An Improved Delta-Centralization Method for Population Stratification

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

[a]Division of Statistical Genetics, Department of Biostatistics, Mailman School of Public Health, Columbia University,
[b]Division of Epidemiology, New York State Psychiatric Institute, and Department of Psychiatry, Columbia College of
Physicians & Surgeons, New York, N.Y., and [c]Department of Genetics, Rutgers University, Rutgers, N.J., USA

Abstract
Dadd et al. [Hum Hered 2010;69:285–294] recently criticized
our delta-centralization (DC) method of controlling for pop-
ulation stratification (PS) and concluded that DC does not
work. To explore our method, the authors simulated data un-
der the Balding-Nichols (BN) model, which is more general
than the model we had used in our simulations. They deter-
mined that the DC method underestimated the PS parame-
ter (δ) and inflated the type I error rates when applied to BN-
simulated data, and from this they concluded that the DC
method is invalid. However, we argue that this conclusion is
premature. In this paper, we (1) show *why* δ is underestimat-
ed and type I error rates are inflated when BN-simulated data
are used, and (2) present a simple adjustment to DC that
works reasonably well for data from both kinds of simula-
tions. We also show that the adjusted DC method has appro-
priate power under a range of scenarios.

Copyright © 2011 S. Karger AG, Basel

In a recent issue of *Human Heredity*, Dadd et al. [1] questioned the appropriateness of the simulation meth-ods we employed in our two previous publications [2, 3], concerning our delta-centralization (DC) method of dealing with population stratification (PS). Based on the results of their simulations, the authors concluded that DC does not control for PS. The aim of this paper is to show that a small adjustment to the original DC method remedies the situation, giving DC reasonable type I error rates in realistic situations. We also investigate the power of this adjusted method.

Dadd et al. [1] noted that the simulations we performed did not follow the Balding-Nichols (BN) model [4], a widely used method (especially in forensics) for generat-ing allele frequencies at loci in structured populations. The major difference between the BN model and our sim-ulations, as correctly explained by Dadd et al., is that the BN model assumes that the subpopulation allele frequen-cies at both the test locus and the null loci are distributed according to a

$$\mathrm{beta}\left[\frac{1-F_{ST}}{F_{ST}}\,p_{ref}\,,\,\frac{1-F_{ST}}{F_{ST}}\left(1-p_{ref}\right)\right]$$

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

distribution, where F_{ST} is Wright's coefficient of genetic differentiation and p_{ref} is a reference allele frequency that Dadd et al. equated to *our population-average allele frequency at the test locus* [in ref. 3]. In contrast, for the model we used [in ref. 3]: (i) the subpopulation allele frequencies at the test locus were pre-specified, and (ii) p_{ref} is equated to *our subpopulation-allele frequency at the test locus*. The practical consequence of this difference is that, under our original model, the allele frequency at a null locus in a given subpopulation is closer on average to the allele frequency at the test locus for that subpopulation, compared to the BN model. Because of the different beta distributions, Dadd et al. [1] used the terms 'subpopulation allele frequency matching' for the matching we performed [in our studies, 2, 3], and 'population-level allele frequency matching' for the matching they performed.

The reason we pre-specified subpopulation allele frequencies at the test locus in (i) above is so that we could fix different values of our PS parameter (δ) and then be able to compare the performances of DC and genomic control (GC) under different levels of PS. It is the PS at the test locus that needs to be corrected, and only by fixing allele frequencies at the test locus were we able to investigate the performance of these methods under different levels of PS. Our original aim was to show that GC fails to control for PS under high levels of PS. Irrespective of the simulations used, this statement remains true and was proved in our earlier paper [2] by using distribution theory. Regarding the second difference in (ii) above, in all of our simulations (see below), we let p_{ref} be the population-average allele frequency at the test locus. We performed two types of simulations. First, we allowed the subpopulation allele frequencies at both the test and null loci to vary according to a beta distribution, with p_{ref} as the population-average allele frequency at the test locus, as in the BN model used by Dadd et al. [1]. Second, we implemented a modified BN model, which shares all the features of the first type of simulations, except that it pre-specifies the subpopulation allele frequencies at the test locus. This is done so as to be able to compare the performance of the different procedures under different levels of PS. Both of these simulation procedures allow for extra variability in allele frequencies at the null loci, and correcting for PS under these models requires an *adjusted DC method*. Before we describe this adjustment, we make an important observation about the simulation results of Dadd et al., as shown in their table 3.

