Invited Original Paper

Human Heredity

Hum Hered 2007;64:149–159
DOI: 10.1159/000102988

Received: November 27, 2006
Accepted after revision: March 21, 2007
Published online: May 25, 2007

A Unified Approach for Quantifying, Testing and Correcting Population Stratification in Case-Control Association Studies

Prakash Gorroochurn[a] Susan E. Hodge[a, c] Gary A. Heiman[b]
David A. Greenberg[a, c]

[a]Division of Statistical Genetics, Department of Biostatistics, and [b]Department of Epidemiology, Mailman School of Public Health, Columbia University, and [c]Clinical-Genetic Epidemiology Unit, New York State Psychiatric Institute, New York, N.Y., USA

Key Words
Association studies · Population stratification · Genomic control · Delta-centralization · False positive rate

Abstract
The HapMap project has given case-control association studies a unique opportunity to uncover the genetic basis of complex diseases. However, persistent issues in such studies remain the proper quantification of, testing for, and correction for population stratification (PS). In this paper, we present the first unified paradigm that addresses all three fundamental issues within one statistical framework. Our unified approach makes use of an omnibus quantity (δ), which can be estimated in a case-control study from suitable null loci. We show how this estimated value can be used to quantify PS, to statistically test for PS, and to correct for PS, all in the context of case-control studies. Moreover, we provide guidelines for interpreting values of δ in association studies (e.g., at $\alpha = 0.05$, a δ of size 0.416 is small, a δ of size 0.653 is medium, and a δ of size 1.115 is large). A novel feature of our testing procedure is its ability to test for either strictly any PS or only 'practically important' PS. We also performed simulations to compare our correction procedure with Genomic Control (GC). Our results show that, unlike GC, it maintains good Type I error rates and power across all levels of PS.

Introduction

Recently, case-control association studies have come back into favor as a means to uncover the genetic basis of common complex diseases [1–3]. However, such studies may be undermined by population stratification (PS), resulting in an excess of false-positives. PS is a form of confounding that arises when cases and controls are sampled from genetically distinct populations [4]. PS can still occur if cases and controls are sampled from the same population when the latter is made up of genetically distinct subpopulations [5, 6]. In both situations, the result is that diseased and non-diseased individuals have inherent genetic differences because of the inherent genetic distance between the subpopulations [7].

PS is potentially a serious problem in association studies [4, 8] and is perhaps the most often cited reason for their non-replicability [9, 10]. Three questions are in order: How can PS in association studies (a) be quantified? (b) be tested for? (c) be corrected for?

In this paper, we provide the first unified approach that is able to answer all three questions within the same statistical framework. In an earlier paper [11], we derived a quantity δ to address the problem of correcting for PS in considerable detail, but there we investigated only Type I errors. Therefore, correcting for PS (issue c) will be mentioned only briefly here for the sake of completeness, although we do describe new simulations for investigating

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0001–5652/07/0643–0149$23.50/0

Accessible online at:
www.karger.com/hhe

Prakash Gorroochurn, PhD
Division of Statistical Genetics, R620, Department of Biostatistics
Columbia University, 722 W 168th Street
New York, NY 10032 (USA)
Tel. +1 212 342 1263, Fax +1 212 342 0484, E-Mail pg2113@columbia.edu

power. In that same paper, we also addressed the issue of quantification of PS through δ (issue a). However, our treatment of this issue was incomplete because we had not yet appreciated the full potential of δ. Therefore, we will spend some time here further elaborating on the usefulness of δ as an ideal measure of PS in case-control association studies, and we will refine some of our conclusions in that paper. New in this paper will also be guidelines for interpreting values of δ in case-control association studies. Finally, we will show how PS can be tested for (issue b) by using the estimated value of δ, thus providing the missing component in the 'quantifying-testing-correcting paradigm'. We will describe how to test either for strictly any PS or only PS that is deemed to be practically important. This distinction is important inasmuch as it lies at the heart of one of the conundrums in classical hypothesis testing, namely the inability to demarcate biologically important (or meaningful) results from statistically significant ones (e.g. [12, 13]). We provide a resolution of this issue here through a method that enables one to a priori decide which values of parameters would be important and then to test only for these. We stress that a full appreciation of the usefulness of δ as a unifying, omnibus quantity can be gleaned only by looking at all three problems concomitantly.

Summary of Previous Approaches

Regarding quantifying PS and the extent of its importance in association studies, two quantities have been often used. The first is Wright's F_{ST} [14] which measures the proportionate reduction in heterozygosity in a total population due to subdivision alone [15, p. 319]. Intuitively, the greater the variation in allele frequencies across subpopulations, the more 'population structure' there is, and the larger the value of F_{ST}. For a more thorough discussion of F_{ST} and related '*F*-statistics', see [15, 16]. The important point to note, however, is that F_{ST} is a function of allele frequencies and subpopulation sampling fractions only, and not of disease prevalences and sample size. The relevance of this observation will become clear in the next paragraph. The second quantity is the confounding risk ratio [CRR; 17], which represents the ratio of the unadjusted (for PS) relative risk of disease to the adjusted relative risk. The larger the confounding effects of PS, the larger the CRR. An excellent treatment of this quantity can be found in [18]. It is useful to note that CRR depends on allele frequencies, disease prevalences and subpopulation sampling fractions, but not on the sample size.

Wacholder et al. [19] used the confounding risk ratio (CRR) to measure the effects of PS, and found that CRR was biased by less than 10% in US studies involving non-Hispanic US Causasians of European origin. On the other hand, Heiman et al. [20] correlated CRR with the elevation in false positive rate (FPR), and found that CRR can be a poor predictor of FPR. They also used a standardized measure of the difference of marker allele frequencies which they showed to be highly predictive of FPR. Khlat et al. [21] used both CRR and FPR to gauge the effects of PS, and found PS to be of relatively small concern. Finally and more recently, Xu and Shete [22] quantified the degree of PS through Wright's F_{ST}, and found that PS could be a major problem. However, while CRR does increase with PS, and F_{ST} is an often-used measure of PS, these two measures are inadequate for capturing all the parameters that determine PS in association studies, namely subpopulation marker frequencies, subpopulation disease prevalences, subpopulation sampling fractions, and the case-control sample size [8, 11, 20, 23]. A natural way of quantifying PS in association studies would be in terms of a quantity which would contain all these parameters, and which would predict by how much such PS elevates the FPR in an association study. In what follows, we present such a quantity.

