Am J Hum Genet 66:1933–1945.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300.

Chen HS, Zhu X, Zhao H, Zhang S. 2003. Qualitative semiparametric test for genetic associations in case-control designs under structured populations. Ann Hum Genet 67:250–264.

Devlin B, Roeder K. 1999. Genomic control for association studies. Biometrics 55:997–1004.

Devlin B, Roeder K, Wasserman L. 2001. Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 60:155–166.

Devlin B, Bacanu SA, Roeder K. 2004. Genomic control to the extreme. Nat Genet 36:1129–1130.

Gorroochurn P, Hodge SE, Heiman GA, Greenberg DA. 2004. Effect of population stratification on case-control association studies: II: false-positive rates and their limiting behavior as the number of subpopulations increases. Hum Hered 58:40–48.

Heiman GA, Hodge SE, Gorroochurn P, Zhang J, Greenberg DA. 2004. Effect of population stratification on case-control association studies: I: elevation in the false positive rates and comparisons to confounding risk ratios (a simulation study). Hum Hered 58:30–39.

Marchini J, Cardon LR, Phillips MS, Donnelly P. 2004a. The effects of human population structure on large genetic association studies. Nat Genet 36:512–728.

Marchini J, Cardon LR, Phillips MS, Donnelly P. 2004b. Reply to "Genomic Control to the extreme". Nat Genet 36:1131.

Pritchard JK, Rosenberg NA. 1999. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220–228.

Pritchard JK, Stephens M, Donnelly P. 2000a. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. 2000b. Association mapping in structured populations. Am J Hum Genet 67:170–181.

Reich DE, Goldstein DB. 2001. Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol 20:4–16.

Sasieni PD. 1997. From genotypes to genes: doubling the sample size. Biometrics 53:1253–1261.

Satten GA, Flanders WD, Yang Q. 2001. 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.

Shmulewitz D, Zhang J, Greenberg DA. 2004. Case-control association studies in mixed populations: correcting using genomic control. Hum Hered 58:145–153.

Strachan T, Read AP. 2004. Human Molecular Genetics, 3rd edition. New York: Garland Science/Taylor & Francis Group.

Stuart A, Ord JK, Arnold A. 1999. Advanced Theory of Statistics, Vol. 2A: Classical Inference and the Linear Model, 6th edition. London: Hodder Headline Group.

Wright S. 1951. The genetical structure of populations. Ann Eugen 15:323–354.