Report

Improved Power Offered by a Score Test for Linkage Disequilibrium Mapping of Quantitative-Trait Loci by Selective Genotyping

Chris Wallace,[1] Juliet M. Chapman,[2] and David G. Clayton[3]

[1]Department of Clinical Pharmacology, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, and [2]Epidemiology and Public Health, London School of Hygiene and Tropical Medicine, London; and [3]Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom

Selective genotyping is used to increase efficiency in genetic association studies of quantitative traits by genotyping only those individuals who deviate from the population mean. However, selection distorts the conditional distribution of the trait given genotype, and such data sets are usually analyzed using case-control methods, quantitative analysis within selected groups, or a combination of both. We show that Hotelling's T^2 test, recently proposed for association studies of one or several tagging single-nucleotide polymorphisms in a prospective (i.e., trait given genotype) design, can also be applied to the retrospective (i.e., genotype given trait) selective-genotyping design, and we use simulation to demonstrate its improved power over existing methods.

Selective genotyping, first proposed for use in linkage studies by Lander and Botstein (1989), can also be used to increase the efficiency of genetic studies of quantitative traits. In selective genotyping, instead of individuals being sampled at random from a population, those whose trait value deviates from the population mean are preferentially sampled. Its application to the area of quantitative-trait association studies has been studied theoretically and with the use of simulation (Slatkin 1999; van Gestel et al. 2000; Xiong et al. 2002*a*; Chen et al. 2005), and it has recently been employed to identify association between attention-deficit/hyperactivity disorder and the dopamine transporter gene, *DAT1* (Cornish et al. 2005). However, case-control methods (e.g., testing for allele-frequency differences between selected groups) are commonly used for analysis, and information from the quantitative-trait values is neglected.

Slatkin (1999) proposed a method that incorporates trait values when one allele of the marker locus is sufficiently rare that homozygotes may be ignored. His strategy is to recruit a truncation-selected (TS) sample of individuals whose trait value exceeds some predetermined threshold and to compare them with a random sample. Slatkin (1999) noted that two tests that are asymptotically independent could be performed. The first test (S_1) is a standard χ^2_1 test for differences in heterozygous/homozygous frequencies between the selected group and the random group. The second test (S_2) is a *t* test for differences in trait means between individuals who are heterozygous and those who are homozygous in the selected group. Note that, in both tests, any individuals homozygous for the rare allele are dropped.

Since these tests are asymptotically independent under the null hypothesis, Slatkin proposed a third test that uses Fisher's product-of-*P*-values method to calculate a combined test statistic:

$$S_3 = -2 \ln P_{S_1} - 2 \ln P_{S_2} ,$$

where P_{S_1} and P_{S_2} are the *P* values from the first and second tests, respectively, and S_3 has a χ^2_4 distribution.

Recently in this journal, Chen et al. (2005) described two alternative strategies for selective genotyping. The first, extended truncation selection (TS-II), has been used in other studies and compares individuals in the upper and lower tails of the trait distribution. The TS and TS-II methods require that thresholds for inclusion be determined in advance, which is not always possible unless the trait distribution in the population is well known. Chen et al. (2005) proposed a practical alternative, extreme rank selection (ERS), in which small "batches" of (perhaps 10) individuals are surveyed, and the 2 individuals with the highest and lowest trait values are selected for genotyping. Chen et al. showed that, using

Received November 4, 2005; accepted for publication December 16, 2005; electronically published January 5, 2006.

Address for correspondence and reprints: Dr. Chris Wallace, Clinical Pharmacology, William Harvery Research Institute, Barts and the London School of Medicine and Dentistry, Charterhouse Square, London EC1M 6BQ, United Kingdom. E-mail: c.wallace@qmul.ac.uk