Dadd et al. [1] used the BN model to compare the true values of δ against their estimated values, $\hat{\delta}_{2006}$ and $\hat{\delta}_{2007}$. However, as we explained above, the BN model assumes that the allele frequencies at the test locus are generated according to a beta distribution. Since the value of δ cannot be fixed under this type of simulation, the computation of $\hat{\delta}_{2006}$ and $\hat{\delta}_{2007}$ becomes meaningless. In other words, under each replication of the simulation procedure, a different value of δ is generated at the test locus, yielding an estimate of this quantity from the null loci. Being an average of these estimates across all replications, neither $\hat{\delta}_{2006}$ nor $\hat{\delta}_{2007}$ therefore estimates any given δ value, making the δ comparisons in table 3 of Dadd et al. [1] misleading.

We now explain why the original DC method we used [in our studies, 2, 3], does not work under the BN model, and how the method can be made to work through a simple adjustment of the DC test statistic. This adjusted method performs reasonably well under both simulation models. Consider the hypothetical scenario shown in table 1.

Both null loci 1 and 2 in table 1 match the test locus in genotype frequencies (in the controls) to within a window of ± 0.15, but only the estimated δ at null locus 1 has the same sign as the δ to be estimated at the test locus. Therefore, δ should be estimated by using the estimated δ at the first locus only, *not* by simply averaging the two estimated values. Thus, if one simply selects all null loci that match without taking into account the sign of the corresponding δ, the overall δ is always considerably underestimated since, at many of the null loci, the estimated δs are negative. In our original simulations, the sign 'mismatch' hardly ever arose because the subpopulation allele frequencies at all the null loci were fairly close to those at the test locus (as we explained above). The strategy of simply averaging at matched loci performed well for the type of simulations we performed, but is inadequate for more general simulations, such as those Dadd et al. performed. Thus, the adjusted DC statistic is

$$T_{adj}^* = \left\{ \mathrm{sign}\left(\frac{a_0}{m} - \frac{c_0}{n} \right) \sqrt{\chi_{test}^2} - \hat{\delta}_{adj} \right\}^2$$

where a_0, c_0 are the number of cases and controls out of a total of m, n, respectively, with at least one marker genotype, χ_{test}^2 is the (uncorrected) χ^2 test statistic for association at the test locus, and $\hat{\delta}_{adj}$ is the δ estimate obtained by averaging across only those matched loci whose δ estimates have the same sign as the δ at the test locus. The adjusted statistic looks the same as the original DC statistic [2, 3], except that, in the latter, δ was estimated by averaging across all matched loci.

Table 1. The 2 × 2 tables at the candidate locus and two unlinked loci

Candidate locus				Null locus 1				Null locus 2			
	M	$\overline{\text{M}}$			M	$\overline{\text{M}}$			M	$\overline{\text{M}}$	
D	30	10	40	D	33	7	40	D	15	25	40
$\overline{\text{D}}$	20	20	40	$\overline{\text{D}}$	19	21	40	$\overline{\text{D}}$	19	21	40
	$t = 20/40 = 0.5, \delta > 0$				$\hat{t}_1 = 19/40 = 0.475, \hat{\delta}_1 > 0$				$\hat{t}_2 = 19/40 = 0.475, \hat{\delta}_2 < 0$		

D and $\overline{\text{D}}$ denote cases and controls, respectively. M and $\overline{\text{M}}$ denote genotypes containing at least one marker allele and no marker allele, respectively. Both null loci match the candidate locus in genotype frequencies to within ±0.15 but only null locus 1 should be included in the estimation of δ since only $\hat{\delta}_1$ has the same sign as the candidate δ.

We used two types of simulation procedures: a full BN simulation (the same as Dadd et al. used), and a modified BN simulation, which we will describe below. With these, we show that when null loci are selected that both match (say, to within ±0.15 and ±0.10) and whose δ estimates have the correct sign (i.e. have the same sign as the δ at the test locus), DC does maintain reasonable type I error rates compared to GC. In all of our simulations, we performed matching at the population level, as in Dadd et al. [1]. All scenarios we used are listed in the Appendix.