Concerning testing for PS, Pritchard and Rosenberg [24] proposed the use of a panel of biallelic 'null' marker loci (i.e. marker loci not associated with the trait and not in LD with the test locus). A usual χ^2 test statistic for the cases and controls is calculated for each null marker. Under the null hypothesis of no PS, the sum of the χ^2 at N null loci has a χ^2_N distribution. This sum can thus be used as a test statistic for PS. In this paper, we propose an alternative test for PS, when the importance of the latter is gauged by how much it elevates the FPR of the association test relative to its nominal α. Our test has the distinctive property that it can test for either strictly any PS or only PS that is practically important (we explain these two concepts later).

Finally, regarding correcting for PS while performing an association test, two main methods have been proposed. In the first major method, Genomic Control (GC) [25, 26], the assumption is that, in the presence of PS, the χ^2 test statistic for association has a standard χ^2_1 distribution when divided by an appropriate constant λ. The utility of the method lies in the fact that λ can be estimated from the empirical distribution of the null loci, usually either as the average of the test statistics at the null loci [27] or as their median divided by 0.456 [25]. To counteract possible anti-conservativeness of GC, Devlin et al. [28] also proposed GCF, which assumes an F_{1,L^*} distribution,

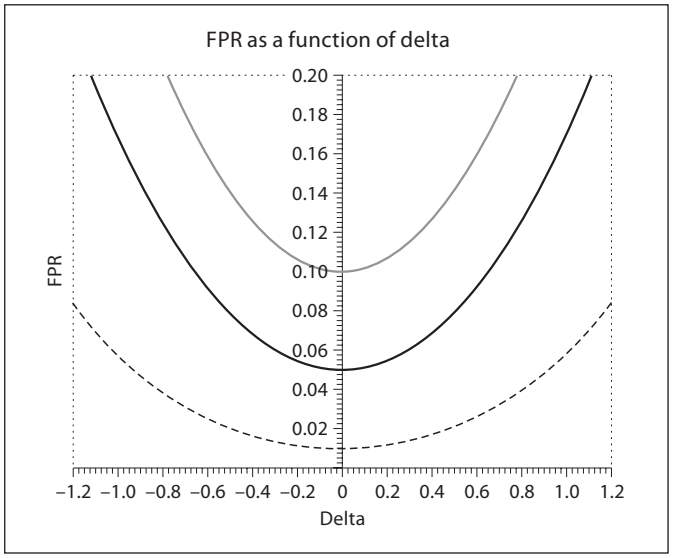

Fig. 1. Quantification of PS: Variation of the inflated FPR due to PS of the association test as a function of δ for $\alpha = 0.01$ (bottom curve), 0.05 (middle curve) and 0.1 (top curve) (see Equation 3).

instead of χ^2_1, where L^* is the total number of null loci used. In the second method, Pritchard et al. [29, 30] applied a Bayesian clustering method to a panel of null loci to infer the number of subpopulations and also the subpopulation membership of subjects. This information is then used to perform a test of association. This method is called structured association (SA). Although these methods can perform well, they also have drawbacks. For example, GC assumes that the distribution of the original test statistic can be transformed from a non-central χ^2 to a central one. However, we showed [11] this assumption is approximately true only for low levels of PS. The second method, SA, is very computationally intensive and can be problematic if the number of subpopulations is misspecified, although improvements have been made by Hoggart et al. [31]. Other recently proposed methods that correct for PS include the use of logistic regression [32–34] and principal component analysis [35, 36].

The Unified Approach

Quantifying PS

There exists a single quantity that completely captures the confounding effects of PS in a case-control study, when measured by the resulting elevation in FPR of the association test used. Consider any population which is made up of discrete subpopulations. Assume there is no true association between marker genotype (i.e. a genotype containing at least one marker allele) and disease in any subpopulation (we later in this section relax this assumption). For a case-control study with m cases and n controls sampled from the population, suppose we use the standard χ^2 test (without continuity correction) for association in 2×2 tables:

$$\chi^2_{\text{test}} = \frac{(m+n)\left(a_0 d_0 - b_0 c_0\right)^2}{mn\left(a_0 + c_0\right)\left(b_0 + d_0\right)}, \tag{1}$$

where, at the test locus, a_0 and b_0 are the number of cases with and without the marker genotype, and c_0 and d_0 are the number of controls with these respective genotypes. Let s be the proportion of diseased individuals in the total population with a marker genotype, and t the proportion of non-diseased individuals with a marker genotype. Define [11]

$$\delta \equiv \frac{s-t}{\sqrt{\frac{s(1-s)}{m} + \frac{t(1-t)}{n}}}. \tag{2}$$

Then, for a test with nominal α and reasonably large samples, the inflated FPR due to PS of the χ^2 test is given by [11]

$$\text{FPR} = 1 - \Phi(z_{\alpha/2} - \delta) + \Phi(-z_{\alpha/2} - \delta), \tag{3}$$

where $\Phi(.)$ is the standard normal distribution function and $z_{\alpha/2}$ is the $(\alpha/2)$th upper percentile of the normal distribution. This relationship is correct to the extent that the use of the χ^2 test in Equation (1) can be justified, i.e. for reasonably large sample sizes (e.g. no expected frequency in the 2×2 table should be less than 5 [37]). When there is no PS, $\delta = 0$ and FPR = α. In the presence of *any* PS, $\delta \neq 0$ and FPR $> \alpha$. The quantity δ thus quantifies PS in a case-control association study because it predicts exactly by how much such PS will inflate the FPR of the association test. In this paper therefore, we equate the existence of PS in association studies to the condition that $\delta \neq 0$. Figure 1 shows some graphs for various test sizes $\alpha = 0.01$, 0.05 and 0.1. Equation 3 can be used to obtain guidelines for interpreting values of δ, at given values of α. Some examples are given in table 1 below for $\alpha = 0.001$, 0.01, and 0.05. In Appendix A, we also include a Maple program that can be used to estimate δ for a given α and FPR.