Am. J. Hum. Genet. 2006;78:498–504. 0002-9297/2006/7803-0017$15.00

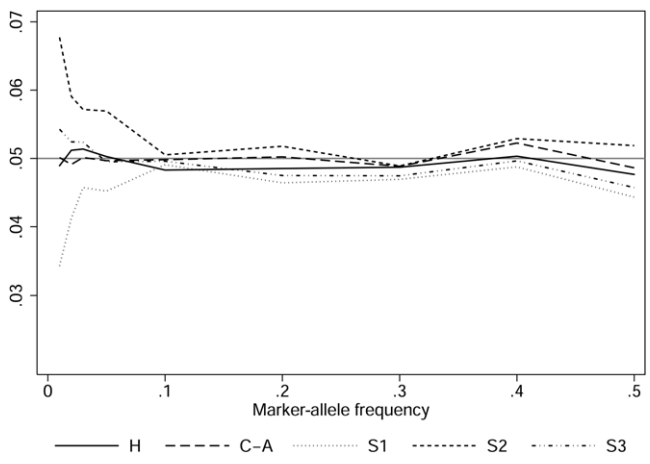

Figure 1 Type I error of the five tests at the 5% level. H = Hotelling's T^2 test; C-A = Cochran-Armitage test; S1, S2, and S3 = Slatkin's first, second, and third tests, respectively.

Slatkin's three test statistics, both ERS and TS-II have increased power over the TS strategy.

However, Slatkin's statistics still do not make use of all information in the sample—individuals who are homozygous for the rare allele are discarded, and the distribution of trait values in the low-trait-value group or random group and in the sample as a whole is not used. The motivation behind the use of case-control–type analysis or statistics, such as those proposed by Slatkin (1999), is that the selective-genotyping strategy distorts the distribution of the trait conditional on genotype. However, in case-control studies of dichotomous traits, analysis using the prospective (i.e., trait given genotype) likelihood has been shown to be valid even for retrospective (i.e., genotype given trait) sampling designs (Prentice and Pyke 1979).

A generalized Hotelling's T^2 test has recently been proposed for the analysis of quantitative and qualitative traits with the use of tagging SNPs (tSNPs) (Xiong et al. 2002*b*; Chapman et al. 2003). This test has the advantage that it can accommodate one or multiple tSNPs and uses all quantitative information about the trait in the sample. Following Chapman et al. (2003), let individual i with trait value Y_i have genotypes X_i^+ and Z_i^+ at the marker and trait loci, respectively, each of which is coded 0, 1, or 2. Note that X_i^+ may be a vector in the case of multiple tSNPs, that the scalar Z_i^+ is unobserved, and that X_i^+ and Y depend on each other only through Z_i. Assume that a region of high linkage disequilibrium (LD) is under study and that the tSNPs have been chosen with the intention that they can predict Z_i^+. Thus, Z^+ may be predicted by $E(Z^+) = \beta^T X^+$. Under a codominant model, the dependence of Y on Z can be described by a generalized linear model of the form $\text{link}(Y) = \gamma_0 + \gamma_1 Z^+$, with the null hypothesis of no association defined by $\gamma_1 = 0$. Chapman et al. (2003) showed that, for the prospective likelihood

$$L = \prod_i P(Y_i|X_i^+) = \prod_i \sum_{Z_i^+} P(Y_i|Z_i^+)P(Z_i^+|X_i^+) ,$$

the score is

$$U = \sum_i Y_i(X_i^+ - \overline{X}^+) \qquad (1)$$

and that the variance of U can be estimated by

$$V = \frac{1}{n-1}\sum_i (Y_i - \overline{Y})^2 \sum_i (X_i^+ - \overline{X}^+)^T(X_i^+ - \overline{X}^+) . \qquad (2)$$

The test statistic is then given by $T_H^2 = U^T V^{\oplus} U$, where $V^{\oplus}$ denotes the generalized inverse of V. T_H^2 may be referred to as a χ^2 distribution with df equal to the rank of V (Rao 1962).

Because selection, S, depends on the phenotype but is conditionally independent of genotype given Y, $P(X|Y,S) = P(X|Y)$, and it is clear that the "retrospective" likelihood

$$L = \prod_i P(X_i^+|Y_i) \qquad (3)$$

is appropriate under selection. However, this expression is relatively complicated in general. On the other hand, $P(Y|X,S) \neq P(Y|X)$, so that the prospective likelihood is not obviously applicable. In appendix A, however, we show that the score test for a normally distributed Y based on the prospective likelihood, ignoring selection, is asymptotically equivalent to the score test based on the retrospective likelihood and, therefore, is appropriate for analysis in studies with a selective-genotyping design. Even though the distribution of the observed trait values will be markedly nonnormal because of the sampling design, in large samples U is approximately multivariate normal. Note also that this χ^2 test is an approximation of the permutation distribution obtained by permuting the Y vector, so that, in moderately large samples, the test does not depend on distributional assumptions about the trait (Stuart et al. 1999). A closer approximation is given by the usual distribution of Hotelling's T^2, which is

$$\frac{p(n-1)}{n-p} F_{(p,n-p)} ,$$

where n is the number of observations and p is the number of marker loci.