First, we used the full BN simulation procedure as Dadd et al. [1] did, with exactly the same parameters. Complete simulation details can be found in that paper. We evaluated the DC test using two procedures for estimating δ: (i) as in Dadd et al. [1], simply by averaging across matched loci (the original DC method, using $\hat{\delta}_{2007}$), and (ii) as we explained above, by averaging across matched loci 'that have the correct sign' (the adjusted DC method, using $\hat{\delta}_{adj}$). We also evaluated the uncorrected χ^2 test (UN), GC using a mean correction (GC$_{mu}$), and F-based GC (GCF). (Further details on the latter two methods can be found in Gorroochurn et al. [2] and Dadd et al. [1].) We tested exactly the same scenarios as Dadd et al. used in their tables 2 and 3, as shown in our table 2. The only difference is that we used 5,000 simulations instead of 10,000. We also varied the number of null loci ($L = 100, 150, 200$), and we tried two different matching criteria (±0.15 and ±0.10), as opposed to Dadd et al. who used 100 null loci and ±0.15 matching.

Table 2 confirms Dadd et al.'s claim that, for the original DC method (i.e. using $\hat{\delta}_{2007}$), the type I errors (Type I ($\hat{\delta}_{2007}$)) are substantially larger than the nominal α, for BN-simulated data. However, the adjusted DC method (i.e. using $\hat{\delta}_{adj}$) yields improved error rates, although the results are somewhat mixed. Although the adjusted DC method no longer has the overly anti-conservative error rates of the original DC method, it is very conservative for extremely small δs (≈ 0.5, i.e. when there is very little PS). On the other hand, GC does not become very conservative for small δs.

Similar results were obtained when the alleles at null loci are generated according to a fairly broad range of distributions (results not shown). Some of the distributions we examined include: (i) a uniform (0.1, 0.9) distribution, and (ii) a BN model with p_{ref} generated from a uniform (0.1, 0.9) distribution. We also investigated simulations where those discarded markers that matched but whose δ gave the incorrect sign had their δ signs changed (results not shown). Overall, this did not change our main conclusions. Note also that table 2 contains neither the true δ values nor their estimates. Again, since the allele frequencies at the test locus vary under each replication of the simulation procedure under the BN model, calculation of these values is meaningless. Moreover, we note that, overall, there are no significant changes in the performance of the adjusted method either when the number of null markers is increased or when the window is made narrower.

We also investigated the power of the adjusted DC method when applied to BN-simulated data. This uses the same parameters as for the type I errors, except that, within each subpopulation, the marker genotype in the cases was assumed to have a frequency 1.4 times that in the controls. Table 3 shows that all three correction methods (GC$_{mu}$, GCF and adjusted DC) have comparable power values, although GC$_{mu}$ and GCF have slightly higher power.

Secondly, we performed modified BN simulations. These were the same as the first, except that we pre-specified the allele frequencies at the test locus. Again, this allows us to fix the amount of PS at the test loci, and en-

Table 2. Type I error rates under the full BN model at $\alpha = 0.05$ and for number of null loci $L = (100, 150, 200)$

	Scenario			
	17′	15′	13′	8′
(a) Window = ±0.15				
Type I (UN)	(0.228, 0.220, 0.234)	(0.152, 0.169, 0.172)	(0.107, 0.104, 0.120)	(0.057, 0.062, 0.058)
Type I (GC$_{mu}$)	(0.049, 0.054, 0.048)	(0.049, 0.056, 0.054)	(0.050, 0.048, 0.060)	(0.051, 0.056, 0.050)
Type I (GCF)	(0.046, 0.052, 0.046)	(0.046, 0.054, 0.053)	(0.046, 0.046, 0.057)	(0.048, 0.055, 0.048)
Type I ($\hat{\delta}_{2007}$)	(0.187, 0.182, 0.196)	(0.120, 0.130, 0.138)	(0.102, 0.106, 0.108)	(0.060, 0.064, 0.062)
Type I ($\hat{\delta}_{adj}$)	(0.041, 0.032, 0.034)	(0.020, 0.019, 0.021)	(0.016, 0.013, 0.013)	(0.008, 0.006, 0.008)
(b) Window = ±0.10				
Type I (UN)	(0.222, 0.239, 0.226)	(0.176, 0.150, 0.154)	(0.110, 0.121, 0.104)	(0.059, 0.056, 0.052)
Type I (GC$_{mu}$)	(0.038, 0.045, 0.054)	(0.065, 0.045, 0.056)	(0.052, 0.058, 0.048)	(0.056, 0.048, 0.046)
Type I (GCF)	(0.034, 0.042, 0.052)	(0.061, 0.044, 0.056)	(0.050, 0.057, 0.046)	(0.052, 0.048, 0.045)
Type I ($\hat{\delta}_{2007}$)	(0.168, 0.174, 0.164)	(0.124, 0.098, 0.108)	(0.112, 0.118, 0.098)	(0.065, 0.059, 0.051)
Type I ($\hat{\delta}_{adj}$)	(0.038, 0.036, 0.038)	(0.026, 0.020, 0.019)	(0.012, 0.012, 0.013)	(0.009, 0.006, 0.006)