Based on the derivation of Equation (3) in [11], we stated there that δ can be used as a measure of PS only when there is no true association between disease and marker genotype. In point of fact, this is not completely true because of the way in which δ is estimated in a case-control study. We now explain the latter by allowing for the possibility of true associations between marker genotype and disease. Suppose subjects in our case-control study are

Table 1. Guidelines for interpreting the magnitude of δ, at a given nominal α of the association test, and the corresponding inflated FPR of the association test due to PS

	Size of effect	Magnitude of δ	Approx. FPR due to PS
$\alpha = 0.001$	small	0.256	0.0014
	medium	0.394	0.002
	large	0.635	0.004
$\alpha = 0.01$	small	0.323	0.014
	medium	0.500	0.02
	large	0.821	0.04
$\alpha = 0.05$	small	0.416	0.07
	medium	0.653	0.1
	large	1.115	0.2

genotyped at a set of null loci. Recall that, at the test locus, the number of cases and controls with the marker genotype are a_0 and c_0, respectively. First, estimate t at the test locus by c_0/n. Next select those null loci whose estimates of t are within, for example, $\pm$ 0.15 of c_0/n. Suppose there are L such loci, and that the number of cases and controls who have been genotyped at the j-th locus are m_j and n_j, respectively. For the j-th (j = 1, 2, …, L) such locus, the estimate of s_j is a_j/m_j, that of t_j is c_j/n_j, and that of δ_j is

$$\hat{\delta}_j = \frac{\left(a_j/m_j\right) - \left(c_j/n_j\right)}{\sqrt{\frac{\left(a_j/m_j\right)\left\{1-\left(a_j/m_j\right)\right\}}{m_j} + \frac{\left(c_j/n_j\right)\left\{1-\left(c_j/n_j\right)\right\}}{n_j}}}. \quad (4)$$

Combining all L matched loci, the overall estimate of δ and that of the inflated FPR due to PS are

$$\hat{\delta} = \frac{1}{L}\sum_{j=1}^{L}\hat{\delta}_j,$$

$$\widehat{FPR} = 1 - \Phi\left(z_{\alpha/2} - \hat{\delta}\right) + \Phi\left(-z_{\alpha/2} - \hat{\delta}\right). \quad (5)$$

This shows that, by matching the genotype frequencies of null loci approximately to that of the test locus in the controls, the above method for estimating δ ensures that the latter captures only the effect of PS and not that of any true association. Thus, δ can be used as a measure of PS when true association between disease and marker genotype is either absent or present, as will be seen later in our simulations of power when correcting for PS.

Testing for PS

Is there a simple way to statistically test for PS in a case-control study? We now propose an alternative test to that of Pritchard and Rosenberg [24], when the importance of PS is gauged by how much it elevates the FPR of the association test relative to its nominal α. Our test is based on the estimated value of δ, and has the distinctive feature that it can test for either strictly any PS or only PS that is practically important. We now explain these two concepts.

Strictly Any PS and Practically Important PS

Any $\delta \neq 0$ implies the presence of PS, so that the two-sided test H_0: $\delta = 0$ versus H_a: $\delta \neq 0$ is a test for strictly any PS. However, when performing such a test, we could face a classical dilemma in hypothesis testing [e.g. 12, 13], where our test could be highly significant, yet hardly practically important. To illustrate, the true value of δ could be only 0.01, and statistically significant in a 'large sample,' but of no scientific or practical importance. For example, if $\alpha = 0.05$, Equation 3 shows that $\delta = 0.01$ corresponds to PS that inflates FPR to only 0.0512.

Since, for a given nominal α, we have an explicit relationship (Equation 3) between δ and the inflated FPR due to PS of the association test to be used, we can avoid the above-mentioned problem. Depending on how large we allow FPR due to PS to be (relative to the nominal α), without being practically important, we can determine the magnitude of δ (say, δ_p) from Equation 3. The test we should then perform is an equivalence test [38] H_0: $|\delta| \geq \delta_p$ vs. H_a: $0 \leq |\delta| < \delta_p$. This is a test for only PS that is practically important. The choice for δ_p depends on the nominal α of the association test to be used, on the actual FPR of this association test when there is no PS and no true disease-marker association, and on the perceptions of the investigator. For example, if α = 0.05, the investigator might regard an inflated FPR due to PS of up to 0.07 to be practically unimportant and therefore use $\delta_p = 0.416$ (obtained from Equation 3) in the test for PS. On the other hand, if $\alpha = 0.01$, and only FPR >0.014 is to be regarded as practically important, then $\delta_p = 0.323$ should be used.

Statistical Tests

Let us assume that the L selected matched loci are in linkage equilibrium with each other. (This is likely to be true if, for example, SNPs at a distance ~1 cM from each other are chosen as the null markers). The test for strictly no PS is two-sided. Under H_0: $\delta = 0$,

$$T = \frac{\hat{\delta}}{SD/\sqrt{L}} \sim t_{L-1} \quad (6a)$$

approximately, where $SD^2 = \{\sum_{j=1}^{L}\hat{\delta}_j^2 - (\sum_{j=1}^{L}\hat{\delta}_j)^2/L\}/(L - 1)$ and t_{L-1} is the Student's t-distribution with $L - 1$ de-

grees of freedom (the rationale for the use of the *t*-distribution stems from the fact that $\hat{\delta}$ is approximately normally distributed). Based on Equation 6a, for a test of size α', reject H_0: $\delta = 0$ if $|T| \geq t_{L-1,\,\alpha'/2}$, where $t_{L-1,\,\alpha'/2}$, is the $(\alpha'/2)$th upper percentile of t_{L-1}. Note that α' here is the level of significance of the test for *PS* and must not be confused with the nominal α in Equation 3, the latter being the level of significance of the test for association and is used to obtain δ_p.

On the other hand, to test for only practically unimportant PS:

(i) Depending on the nominal α of the association test, decide how large the inflated FPR due to PS can be.

(ii) Determine the corresponding δ_p from Equation 3 by using the program in Appendix A or contacting the first author (P.G.).

(iii) For a test of size α', reject the null hypothesis of practically important PS if the $(1–2\alpha')$ confidence interval for the true δ falls completely within the 'zone of indifference' $(-\delta_p, \delta_p)$ [38], i.e. if

$$\hat{\delta} - t_{L-1,\,\alpha'}\frac{SD}{\sqrt{L}} > -\delta_p \quad \text{and} \quad \hat{\delta} + t_{L-1,\,\alpha'}\frac{SD}{\sqrt{L}} < \delta_p. \tag{6b}$$

Note: (a) because this is an equivalence test, the null and alternative hypotheses have been switched: the alternative hypothesis is that of equivalence (i.e. PS is practically unimportant); (b) the equivalence test based on Equation 6b above has significance level α', as required, rather than $2\alpha'$ [see 39, p 172].