We now investigate the power of T_H^2 in a selective-

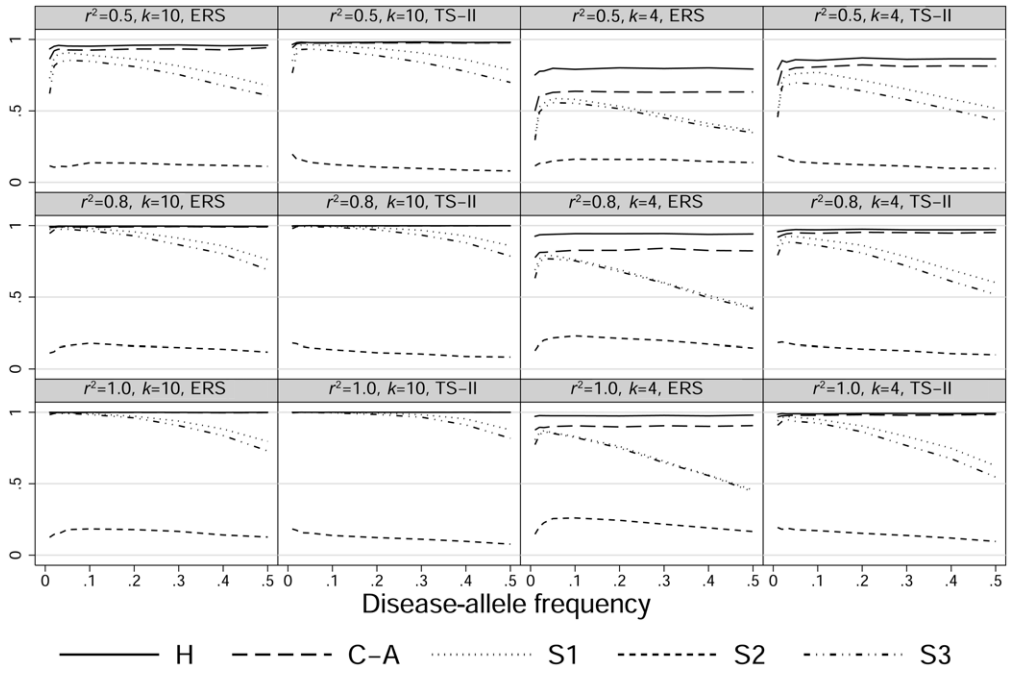

Figure 2 Power of the five tests under codominant models. k is the batch size for an ERS strategy and corresponds to selecting the most extreme 1/kth quantiles under a TS-II strategy. H = Hotelling's T^2 test; C-A = Cochran-Armitage test; S1, S2, and S3 = Slatkin's first, second, and third tests, respectively.

genotyping design using simulation, under both TS-II and ERS, and also calculate Slatkin's S_1, S_2, S_3 and a Cochran-Armitage test (χ^2_{CA}) for trend in allele frequencies between groups, for comparison. We chose to employ the latter test instead of the more common χ^2_1 test for allele-frequency difference, because the two are equivalent when Hardy-Weinberg equilibrium (HWE) holds, but, unlike the standard χ^2_1 test, χ^2_{CA} is valid even when a sample deviates from HWE (Sasieni 1997).

