to generate table 1 in the letter from Curtis and Sham (2006)—the average values of the χ^2 statistic and the entropy-based statistic are much larger than that of the likelihood-ratio–based heterogeneity test statistic. The range of the haplotype frequencies leading to this result is not large. Both figures 1 and 2 demonstrate that the average values of the χ^2 statistic and the entropy-based statistic are similar and that, in most cases, the average values of both the χ^2 statistic and the entropy-based statistic are smaller than that of the likelihood-ratio–based heterogeneity test statistic.

If the covariance matrices of the moment estimates of haplotype frequencies in the standard χ^2 statistic and the entropy-based statistic are replaced with the covariance matrix of maximum-likelihood estimates of the haplotype frequencies, the average values of the three statistics will be asymptotically the same. Therefore, if we were to use the covariance matrix from the maximum-likelihood estimates, any differences among the three test statistics would be, on average, small.

In summary, (1) we formulate the entropy-based statistic in terms of the estimated frequencies, not counts as implemented in the simulation performed by Curtis and Sham (2006); (2) the estimation error for haplotype frequencies will have an impact on the test statistic under the null hypothesis, and the magnitude of this effect will depend on the haplotype frequencies; and (3) asymptotically, the impact of the haplotype-frequency estimation error on the standard χ^2 statistic and the entropy-based statistic is smaller than the impact on the likelihood-ratio–based heterogeneity test statistic. Therefore, the claim that type I error rates of the heterogeneity test are always much smaller than those of the standard χ^2 statistic and the entropy-based statistic is incorrect. The effects of the haplotype-frequency estimation errors on type I error rates of a test are complex and should be investigated by both theoretical analysis and intensive simulation studies over large parameter spaces, not just over a small range of haplotype frequencies.

JINYING ZHAO, ERIC BOERWINKLE, AND MOMIAO XIONG

Human Genetics Center
University of Texas Health Science Center at Houston
Houston

Reference

Curtis D, Sham PC (2006) Estimated haplotype counts from case-control samples cannot be treated as observed counts. Am J Hum Genet 78:729–730 (in this issue)

Address for correspondence and reprints: Dr. Momiao Xiong, Human Genetics Center, The University of Texas Health Science Center at Houston, Houston, TX 77030. E-mail: Momiao.Xiong@uth.tmc.edu

0002-9297/2006/7804-0020$15.00

Am. J. Hum. Genet. 78:729–730, 2006

Estimated Haplotype Counts from Case-Control Samples Cannot Be Treated as Observed Counts

To the Editor:

Although the entropy-based method described by Zhao et al. (2005) provides a sensitive way to detect kinds of departure from a random distribution of haplotype counts between two data sets, we cannot see how it can be applied in practice to case-control samples. This is because tests that treat haplotype counts estimated from unphased data as if they were actually observed haplotypes are inherently anticonservative.

To illustrate in principle why this is, let us consider a sample in which all subjects happen to be doubly heterozygous at two loci, with genotype Aa/Bb. The maximum-likelihood estimates for the haplotype frequencies do not consist, as one might think intuitively, of each possible haplotype having frequency 0.25. Instead, there are two equally likely solutions: that haplotypes AB and ab each occur with frequency 0.5 or that, conversely, haplotypes Ab and aB both have frequency 0.5. (With N subjects, the solution for four haplotypes has likelihood 0.25^{2N}, whereas the solution for just two haplotypes has likelihood 0.5^{2N}.) If a few other genotypes are added to the data set, they will push the solution one way or the other. For example, if the sample consists of a mixture of cases and controls, and one case has genotype AA/Bb and one control has genotype aa/Bb, then the estimated haplotype frequencies will suggest that almost all cases have haplotypes AB and ab, whereas almost all controls have haplotypes Ab and aB. Although such an extreme example would not occur in practice, it is important to understand that maximum-likelihood estimation of haplotype frequencies favors solutions containing a small number of different haplotypes. This implies that, when frequencies of multilocus haplotypes are estimated separately in cases and controls, small random effects can produce quite large, apparently notable differences. In a real situation, one might estimate, for instance, that a particular haplotype occurred in a small percentage of cases but never in controls, leading to the possibly erroneous deduction that this indicates the presence of a pathogenic mutation.