Correcting for PS

Is there a simple way δ can be used to correct for PS while testing for association? Let χ^2_{test} in Equation 1 denote the value of the χ^2 test-statistic for the 2×2 table of marker genotype versus disease status at the test locus. Using the estimated value of δ from the null loci as in Equation 5, we showed [11] that the test statistic

$$T^*_{DC} \equiv \left\{\text{sign}\left(a_0/m - c_0/n\right) \times \sqrt{\chi^2_{\text{test}}} - \hat{\delta}\right\}^2 \tag{7}$$

has the standard χ^2_1 distribution, where:

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

T^*_{DC} is therefore the association test statistic that corrects for PS. We have called this method δ-centralization (DC). It is computationally simpler that SA and, unlike GC, its validity does not depend on the degree of confounding due to PS (i.e. on the value of δ).

Simulations

We performed computer simulations in order to investigate Type I errors and power. We first describe how the candidate locus and null loci were generated in a discrete population substructure model. We assume there is no true association between marker genotype and disease in each subpopulation. Using a given value of p_k as the marker allele frequency at the test locus in both cases and controls in subpopulation k ($k = 1, 2, \ldots, K$), the allele frequency (p_{jk}) of the j-th null locus in subpopulation k was generated from a

$$\text{Beta}\left(\frac{1-F_k}{F_k}\,p_k,\ \frac{1-F_k}{F_k}\,q_k\right)$$

distribution for both cases and controls, where $q_k = 1 - p_k$ and F_k is Wright's coefficient of inbreeding in subpopulation k. The values of F_k chosen were in the range 0.01–0.05, which is typical of most human populations [see 7, table 2.3.1A]. Using the values of these allele frequencies, we calculated s_j and t_j for the j-th null locus, as follows:

$$s_j = \frac{\sum_{k=1}^{K}\pi_k d_k r_{jk}}{\sum_{k=1}^{K}\pi_k d_k}, \quad t_j = \frac{\sum_{k=1}^{K}\pi_k (1-d_k) r_{jk}}{1-\sum_{k=1}^{K}\pi_k d_k}. \tag{8}$$

Here π_k, d_k, and r_{jk} are the subpopulation sampling fraction, disease prevalence and marker genotype frequency at the j-th null locus (calculated from p_{jk} by assuming Hardy-Weinberg equilibrium) of the k-th subpopulation. Let a_j and c_j be the number of cases and controls, respectively, having the marker genotype at the j-th locus. The values s_j and t_j for the j-th null locus were used to generate a 2×2 table at that locus, since $a_j \sim \text{Bin}(m_j, s_j)$ and $c_j \sim \text{Bin}(n_j, t_j)$. Once these tables were generated, confounding due to PS was quantified, tested and corrected for, using the methods as we described above. If we assume there is true association between marker genotype and disease in each subpopulation, we then use a different marker genotype frequency between diseased and non-diseased subjects in each subpopulation. For comparison purposes, we applied DC, GC and GCF (see tables 4 and 5). Comparisons were done for increasing values of δ.

We performed the above simulations to compare the operating characteristics of the *t* test for strictly no PS in Equation 6a (table 2) and for practically unimportant PS in Equation 6b (table 3). In the latter case, we allowed an FPR of up to 0.07 to be practically unimportant (at $\alpha = 0.05$) and therefore used $\delta_p = 0.416$. Table 2 shows that matching to within ± 0.15, rather than ± 0.1, gives the *t* test better operating characteristics. Thus the *T*-statistic

Table 2. Testing for strictly any PS: Type I error and power of *t*-test (to detect any PS) when genotypes are matched to within ±0.1, ±0.125, and ±0.15

	Configuration numbers	Value of δ	Type I error of *t*-test		
			±0.1 matching	±0.125 matching	±0.15 matching
No PS	1	0	0.10690	0.07360	0.05300
	2	0	0.11900	0.07380	0.06440
	3	0	0.10090	0.07100	0.05300
		Approx. value of δ	Power of *t*-test		
			±0.1 matching	±0.125 matching	±0.15 matching
PS	9	0.5	0.95340	0.98160	0.98980
	14	1.0	1.0000	1.0000	1.0000
	18	2.0	1.0000	1.0000	1.0000

L^* = 100, $m = n$ = 200, α' = 0.05, K = 2 subpopulations, F_1 = 0.05, F_2 = 0.01. See configuration numbers from table A1 in Appendix B for description of actual values of $\boldsymbol{\pi}$, **d**, and **r** used.

Table 3. Testing for practically important PS: Type I error and power of *t*-test (to detect practically important PS) when genotypes are matched to within ±0.15

	Configuration numbers	Approx. value of δ	Type I error of *t*-test
Practically important PS	9	0.5	0.01100
	10	0.6	0.00780
	11	0.7	0.00020
	12	0.8	0.00000
	14	1.0	0.00000
	18	2.0	0.00000
			Power of *t*-test
Practically unimportant PS	1	0.0	0.98040
	4	0.1	0.91940
	5	0.2	0.67420
	6	0.3	0.23880
	7	0.4	0.05120

L^* = 100, $m = n$ = 200, α' = 0.05, K = 2 subpopulations, F_1 = 0.05, F_2 = 0.01. Matching was done to within ±0.15. We have here defined practically unimportant PS as PS that inflates the FPR of an association test up to 0.07 (at a nominal α = 0.05), and therefore used δ_p = 0.416. See configuration numbers from Table A1 in Appendix B for description of actual values of $\boldsymbol{\pi}$, **d**, and **r** used.

in Equation 6a fails to have the nominal Student's *t*-distribution when matching is too narrow. Table 3 shows that power is relatively high for very small δ ($\approx$ 0 to 0.1), but starts decreasing for $0.2 \leq \delta < 0.416$.

Simulation results comparing DC, GC and GCF are shown in tables 4 and 5. The performance of GC depends very much on the degree of confounding, i.e. on the value of δ. With increasing δ, GC becomes more and more conservative. At δ = 0.5, GC performs well, but at δ = 1.0 it becomes over-conservative, rejecting at an average rate of about 2.5% for a nominal α = 0.05. The situation is worse for larger δ [see also 40]. Concerning power, GC seems to lose considerable power only for $\delta > 2.0$. In all these situations, DC performs well with respect to both Type I and Type II errors. GCF always has a smaller rejection rate than GC since it is based on an F_{1, L^*} distribution, instead of χ^2_1, and $F_{1, L^*} > \chi^2_1$ for all L^*. Therefore when GC is either over-conservative or 'low-powered', GCF is even more so.