We assume only a single tSNP locus, partly for simplicity and partly because Slatkin's tests are only defined for a single marker. Note that this means that V in equation (2) always has a proper inverse, so we use this in place of $V^{\oplus}$. We extend the range of models studied by Chen et al. (2005) to include more common alleles and to allow deviation from an additive model. Our genetic models are defined by the frequencies of the minor alleles at the trait and marker loci, P_Q and P_M, respectively; by the heritability of the trait due to locus Q, h^2; by the relative dominance of the trait, $d \in [0,1]$, where $d = 0.5$ corresponds to codominance; and by the LD between the trait and marker loci, which we summarize by the coefficient of determination r^2, which is defined as

$$\frac{(\pi_{11}\pi_{22} - \pi_{12}\pi_{21})^2}{(\pi_{11} + \pi_{12})(\pi_{21} + \pi_{22})(\pi_{11} + \pi_{21})(\pi_{12} + \pi_{22})},$$

where π_{ij} is the frequency of haplotype ij. r^2 is a useful measure of LD, and it has been shown that typing a SNP that "tags" a causative SNP with a particular r^2 value requires an expected sample size increase of $1/r^2$ in association studies (Pritchard and Przeworski 2001). Although the decay of r^2 with physical distance depends on local genetic architecture, several software packages exist to assist in choosing tSNPs according to the r^2 measure (e.g., see de Bakker et al. 2005).

We consider sampling 1,000 individuals from each extreme. For the total screening size to remain equal between strategies, the ERS approach with a batch size of k is compared with the TS-II approach in which the upper and lower τth quantiles are sampled, where $\tau = 1/k$. For each model, we simulated 5,000 data replicates. We considered P_Q in the range 0.01–0.5; an h^2 of 0 (for type I error rates), 0.01, or 0.02; an r^2 of 0.5, 0.8, or 1.0; and a k of 4 or 10 (corresponding to a τ of 0.25 or 0.1).

Preliminary analysis showed that, as might be expected, the power of the Cochran-Armitage and Hotelling's tests did not vary with P_M when P_Q and r^2 were fixed. However, the power of Slatkin's tests for fixed P_Q and r^2 decreased as P_M increased, because this required an increasing proportion of individuals (rare-allele homozygotes) to be dropped. To avoid unfair comparison

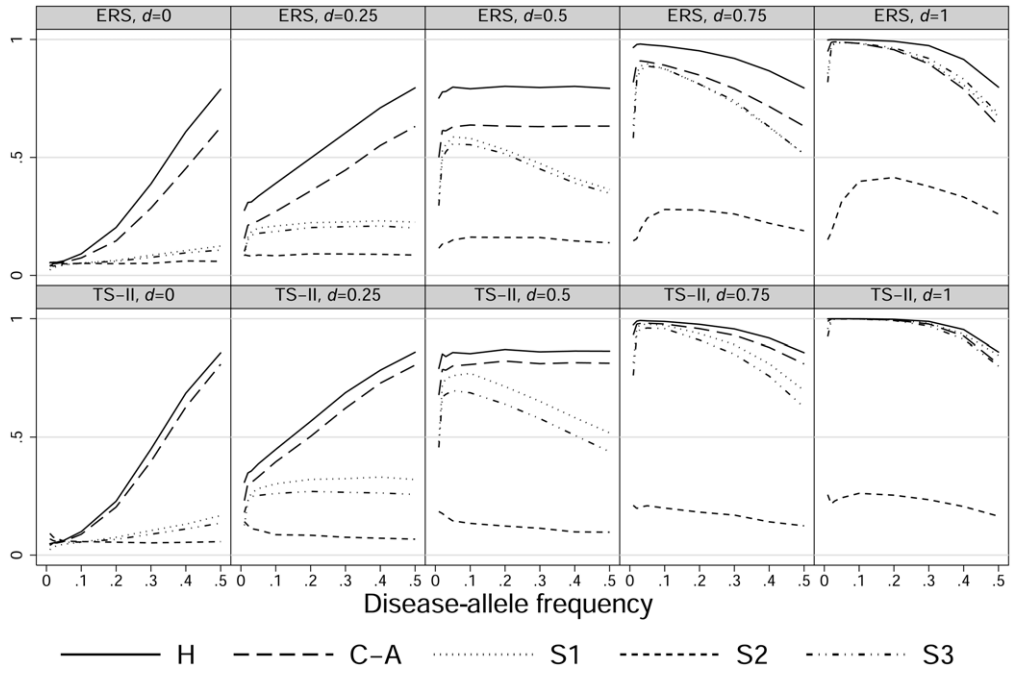

Figure 3 Effects of deviation from a codominant model on the power of the five tests, for ERS and TS-II strategies. The dominance parameter d ranges from 0 (recessive) to 1 (dominant). H = Hotelling's T^2 test; C-A = Cochran-Armitage test; S1, S2, and S3 = Slatkin's first, second, and third tests, respectively.

with Slatkin's tests, we fixed P_M at its minimum value given P_Q and r^2: $P_{M,\mathrm{MIN}} = P_Q r^2/[1 + P_Q(r^2 - 1)]$.