To determine whether haplotype frequencies differ sig-

nificantly between cases and controls, the correct approach is to perform a heterogeneity test, in which one calculates whether the overall likelihood is significantly higher if different frequencies are allowed than if the same frequencies apply to both groups. An incorrect approach is to estimate haplotype counts by multiplying the frequencies by twice the sample size and then to treat these counts as if they were actually observed. The counts may be compared using a Pearson χ^2 test on a contingency table, by a permutation test as implemented in the CLUMP program (Sham and Curtis 1995) or by the newly described entropy method (Zhao et al. 2005). In every case, the test based on estimated counts will be anticonservative.

To illustrate that this is the case, we randomly generated case-control samples genotyped for two markers, assuming that the population frequencies of the haplotypes were the same for all subjects, under the assumption of random mating. For each data set, we applied a Pearson χ^2 test and the entropy test to the counts of the simulated haplotypes. We then combined pairs of haplotypes into two-locus genotypes, and we used the GENECOUNTING program (Zhao et al. 2002) to obtain estimated haplotype frequencies in the cases, controls, and combined sample, along with the associated likelihoods. We applied a heterogeneity test to these likelihoods and again applied the Pearson χ^2 and entropy tests, this time to the estimated counts. Illustrative results are given in table 1, for which the population frequencies of the four haplotypes were set at 0.5, 0.2, 0.2, and 0.1, and a sample size of 500 cases and 500 controls was used. The Pearson χ^2 and entropy tests perform appropriately when applied to the actual haplotype counts, as does the heterogeneity test using likelihoods based on estimated frequencies. However, both of the tests that use estimated counts are markedly anticonservative.

It is not appropriate to treat estimated haplotypes as if they were observed, and tests that do so will produce unacceptably high type I error rates. As we have said, this will apply even if a permutation test is performed on the estimated haplotypes—for example, by inputting them into the CLUMP program (Sham and Curtis 1995). However, a valid test can be devised if, instead, the original data are repeatedly permuted and then, for each permuted data set, haplotypes are estimated and a test statistic is derived. The rank of the test statistic obtained from the original data set can then be used to obtain an empirical significance level (North et al. 2003), and such an approach could be used for the entropy-based statistic. Without such a permutation procedure, we do not see how the entropy test can be applied to case-control data.

Table 1

Number of Times Each Statistic Reaches a Given *P* Value in 100,000 Simulations

	Real Counts			Estimated Counts	
P	χ^2	Entropy Test	Heterogeneity Test	χ^2	Entropy Test
.05	5,013	4,971	5,072	11,115	11,089
.01	970	968	1,034	3,678	3,678
.001	103	107	114	797	813
.0001	10	7	11	206	209
.00001	0	0	1	51	54
.000001	0	0	1	16	17

David Curtis[1] and Pak C. Sham[2]
[1]*Academic Department of Psychiatry, Queen Mary's School of Medicine and Dentistry, and* [2]*Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, London*

References

North BV, Curtis D, Sham PC (2003) A note on the calculation of empirical *P* values from Monte Carlo procedures. Am J Hum Genet 72:498–499

Sham PC, Curtis D (1995) Monte Carlo tests for associations between disease and alleles at highly polymorphic loci. Ann Hum Genet 59: 97–105

Zhao J, Boerwinkle E, Xiong M (2005) An entropy-based statistic for genomewide association studies. Am J Hum Genet 77:27–40

Zhao JH, Sham PC (2002) Faster haplotype frequency estimation using unrelated subjects. Hum Hered 53:36–41

Address for correspondence and reprints: Dr. David Curtis, Academic Department of Psychiatry, Royal London Hospital, Whitechapel, London E1 1BB, United Kingdom. E-mail: david.curtis@qmul.ac.uk

0002-9297/2006/7804-0021$15.00

Am. J. Hum. Genet. 78:730–731, 2006

Reply to Wirtenberger et al.

To the Editor:

Wirtenberger et al. (2006) analyzed the SNP content of 82 large (median length 157 kb) common copy-number polymorphisms (CNPs), selected from the Database of Genomic Variations, and determined the number of SNPs included in the GeneChip Mapping 100K arrays (Affymetrix). The data they presented showed that the density of these SNPs within the CNPs is lower than would be expected, with 52.4% of CNPs having no SNP coverage (median length 120 kb) and only 8.5% having