Type I ($\hat{\delta}_{2007}$) and Type I ($\hat{\delta}_{adj}$) are the type I error rates using $\hat{\delta}_{2007}$ and $\hat{\delta}_{adj}$, respectively, where $\hat{\delta}_{adj}$ is the estimated δ by matching (at the population level) and making sure all matched loci have the correct sign. The four scenarios represent situations with no true association (but with PS) and are listed in the Appendix.

Table 3. Power under the full BN model at $\alpha = 0.05$

	Scenario			
	18′	16′	14′	9′
Power (UN)	0.681	0.704	0.361	0.490
Power (GC$_{mu}$)	0.418	0.512	0.253	0.471
Power (GCF)	0.412	0.504	0.244	0.463
Power ($\hat{\delta}_{2007}$)	0.680	0.705	0.362	0.491
Power ($\hat{\delta}_{adj}$)	0.396	0.426	0.128	0.243

Power ($\hat{\delta}_{2007}$) and Power ($\hat{\delta}_{adj}$) are the power using $\hat{\delta}_{2007}$ and $\hat{\delta}_{adj}$, respectively, where $\hat{\delta}_{adj}$ is the estimated δ by matching (at the population level) and making sure all matched loci have the correct sign. The four scenarios represent situations with true associations (and with PS) and are listed in the Appendix. They use the same parameters as in table 2, except that within each subpopulation, marker genotype in cases is assumed to have a frequency 1.4 times that in controls.

ables correction procedures to be compared under different levels of PS. Tables 4 and 5 give type I error rates and power values, respectively. Note that here, as opposed to previously, there is a true δ value and it is meaningful to compare it with its estimated values. Both tables show that the strategy of matching and keeping the correct sign results in much better δ estimates and type I errors than

the strategy of matching only. However, the overall results are again somewhat mixed. The adjusted DC method maintains reasonable type I error rates except for very high levels of PS ($\delta = 2.00$) when it becomes anti-conservative. On the other hand, GC maintains reasonable type I error rates except for moderate to high levels of PS ($\delta \geq 1.0$) when it becomes even more anti-conservative than the adjusted DC method. We previously proved GC performs well only for small δs (i.e. small levels of PS) [2]. Again, there are no significant changes in the performance of the adjusted method either when the number of null markers is increased or when the window is made narrower.

We have shown that, when δ is estimated by matching at null loci and by selecting only those null loci whose δ estimates have the same sign as δ at the test locus, the problem recognized by Dadd et al. [1] is resolved to a reasonable extent. We are grateful to Dadd et al. for pointing this problem out.

The other important issue concerns the value of the GC method. Dadd et al. [1] state that GC is computationally simple and should be favored. However, as several authors [5–7] have pointed out, GC applies a uniform correction factor for all test loci, despite the fact that these loci may have different allele frequencies and thus may require locus-specific correction factors. In contrast, DC, though computationally more intensive, provides a locus-specific correction method.