Simulations showed that all the tests had good control of the 5% error rate when P_M was in the range 0.05–0.5 (fig. 1). However, like Chen et al. (2005), we found evidence that Slatkin's S_2 test tended to be anticonservative when P_M was <0.05, whereas S_1 tended to be conservative; no such evidence was observed for S_3, χ^2_{CA}, or T^2_{H}. The poor control of the type I error rate for S_2 and S_1 is likely the result of small sample violations of asymptotic theory. For $P_M = 0.01$, we expect the heterozygote group to contain only 20 individuals, even when a total of 1,000 individuals are sampled in each extreme-trait-value group. χ^2 tests are known to fail to maintain type I error rates when the count in any one cell is small. For t tests, the selection leads to violation of the normality assumption, and it is only for larger values of P_M (and hence for larger sample sizes) that the central limit theorem ensures that the asymptotic theory holds.

The results of power calculations for $h^2 = 0.01$ under codominant ($d = 0.5$) models are shown in figure 2. The results for higher h^2 values displayed a similar pattern but achieved 100% power across all tests except S_1 a considerable proportion of the time (data not shown).

Using codominant models, we found that T^2_{H} was uniformly more powerful than Slatkin's tests, with χ^2_{CA} between these in terms of power. We suspect that the generally reduced power of Slatkin's tests results largely from neglecting the information from the rare homozygotes, who are likely to have the most-extreme trait values. Additionally, for S_3, the two P values that contribute to the test are given equal weight, which is not optimal. As smaller fractions of the population are studied (i.e., as k increases), so trait variation within a selected group will decrease and the test for trait-genotype association will have less information.

Figure 3 shows the effect of deviation from a codominant model, with $k = 4$, $r^2 = 0.5$, and $h^2 = 0.01$. For rare disease alleles, power tends to increase with increasing dominance and to decrease with increasing recessiveness. As P_Q approaches 0.5, power remains relatively constant across models. T^2_{H} remains uniformly the most powerful.

The pattern for changes in dominance is as might be expected for Slatkin's tests, which only examine differences between common homozygotes and heterozygotes. Thus, under dominant models, the two trait extremes are studied, and the power of Slatkin's tests approach that of Hotelling's test for rare alleles, but their power falls as the allele frequency increases and the sample size thus decreases. Conversely, for recessive models, the expected trait value does not differ between homozygotes

and heterozygotes, and so the power of Slatkin's tests approaches their type I error rates.

Both T_H^2 and χ_{CA}^2 are most powerful under codominant models, so the increase in power for rarer alleles with increasing dominance is perhaps surprising. However, the pattern can be understood if we consider the changing effects of selective genotyping with increasing dominance. Selective genotyping increases the proportion of sampled individuals who carry the extreme genotypes (in the case of codominant models, these are the homozygotes). However, for rare alleles, modest heritability, and modest thresholds (relatively small k), the low frequency of rare homozygotes means that a proportion of the selected individuals will be heterozygotes with less extreme phenotypes. As dominance increases, heterozygotes have more-extreme phenotypes, and so power increases. Conversely, as the rare allele operates in a more recessive manner, heterozygotes become phenotypically more similar to common homozygotes, and power decreases.

More generally, and as expected, power always increases with increasing LD between the marker and the disease locus and with more-stringent selective-genotyping procedures. Across the range of models considered here, TS-II strategies tend to be more powerful than ERS ones, but this must be weighed against the costs of a screening survey to determine TS-II thresholds, which may be required in some cases. Our results demonstrate that, with the use of a generalized Hotelling's T^2 test, ERS remains a practical alternative to TS-II, with only a relatively small difference in power, considering the cost of a potentially required survey to determine TS-II limits.