Discussion

PS in association studies cannot be disregarded outright and may be a major reason why many positive findings fail to replicate. In this paper, we have provided, for the first time, one statistical framework that is able to address all aspects of PS in case-control association studies: we have shown that, through an omnibus quantity (δ), PS can be (a) quantified, (b) tested for, and (c) corrected for. We summarize our findings for these three issues in turn.

Summary of Results for the Three Issues

The quantity δ is ideally suited as a measure of PS (issue a) in case-control association studies. This is because, by estimating δ, one is able to exactly predict by how much PS in a case-control study inflates the FPR of an association test relative to its nominal α (see Equation 3 and fig. 1). In general, for a given α, FPR (and hence the extent of PS) increases with the magnitude of δ. Moreover, we can use Equation 3 to better interpret values of δ (see table 1): for example, at α = 0.05, a δ of size 0.416 (or less) is small because it corresponds to an inflated FPR of the association test of only 0.07, a δ of size about 0.653 is medium (the FPR is twice the nominal α), and a δ of size 1.115 (or more) is large (the FPR is four times the nominal α). We stress that δ is not a global measure of stratification between subpopulations; rather, it is a locus-specific measure. A global measure of PS cannot correctly indicate the extent to which a test for association will be affected by PS at a particular locus because the 'global' difference between allele frequencies at several markers will often be

Table 4. Correcting for PS: Type I error and power of GC, GCF and DC (to detect any true association) with 2 subpopulations

Approximate value of δ		CONFIG #	Type I error	CONFIG #	Power
0.5	uncorrected FPR	8	0.07900	9	
	FPR using GC with median		0.04700		0.99880
	FPR using GC with mean		0.04820		0.99980
	FPR with GCF with mean		0.04540		0.99980
	FPR using DC		0.04700		0.99900
1.0	uncorrected FPR	13	0.16920	14	
	FPR using GC with median		0.02520		0.88740
	FPR using GC with mean		0.03200		0.92940
	FPR with GCF with mean		0.03020		0.92540
	FPR using DC		0.05260		0.92060
1.5	uncorrected FPR	15	0.33480	16	
	FPR using GC with median		0.00420		0.81520
	FPR using GC with mean		0.00480		0.90680
	FPR with GCF with mean		0.00400		0.89740
	FPR using DC		0.04460		0.98500
2.0	uncorrected FPR	17	0.51920	18	
	FPR using GC with median		0.00080		0.64800
	FPR using GC with mean		0.00100		0.84580
	FPR with GCF with mean		0.00080		0.82760
	FPR using DC		0.04820		0.99400

$L^* = 100$, $m = n = 200$, $\alpha = 0.05$, $K = 2$ subpopulations, $F_1 = 0.05$, $F_2 = 0.01$. Matching was done to within ± 0.15. See configuration numbers from Table A1 in Appendix A for description of actual values of $\boldsymbol{\pi}$, $\mathbf{d}$, and $\mathbf{r}$ used. (Note that, unlike Type I errors, power values cannot be compared across values of δ because the effect size, i.e. 'strength of true association', is different for each value of δ).

significantly different from a 'locus-specific' difference. Therefore, our argument is, since we are considering association tests performed on a locus-to-locus basis, a correct measure of PS for an association study should be a local measure, specific to the locus being tested for association. Thus, δ is not intended as a replacement of global measures of population structure such as F_{ST}.

We make three additional observations: (i) From Equation 3, we note that the value of δ must be used in conjunction with the nominal α of the association test when gauging the extent of PS. For example, $\delta = 0.4$ is 'huge' when $\alpha = 10^{-5}$, because the FPR ($= 3 \times 10^{-5}$) is tripled relative to the nominal α (a relative increase of 200%); however, the same $\delta = 0.4$ is 'very small' when $\alpha = 0.05$ because the FPR is only 0.0685 (a relative increase of only 37%). (ii) A second point regards a statistical connection of δ: in [11] we showed that, in the presence of PS and in the absence of any true association, χ^2_{test} in Equation 1 has a non-central χ^2 distribution with 1 degree of freedom and approximate non-centrality parameter δ^2. (iii) The third observation concerns a desirable relationship between δ and sample size: from Equation 2, we see that, for a fixed value of the numerator $s - t$ (which can be thought of as a measure of the inherent genetic distance between diseased and non-diseased individuals due to subpopulations), δ and hence the confounding effect of PS can be made arbitrarily large by increasing the number of cases and controls. This is a classic case whereby bias is made even worse by increasing the sample size [5, 20, 41, 42]. This is because, in essence, the biased procedure is actually testing a different null hypothesis, so increasing the sample size raises the 'power' to reject that null hypothesis.

Our second use of δ relates to the testing for PS (issue b). We have shown how to test for either strictly any PS or only PS that is practically important PS. By practically important PS, we mean PS that inflates the FPR of an association test beyond a certain amount, relative to its nominal α. By deciding how big an FPR is large enough, relative to the nominal α, one can use Equation 3 to estimate the corresponding δ_p and perform an equivalence test for practically important PS. Our guidelines are to

Table 5. Correcting for PS: Type I error and power of GC, GCF and DC (to detect any true association) with 3 subpopulations

Approximate value of δ			Type I error		Power
0.5	uncorrected FPR	19	0.07720	20	
	FPR using GC with median		0.03360		0.45380
	FPR using GC with mean		0.03260		0.46620
	FPR with GCF with mean		0.03040		0.45340
	FPR using DC		0.04720		0.43520
1.0	uncorrected FPR	21	0.16800	22	
	FPR using GC with median		0.01920		0.92600
	FPR using GC with mean		0.02480		0.95900
	FPR with GCF with mean		0.02320		0.95660
	FPR using DC		0.04560		0.96120
1.5	uncorrected FPR	23	0.33280	24	
	FPR using GC with median		0.00200		0.33000
	FPR using GC with mean		0.01380		0.62860
	FPR with GCF with mean		0.01260		0.61080
	FPR using DC		0.04240		0.71400
2.0	uncorrected FPR	25	0.52700	26	
	FPR using GC with median		0.00000		0.34580
	FPR using GC with mean		0.00600		0.82920
	FPR with GCF with mean		0.00520		0.81500
	FPR using DC		0.04500		0.93480

$L^* = 100$, $m = n = 200$, $\alpha = 0.05$, $K = 3$ subpopulations, $F_1 = 0.05$, $F_2 = 0.01$, $F_3 = 0.02$. Matching was done to within ± 0.15. See configuration numbers from Table A2 in Appendix C for description of actual values of $\boldsymbol{\pi}$, $\mathbf{d}$, and $\mathbf{r}$ used. (Note that, unlike Type I errors, power values cannot be compared across values of δ because the effect size, i.e. 'strength of true association', is different for each value of δ).