Table 4. True value of δ and type I error rates under a modified BN model at $\alpha = 0.05$ and for number of null loci $L = (100, 150, 200)$

	Scenario			
	17	15	13	8
(a) Window = ±0.15				
True δ	2.00	1.50	1.00	0.500
$\hat{\delta}_{2007}$	(0.055, 0.066, 0.064)	(0.050, 0.042, 0.049)	(0.029, 0.030, 0.029)	(0.004, 0.001, −0.001)
$\hat{\delta}_{adj}$	(1.31, 1.31, 1.31)	(1.12, 1.12, 1.13)	(0.992, 1.00, 0.998)	(0.850, 0.854, 0.850)
Type I (UN)	(0.506, 0.530, 0.522)	(0.336, 0.326, 0.322)	(0.161, 0.174, 0.168)	(0.082, 0.074, 0.076)
Type I (GC_{mu})	(0.111, 0.121, 0.125)	(0.120, 0.116, 0.110)	(0.087, 0.090, 0.088)	(0.071, 0.066, 0.066)
Type I (GCF)	(0.104, 0.117, 0.120)	(0.114, 0.110, 0.106)	(0.082, 0.084, 0.084)	(0.065, 0.064, 0.066)
Type I ($\hat{\delta}_{2007}$)	(0.484, 0.494, 0.495)	(0.312, 0.305, 0.299)	(0.151, 0.160, 0.156)	(0.079, 0.071, 0.072)
Type I ($\hat{\delta}_{adj}$)	(0.090, 0.104, 0.099)	(0.048, 0.057, 0.058)	(0.024, 0.032, 0.030)	(0.014, 0.011, 0.026)
(b) Window = ±0.10				
True δ	2.00	1.50	1.00	0.500
$\hat{\delta}_{2007}$	(0.091, 0.099, 0.097)	(0.082, 0.084, 0.068)	(0.032, 0.033, 0.037)	(0.003, −0.002, −0.002)
$\hat{\delta}_{adj}$	(1.28, 1.27, 1.27)	(1.09, 1.09, 1.09)	(1.01, 1.01, 1.01)	(0.859, 0.854, 0.857)
Type I (UN)	(0.496, 0.504, 0.510)	(0.320, 0.330, 0.334)	(0.163, 0.177, 0.178)	(0.070, 0.085, 0.088)
Type I (GC_{mu})	(0.118, 0.126, 0.116)	(0.120, 0.110, 0.104)	(0.088, 0.089, 0.087)	(0.062, 0.074, 0.078)
Type I (GCF)	(0.111, 0.120, 0.110)	(0.116, 0.105, 0.100)	(0.084, 0.087, 0.084)	(0.056, 0.070, 0.077)
Type I ($\hat{\delta}_{2007}$)	(0.470, 0.465, 0.470)	(0.280, 0.281, 0.270)	(0.146, 0.165, 0.158)	(0.074, 0.082, 0.088)
Type I ($\hat{\delta}_{adj}$)	(0.102, 0.101, 0.099)	(0.066, 0.047, 0.053)	(0.031, 0.022, 0.023)	(0.014, 0.019, 0.017)

This model shares all features of the usual BN model, except that it pre-specifies the subpopulation allele frequencies at the test locus; however, matching is at the population level, as in Dadd et al. [1].

Type I ($\hat{\delta}_{2007}$) and Type I ($\hat{\delta}_{adj}$) are the type I error rates using $\hat{\delta}_{2007}$ and $\hat{\delta}_{adj}$, respectively, where $\hat{\delta}_{adj}$ is the estimated δ by matching and making sure all matched loci have the correct sign. The four scenarios represent situations with no true association (but with PS) and are listed in the Appendix.

Table 5. True value of δ and power under the modified BN model at $\alpha = 0.05$

	Scenario			
	18	16	14	9
True δ	2.00	1.50	1.00	0.500
$\hat{\delta}_{2007}$	0.058	0.044	0.029	0.000
$\hat{\delta}_{adj}$	1.31	1.12	0.997	0.852
Power (UN)	1.00	1.00	0.993	1.00
Power (GC_{mu})	0.999	0.996	0.976	1.00
Power (GCF)	0.999	0.997	0.974	1.00
Power ($\hat{\delta}_{2007}$)	1.00	1.00	0.993	1.00
Power ($\hat{\delta}_{adj}$)	0.992	0.992	0.920	0.997

This model shares all features of the usual BN model, except that it pre-specifies the subpopulation allele frequencies at the test locus; however, matching is at the population level, as in Dadd et al. [1].