We have shown that T_H^2 is more powerful than other tests used in selective-genotyping designs. It also extends to the study of multiple tSNPs by jointly analyzing the tSNPs (rather than their haplotypes). This has been shown to be the optimal strategy in regions of high LD because it minimizes the df of the test (Clayton et al. 2004). We therefore recommend that studies employing selective-genotyping designs use this test statistic for their analysis.

Acknowledgments

C.W. is supported by British Heart Foundation grant FS/05/061/19501. J.M.C. is supported by Wellcome Trust grant EPNCBC49. D.G.C. is supported by the Wellcome Trust and the Juvenile Diabetes Research Foundation.

Appendix A

Chapman et al. (2003) gave an expression for the score test statistic for a prospective likelihood. We begin by following Chapman (2005) to find an expression for the retrospective score. Note that the likelihood in equation (3) may be rewritten as

$$\begin{aligned} L(X^+|Y) &= \frac{L(Y|X^+)L(X^+)}{L(Y)} \\ &= \frac{\sum_{Z^+} L(Y|Z^+)L(Z^+|X^+)L(X^+)}{\sum_{Z^+,X^+} L(Y|Z^+)L(Z^+|X^+)L(X^+)} \end{aligned}$$

so that the first derivative of the log likelihood is

$$\begin{aligned} \frac{\partial l(X^+|Y)}{\partial \gamma_1} &= \frac{\sum_{Z^+} \frac{\partial L(Y|Z^+)}{\partial \gamma_1} L(Z^+|X^+)L(X^+)}{\sum_{Z^+} L(Y|Z^+)L(Z^+|X)L(X^+)} - \frac{\sum_{Z^+,X^+} \frac{\partial L(Y|Z^+)}{\partial \gamma_1} L(Z^+|X^+)L(X^+)}{\sum_{Z^+,X^+} L(Y|Z^+)L(Z^+|X)L(X^+)} \\ &= \frac{\sum_{Z^+} \frac{\partial l(Y|Z^+)}{\partial \gamma_1} L(Y|Z^+)L(Z^+|X^+)L(X^+)}{L(Y,X^+)} - \frac{\sum_{Z^+,X^+} \frac{\partial l(Y|Z^+)}{\partial \gamma_1} L(Y|Z^+)L(Z^+|X^+)L(X^+)}{L(Y)} \\ &= \sum_{Z^+} \frac{\partial l(Y|Z^+)}{\partial \gamma_1} L(Z^+|Y,X^+) - \sum_{Z^+,X^+} \frac{\partial l(Y|Z^+)}{\partial \gamma_1} L(Z^+,X^+|Y) \ . \end{aligned}$$

For a quantitative trait, let us assume $Y \sim N(\gamma_0 + \gamma_1 Z^+, \sigma^2)$. Then,

$$\frac{\partial l(Y|Z^+)}{\partial \gamma_1} = \sigma^{-2} \sum_i (Y_i - \gamma_0 - \gamma_1 Z_i^+) Z_i^+ \ .$$

So, under H_0, $\gamma_1 = 0$ and

$$\begin{aligned}\frac{\partial l(X^+|Y)}{\partial \gamma_1} &= E_{[Z^+|X^+]}\left[\sigma^{-2} \sum_i (Y_i - \gamma_0) Z_i^+\right] - E_{[Z^+, X^+]}\left[\sigma^{-2} \sum_i (Y_i - \gamma_0) Z_i^+\right] \\ &= \sigma^{-2} \sum_i (Y_i - \gamma_0) E(Z_i^+|X_i^+) - \sigma^{-2} \sum_i (Y_i - \gamma_0) E(Z_i^+) \ .\end{aligned}$$