use $\delta_p = 0.416$ when $\alpha = 0.05$, corresponding to an inflated FPR of 0.07 due to PS, and $\delta_p = 0.323$ when $\alpha = 0.01$, corresponding to an inflated FPR of 0.014 due to PS. Of course, all of this assumes that the association test used has an FPR close to the nominal α when there is no PS and no true association [43], and to this end we recommend using a χ^2 without continuity correction. This is because the latter has an FPR close to the nominal α, as opposed to the χ^2 with continuity correction whose FPR is much less than α [e.g., see 44, p. 137]. Since both quantifying PS and testing for practically important PS depend on the nominal α of the association test to be subsequently performed, it is also natural to ask how to do either of these if the aim is only to quantify or test, and not correct later. Here, we suggest assuming the most widely used value of α, namely 0.05, for both quantifying, and testing by calibrating δ_p accordingly.

Our third use of δ concerns correcting for PS (issue c) while performing an association test (see Equation 8). The method we have proposed (DC) works by directly centralizing the test statistic χ^2_{test} in Equation 1, which has a non-central χ^2 distribution in the presence of PS, through its non-centrality parameter (δ^2). The estimate $\hat{\delta}$ in essence 'distills' χ^2_{test} by removing the effect of PS and giving out only the genetic effect. The reason why χ^2_{test} can be centralized is because, in the presence of PS, its distribution is non-central with only 1 degree of freedom. The test in Equation 7 is two-sided. For one-sided alternatives, the test statistic to use is

$$\text{sign}\left(a_0/m - c_0/n\right) \times \sqrt{\chi^2_{\text{test}}} - \hat{\delta},$$

which has a standard normal distribution. DC is similar to GC in that both use null loci as 'genomic controls'. However, unlike GC, DC is valid for all levels of PS (i.e. all values of δ) and is also computationally simpler than SA. Finally, if narrow frequency matching is performed, as few as 25 matched null loci are enough for DC to perform well [11].

Concluding Remarks

Since our unified approach depends on δ, it is crucial that this quantity be accurately estimated in any case-

control study. We have shown how this can be done by genotyping both cases and controls at a set of null marker loci (preferably SNPs), and then selecting those null loci whose marker genotype (hence allele) frequencies in the controls closely match those of the test locus in the controls. A value of δ can be estimated from each matched null locus, and by averaging, an overall estimate of δ can be obtained. It is thus important that there is a sufficiently large number of controls so that genotype (or allele) frequencies can be accurately estimated at both the test and null loci. Moreover, since δ is obtained by matching, the methods in our unified approach (quantifying, testing, and correcting) are independent of the probability distribution used for generating loci across the genome. Finally, the method of frequency matching by using controls ensures that δ is a measure of PS either in the presence or absence of any true association between disease and marker genotype.

Although the distribution of the estimated δ values at all null loci is approximately normal, that of the matched null loci is roughly uniform. This has an important implication regarding the tests for PS in Equations 6a and 6b, being both based on the Student's *t*-distribution. These tests essentially rely on the Central Limit Theorem (CLT) and, as such, their validity depends on the number of matched loci (L) being large enough. Moreover, our simulations (table 1) show that if the matching is too strict (± 0.1) the CLT does not hold and the *t*-test in Equation 6a becomes over anti-conservative. When matching is done to within ± 0.15, the CLT is more applicable and the *t* test has correct error rates. As far as correcting for PS is concerned, the narrower the matching window, the more accurate the estimation of δ. However, this also means more null loci will have to be genotyped before the suitable ones are found. In general, we recommend a matching window of ± 0.15, since both the test for the PS and the correction for PS then have good operating characteristics.

Should we always test for PS first and then decide whether to apply the DC method or use an uncorrected test-statistic based on the result of the first test? There are at least three problems with such an approach: (i) if a Type I error is made in the test for strictly no PS, then using the DC method for association will result in loss of power; (ii) if a Type II error is made in the test for strictly no PS, then using an uncorrected test for association will result in an increased FPR; (iii) finally, if PS is declared practically unimportant and an association test is performed without correcting for it, the FPR of the association test will increase. Another fact is that most practitioners will feel safer to estimate δ and then directly move to correct for PS, without necessarily going through the testing phase. We therefore advocate flexibility in using the unified approach: testing and correcting could best be thought as independent tasks, and there is no need to do one before the other, once δ is estimated.

An important topic concerns the issue of multiple-tests performed in genome-wide association studies [45–48]. Here the danger is that of a rapid rise in the experimentwise error-rate because of the large number of test loci. On the other hand, a Bonferroni-correction approach results in the opposite problem: substantial loss in power because the typical test size is $\sim 10^{-6}$ [2]. A reasonable compromise can be achieved through: (i) Benjamini and Hochberg's FDR procedure [49] if no correlation between loci is allowed for, (ii) variants of the FDR procedure [e.g. 50] if possible correlation between loci is allowed for. We therefore recommend applying the FDR or its variants when either testing or correcting for PS for a large number of loci.

It is also important to note the genetic model we have used here. We have assumed a dominant genetic model (in the sense that the genotypic relative risk for disease is the same if the genotype at the test locus has one or two marker alleles), and both the definition of δ in Equation 2 and the DC test statistic in Equation 7 are particular to such a model. DC can still be applied if a recessive model is assumed, by changing the definition of a marker genotype to include both marker alleles only. For an additive model, both Equation 2 and Equation 7 change drastically [see 11]: things are more complicated, but can still be done.

Acknowledgments

This work was supported in part by NIH grants R01 AA013654, DK-31813, NS27941, DK31775. We thank Dr Bruce Levin for his continued invaluable advice. We also thank two anonymous referees for considerably improving the paper.