Power ($\hat{\delta}_{2007}$) and Power ($\hat{\delta}_{adj}$) are the power using $\hat{\delta}_{2007}$ and $\hat{\delta}_{adj}$, respectively, where $\hat{\delta}_{adj}$ is the estimated δ by matching and making sure all matched loci have the correct sign. The four scenarios represent situations with true associations (and with PS) and are listed in the Appendix. They use the same parameters as in table 4, except that within each subpopulation, marker genotype in cases is assumed to have a frequency 1.4 times that in controls.

In conclusion, we strongly recommend the strategy of matching at null loci and keeping the correct sign in all future applications of the adjusted DC method. GC does not control for PS under a broad range of PS levels (from moderate to high) under the modified BN simulation model, whereas the adjusted DC method we introduce here does not control for PS for very high levels of PS. In the latter case, the adjusted DC method is anti-conservative but less so than GC. For very small levels of PS, the adjusted DC method is much more conservative than GC.

Appendix

Configuration numbers (#) are as indicated in tables 2–5, and corresponding values of π (subpopulation size), d (subpopulation disease prevalence), and r (subpopulation genotype frequenciesat test locus) for $K = 2$ subpopulations with m (cases) = n (controls) = 200. The first set (8′–18′) represents simulations under the full BN model, the second set (8–18) under the modified BN model. For each pair of simulations (e.g. 8′ and 9′, 13′ and 14′, 17′ and 18′, etc.) the first represents simulations for type I error, the second for power.

#	π	d	r (cases)	r (controls)
8′	(0.7648, 0.2352)	(0.1908, 0.1416)	NA	NA
9′	(0.7648, 0.2352)	(0.1908, 0.1416)	NA	NA
13′	(0.5241, 0.4759)	(0.2818, 0.1654)	NA	NA
14′	(0.5241, 0.4759)	(0.2818, 0.1654)	NA	NA
15′	(0.2620, 0.7380)	(0.2502, 0.05855)	NA	NA
16′	(0.2620, 0.7380)	(0.2502, 0.05855)	NA	NA
17′	(0.2620, 0.7380)	(0.2502, 0.03326)	NA	NA
18′	(0.2620, 0.7380)	(0.2502, 0.03326)	NA	NA
8	(0.7648, 0.2352)	(0.1908, 0.1416)	(0.5614, 0.1484)	(0.5614, 0.1484)
9	(0.7648, 0.2352)	(0.1908, 0.1416)	(0.8421, 0.2226)	(0.5614, 0.1484)
13	(0.5241, 0.4759)	(0.2818, 0.1654)	(0.4298, 0.1505)	(0.4298, 0.1505)
14	(0.5241, 0.4759)	(0.2818, 0.1654)	(0.6447, 0.2258)	(0.4298, 0.1505)
15	(0.2620, 0.7380)	(0.2502, 0.05855)	(0.4770, 0.2882)	(0.4770, 0.2882)
16	(0.2620, 0.7380)	(0.2502, 0.05855)	(0.7, 0.45)	(0.4770, 0.2882)
17	(0.2620, 0.7380)	(0.2502, 0.03326)	(0.4770, 0.2882)	(0.4770, 0.2882)
18	(0.2620, 0.7380)	(0.2502, 0.03326)	(0.7155, 0.4323)	(0.4770, 0.2882)

NA = Not applicable, because under the full BN model, allele frequencies at the test locus are not fixed.

References

1 Dadd T, Lewis CM, Weale ME: Delta-centralization fails to control for population stratification in genetic association studies. Hum Hered 2010;69:285–294.

2 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.

3 Gorroochurn P, Hodge SE, Heiman GA, Greenberg DA: A unified approach for quantifying, testing and correcting population stratification in case-control association studies. Hum Hered 2007;64:149–159.

4 Balding DJ, Nichols RA: A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. Genetica 1995;96:3–12.

5 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.

6 Tiwari HK, Barnholtz-Sloan J, Wineinger N, Padilla MA, Vaughan LK, Allison DB: Review and evaluation of methods correcting for population stratification with a focus on underlying statistical principles. Hum Hered 2008;66:67–86.

7 Gorroochurn P, Hodge SE, Heiman GA, Greenberg DA: Comments on: review and evaluation of methods correcting for population stratification with a focus on underlying statistical principles. Hum Hered 2009;67:145–146.