Using a profile likelihood approach, we replace the nuisance parameters γ_0 and σ^2 by their maximum-likelihood estimates under the null ($\overline{Y}$ and $\hat{\sigma}^2 = \sum (Y_i - \overline{Y})^2/(n-1)$, respectively). Then, after it is noted that $E(Z_i^+)$ is constant for all i, the score at H_0, $\hat{u}$, may be given by

$$\begin{aligned}\frac{\partial l(X^+|Y)}{\partial \gamma_1} &= \beta^T \hat{\sigma}^{-2} \sum_i (Y_i - \overline{Y}) X_i^+ \\ &= \beta^T \hat{\sigma}^{-2} \sum_i Y_i (X_i^+ - \overline{X}^+) = \beta^T \hat{\sigma}^{-2} U \ .\end{aligned}$$

In a retrospective design, Y_i is fixed and X_i^+ is random, so that

$$\mathrm{Var}\,(\hat{u}) = \hat{\sigma}^{-4} \sum (Y_i - \overline{Y})^2 \beta^T \mathrm{Var}\,(X^+) \beta \ .$$

If we estimate $\mathrm{Var}\,(X^+)$ by $[1/(n-1)] \sum (X_i^+ - \overline{X}^+)(X_i^+ - \overline{X}^+)^T$, then we may write $\mathrm{Var}\,(\hat{u}) = \hat{\sigma}^{-4} \beta^T V \beta$, with U and V as given by equations (1) and (2). The score test is then

$$\begin{aligned}T_{\mathrm{H}}^2 &= \hat{u}^T \hat{v}^{\oplus} \hat{u} \\ &= \hat{\sigma}^{-2} U^T \beta (\hat{\sigma}^{-4} \beta^T V \beta)^{\oplus} \hat{\sigma}^{-2} \beta^T U \\ &= U^T \beta (\beta^T V \beta)^{\oplus} \beta^T U\end{aligned}$$

(where $^{\oplus}$ denotes the generalized inverse), the same as that given by Chapman et al. (2003), who show that maximizing with respect to β^T gives

$$T_{\mathrm{H}}^2 = U^T V^{\oplus} U \ .$$

References

Chapman JM (2005) Statistical methods for linkage disequilibrium mapping of disease susceptibility genes in candidate regions. PhD thesis, University of Cambridge, Cambridge

Chapman JM, Cooper JD, Todd JA, Clayton DG (2003) Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. Hum Hered 56:18–31

Chen Z, Zheng G, Ghosh K, Li Z (2005) Linkage disequilibrium mapping of quantitative-trait loci by selective genotyping. Am J Hum Genet 77:661–669

Clayton D, Chapman J, Cooper J (2004) Use of unphased multilocus genotype data in indirect association studies. Genet Epidemiol 27: 415–428

Cornish KM, Manly T, Savage R, Swanson J, Morisano D, Butler N, Grant C, Cross G, Bentley L, Hollis CP (2005) Association of the dopamine transporter (DAT1) 10/10-repeat genotype with ADHD symptoms and response inhibition in a general population sample. Mol Psychiatry 10:686–698

de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005) Efficiency and power in genetic association studies. Nat Genet 37:1217–1223

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199

Prentice RL, Pyke R (1979) Logistic disease incidence models and case-control studies. Biometrika 66:403–411

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. Am J Hum Genet 69:1–14

Rao CR (1962) A note on a generalized inverse of a matrix with

applications to problems in mathematical statistics. J R Stat Soc Ser B 24:152–158
Sasieni PD (1997) From genotypes to genes: doubling the sample size. Biometrics 53:1253–1261
Slatkin M (1999) Disequilibrium mapping of a quantitative-trait locus in an expanding population. Am J Hum Genet 64:1764–1772
Stuart A, Ord K, Arnold S (1999) Linear regression and correlation. In: Stuart A, Ord K, Arnold S. Kendall's advanced theory of statistics, volume 2A: classical inference and the linear model. 6th ed. Arnold, London, pp 486–488
van Gestel S, Houwing-Duistermaat JJ, Adolfsson R, van Duijn CM, Broeckhoven CV (2000) Power of selective genotyping in genetic association analyses of quantitative traits. Behav Genet 30:141–146
Xiong M, Fan R, Jin L (2002*a*) Linkage disequilibrium mapping of quantitative trait loci under truncation selection. Hum Hered 53: 158–172
Xiong M, Zhao J, Boerwinkle E (2002*b*) Generalized T^2 test for genome association studies. Am J Hum Genet 70:1257–1268