Appendix

A. Maple program that uses the Newton-Raphson procedure to estimate δ for a given α and FPR

```
raphson := proc(fpr, alpha, delta_start)
local z_alpha, delta, i, f, f_dash;
    with(stats);
    z_alpha := statevalf[icdf, normald[0, 1]](1 - .5*alpha);
    delta[1] := delta_start;
    for i to 25 do
```

```
        f[i] := 1
            – statevalf[cdf, normald[0, 1]](z_alpha – delta[i])
            + statevalf[cdf, normald[0, 1]](-z_alpha – delta[i])
            – fpr;
        f_dash[i] :=
            statevalf[pdf, normald[0, 1]](z_alpha – delta[i])
            – statevalf[pdf, normald[0, 1]](z_alpha + delta[i]);
            delta[i + 1] := delta[i] - f[i]/f_dash[i];
            if abs(delta[i + 1] – delta[i]) < .0001 then
                RETURN(delta[i])
            fi
        od
    end
```

B. Table A1. Configuration numbers (#) as indicated in tables 2, 3, and 4, and corresponding values of δ, $\boldsymbol{\pi}$, **d**, and **r** for $K = 2$ subpopulations with $m = n = 200$

#	Approx. δ	Actual values				
		δ	$\boldsymbol{\pi}$	**d**	**r** (cases)	**r** (controls)
1	0	0.0000	(0.2620, 0.7380)	(0.2, 0.2)	(0.85, 0.50)	(0.4, 0.2)
2	0	0.0000	(0.2620, 0.7380)	(0.1, 0.2)	(0.85, 0.50)	(0.2, 0.2)
3	0	0.0000	(0.45, 0.55)	(0.034, 0.034)	(0.18, 0.42)	(0.12, 0.32)
4	0.1	0.1001	(0.3447, 0.6553)	(0.2396, 0.1992)	(0.6, 0.4)	(0.4154, 0.3272)
5	0.2	0.2002	(0.5537, 0.4463)	(0.2352, 0.1430)	(0.6, 0.5)	(0.4382, 0.3706)
6	0.3	0.3000	(0.2872, 0.7128)	(0.1050, 0.1634)	(0.3, 0.5)	(0.1779, 0.3194)
7	0.4	0.3998	(0.2872, 0.7128)	(0.1050, 0.1915)	(0.2, 0.4)	(0.1779, 0.3194)
8	0.5	0.4979	(0.7648, 0.2352)	(0.1908, 0.1416)	(0.5614, 0.1484)	(0.5614, 0.1484)
9	0.5	0.4979	(0.7648, 0.2352)	(0.1908, 0.1416)	(0.8421, 0.2226)	(0.5614, 0.1484)
10	0.6	0.5995	(0.7894, 0.2016)	(0.1634, 0.0131)	(0.6, 0.4)	(0.4760, 0.3336)
11	0.7	0.6999	(0.5458, 0.4542)	(0.2713, 0.1603)	(0.6, 0.4)	(0.4592, 0.2477)
12	0.8	0.7990	(0.4789, 0.5211)	(0.1145, 0.2429)	(0.3, 0.5)	(0.1688, 0.3323)
13	1.0	1.0027	(0.5241, 0.4759)	(0.2818, 0.1654)	(0.4298, 0.1505)	(0.4298, 0.1505)
14	1.0	1.0027	(0.5241, 0.4759)	(0.2818, 0.1654)	(0.6447, 0.2258)	(0.4298, 0.1505)
15	1.5	1.5026	(0.2620, 0.7380)	(0.2502, 0.05855)	(0.4770, 0.2882)	(0.4770, 0.2882)
16	1.5	1.5026	(0.2620, 0.7380)	(0.2502, 0.05855)	(0.7, 0.45)	(0.4770, 0.2882)
17	2.0	2.0030	(0.2620, 0.7380)	(0.2502, 0.03326)	(0.4770, 0.2882)	(0.4770, 0.2882)
18	2.0	2.0030	(0.2620, 0.7380)	(0.2502, 0.03326)	(0.7155, 0.4323)	(0.4770, 0.2882)

To obtain the values of $\boldsymbol{\pi}$, **d**, **r** (controls) for a value of δ, we fixed all but one of the values of π_i, d_i, r_i (controls), and then solved for the unknown using Equations 2 and 8. The values of **r** (cases) were made to equal those of **r** (controls) for the Type I errors in Tables 4 and 5, but were otherwise chosen arbitrarily.

C. Table A2. Configuration numbers (#) as indicated in table 5, and corresponding values of δ, $\boldsymbol{\pi}$, **d**, and **r** for $K = 3$ subpopulations with $m = n = 200$

#	$\approx \delta$	Actual values				
		δ	$\boldsymbol{\pi}$	**d**	**r** (cases)	**r** (controls)
19	0.5	0.4984	(0.2036, 0.5375, 0.2589)	(0.2474, 0.2043, 0.0064)	(0.4343, 0.2492, 0.2298)	(0.4343, 0.2492, 0.2298)
20	0.5	0.4984	(0.2036, 0.5375, 0.2589)	(0.2474, 0.2043, 0.0064)	(0.6, 0.3, 0.3)	(0.4343, 0.2492, 0.2298)
21	1.0	1.0028	(0.3242, 0.2027, 0.4730)	(0.1594, 0.1070, 0.3000)	(0.1130, 0.1165, 0.2859)	(0.1130, 0.1165, 0.2859)
22	1.0	1.0028	(0.3242, 0.2027, 0.4730)	(0.1594, 0.1070, 0.3000)	(0.2, 0.2, 0.5)	(0.1130, 0.1165, 0.2859)
23	1.5	1.4979	(0.0858, 0.1566, 0.7576)	(0.1114, 0.1392, 0.05012)	(0.3382, 0.4228, 0.1280)	(0.3382, 0.4228, 0.1280)
24	1.5	1.4979	(0.0858, 0.1566, 0.7576)	(0.1114, 0.1392, 0.05012)	(0.5, 0.6, 0.2)	(0.3382, 0.4228, 0.1280)
25	2.0	2.0001	(0.0559, 0.5979, 0.3462)	(0.2719, 0.2023, 0.03193)	(0.4403, 0.4182, 0.1233)	(0.4403, 0.4182, 0.1233)
26	2.0	2.0001	(0.0559, 0.5979, 0.3462)	(0.2719, 0.2023, 0.03193)	(0.6, 0.6, 0.2)	(0.4403, 0.4182, 0.1233)

To obtain the values of $\boldsymbol{\pi}$, **d**, **r** (controls) for a value of δ, we fixed all but one of the values of p_i, d_i, r_i (controls) and then solved for the unknown using Equations 2 and 8. The values of **r** (cases) were made to equal those of **r** (controls) for the Type I errors in Tables 4 and 5, but were otherwise chosen arbitrarily.

References

1 Cardon LR, Bell JI: Association study designs for complex diseases. Nat Rev Genet 2001;2:91–99.
2 Strachan T, Read A: Human Molecular Genetics, ed 3. KY, Garland Science/Taylor & Francis Group, 2004.
3 Palmer LJ, Cardon LR: Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet 2005;366:1223–1234.
4 Schork NJ, Fallin D, Thiel B, Xu X, Broeckel U, Jacob HJ, Cohen D: The future of genetic case-control studies; in Rao DC, Province MA (eds): Genetic Dissection of Complex Traits. CA, Academic Press, 2001.
5 Thomas DC, Witte JS: Point: population stratification: a problem for case-control studies of candidate-gene associations? Cancer Epidemiol Biomarkers Prev 2002;11:505–512.
6 Cardon LR, Palmer LJ: Population stratification and spurious allelic association. Lancet 2003;361:598–604.
7 Cavalli-Sforza LL, Menozzi P, Piazza A: The History and Geography of Human Genes. NJ, Princeton Univ. Press, Princeton, 1994.
8 Rosenberg NA, Nordborg M: A general population-genetic model for the production by population structure of spurious genotype-phenotype associations in discrete, admixed or spatially distributed populations. Genetics 2006;173:1665–1678.
9 Terwilliger JD, Goring HH: Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. Hum Biol 2000;72:63–132.
10 Weiss KM, Terwilliger JD: How many diseases does it take to map a gene with SNPs? Nat Genet 2000;26:151–157.
11 Gorroochurn P, Heiman GA, Hodge SE, Greenberg DA: Centralizing the non-central chi-square: a new method to correct for population stratification in genetic case-control association studies. Genet Epidemiol 2006;30:277–289.
12 Runyon RP, Haber A, Pittenger DJ, Coleman KA: Fundamentals of Behavioral Statistics, ed 8. New York, McGraw-Hill College, 1995.
13 Salsburg D: The Lady Tasting Tea. New York, W. H. Freeman & Co., 2001.
14 Wright S: The genetical structure of populations. Ann Eugen 1951;15:323–354.
15 Halliburton R: Introduction to Population Genetics. New Jersey, Prentice Hall, 2003.
16 Hedrick PW: Genetics of Populations, ed 3. Sudbury, MA, Jones and Bartlett, 2005.
17 Miettinen OS: Components of the crude risk ratio. Am J Epidemiol 1972;96:168–172.
18 Breslow NE, Day NE: Statistical methods in cancer research. Volume I – The analysis of case-control studies. IARC Sci Publ 5–338, 1980.
19 Wacholder S, Rothman N, Caporaso N: Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Can 2000;92:1151–1158.
20 Heiman GA, Hodge SE, Gorroochurn P, Zhang J, Greenberg DA: Effect of population stratification on case-control association studies. I. Elevation in false positive rates and comparison to confounding risk ratios (a simulation study). Hum Hered 2004;58:30–39.
21 Khlat M, Cazes MH, Genin E, Guiguet M: Robustness of case-control studies of genetic factors to population stratification: magnitude of bias and type I error. Cancer Epidemiol Biomarkers Prev 2004;13:1660–1664.
22 Xu H, Shete S: Effects of population structure on genetic association studies. BMC Genet 2005;6(suppl 1):S109.
23 Gorroochurn P, Hodge SE, Heiman G, Greenberg DA: Effect of population stratification on case-control association studies. II. False-positive rates and their limiting behavior as number of subpopulations increases. Hum Hered 2004;58:40–48.
24 Pritchard JK, Rosenberg NA: Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 1999;65:220–228.
25 Devlin B, Roeder K: Genomic control for association studies. Biometrics 1999;55:997–1004.
26 Devlin B, Roeder K, Wasserman L: Genomic control, a new approach to genetic-based association studies. Theor Popul Biol 2001;60:155–166.
27 Reich DE, Goldstein DB: Detecting association in a case-control study while correcting for population stratification. Genet Epidemiol 2001;20:4–16.
28 Devlin B, Bacanu SA, Roeder K: Genomic Control to the extreme. Nat Genet 2004;36:1129–1130.
29 Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. Genetics 2000;155:945–959.
30 Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: Association mapping in structured populations. Am J Hum Genet 2000;67:170–181.
31 Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM: Control of confounding of genetic associations in stratified populations. Am J Hum Genet 2003;72:1492–1504.
32 Wang Y, Localio R, Rebbeck TR: Evaluating bias due to population stratification in case-control association studies of admixed populations. Genet Epidemiol 2004;27:14–20.
33 Wang Y, Localio R, Rebbeck TR: Bias correction with a single null marker for population stratification in candidate gene association studies. Hum Hered 2005;59:165–175.
34 Setakis E, Stirnadel H, Balding DJ: Logistic regression protects against population structure in genetic association studies. Genome Res 2006;16:290–296.
35 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–909.
36 Patterson N, Price AL, Reich D: Population Structure and Eigenanalysis. PLoS Genet 2006;2:e190.
37 Cochran WG: Some methods of strengthening common chi-squared tests. Biometrics 1954;10:417–451.
38 Wellek S: Testing Statistical Hypotheses of Equivalence, Boca Raton, Florida, 2003.
39 Fleiss JL, Levin B, Paik MC: Statistical Methods for Rates and Proportions, ed 3. New York, Wiley, 2003.
40 Shmulewitz D, Zhang J, Greenberg DA: Case-control association studies in mixed populations: correcting using genomic control. Hum Hered 2004;58:145–153.
41 Shapiro S, Rosenberg L, Palmer LJ: Bias in case-control studies; in Elston RC, Olson JM, Palmer LJ (eds): Biostastical Genetics and Genetic Epidemiology. New York, John Wiley & Sons, 2002.
42 Marchini J, Cardon LR, Phillips MS, Donnelly P: The effects of human population structure on large genetic association studies. Nat Genet 2004;36:512–517.
43 Koller DL, Peacock M, Lai D, Foroud T, Econs MJ: False positive rates in association studies as a function of degree of stratification. J Bone Miner Res 2004;19:1291–1295.
44 Armitage P, Berry G, Matthews JNS: Statistical Methods in Medical Research, ed 4. Blackwell Science LTD, 2002.
45 Carlson CS, Eberle MA, Kruglyak L, Nickerson DA: Mapping complex disease loci in whole-genome association studies. Nature 2004;429:446–452.
46 Hartl DL, Clark AG: Principles of Population Genetics, 3rd. Sinauer Associates, Inc., 1997.
47 Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95–108.
48 Wang WY, Barratt BJ, Clayton DG, Todd JA: Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 2005;6:109–118.
49 Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 1995;57:289–300.
50 Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I: Controlling the false discovery rate in behavior genetics research. Behav Brain Res 2001;125:279–284.