Linkage Disequilibrium Testing When Linkage Phase Is Unknown

Daniel J. Schaid[1]

Department of Health Sciences Research, Mayo Clinic/Foundation, Rochester, Minnesota 55905

Manuscript received April 10, 2003
Accepted for publication September 10, 2003

ABSTRACT

Linkage disequilibrium, the nonrandom association of alleles from different loci, can provide valuable information on the structure of haplotypes in the human genome and is often the basis for evaluating the association of genomic variation with human traits among unrelated subjects. But, linkage phase of genetic markers measured on unrelated subjects is typically unknown, and so measurement of linkage disequilibrium, and testing whether it differs significantly from the null value of zero, requires statistical methods that can account for the ambiguity of unobserved haplotypes. A common method to test whether linkage disequilibrium differs significantly from zero is the likelihood-ratio statistic, which assumes Hardy-Weinberg equilibrium of the marker phenotype proportions. We show, by simulations, that this approach can be grossly biased, with either extremely conservative or liberal type I error rates. In contrast, we use simulations to show that a composite statistic, proposed by Weir and Cockerham, maintains the correct type I error rates, and, when comparisons are appropriate, has similar power as the likelihood-ratio statistic. We extend the composite statistic to allow for more than two alleles per locus, providing a global composite statistic, which is a strong competitor to the usual likelihood-ratio statistic.

LINKAGE disequilibrium (LD), the nonrandom association of alleles from different loci, can provide valuable information on the structure of haplotypes of the human genome. This may prove useful for studying the association of genomic variation with human traits because haplotype-based methods can offer a powerful approach for disease gene mapping (Daly *et al.* 2001; Gabriel *et al.* 2002). The measurement and testing of LD among measured genetic variants is often based on pairs of loci; statistical analyses measure the departure of the joint frequency of pairs of alleles from two loci on a haplotype from random pairing of alleles. Statistical evaluation of LD is well developed when haplotypes are directly observed (Hedrick 1987; Weir 1996). But, it is common to measure genetic markers on unrelated subjects without knowing the haplotype origin (linkage phase) of the marker alleles. In this case, a common way to test for LD is to enumerate all pairs of haplotypes that are consistent with each subject's observed marker phenotypes, calculate maximum-likelihood estimates (MLEs) of the haplotype frequencies, and use these estimates to construct a likelihood-ratio statistic—twice the difference between the log-likelihood based on MLEs and the log-likelihood based on independence of alleles from different loci (Excoffier and Slatkin 1995; Hawley and Kidd 1995; Long *et al.* 1995; Slatkin and Excoffier 1996). This method, however, requires the assumption of random pairing of haplotypes, which implies that each of the loci has genotype proportions that fit Hardy-Weinberg equilibrium (HWE) proportions (see appendix). It has been shown that departure from HWE proportions, which we denote Hardy-Weinberg disequilibrium (HWD), can bias estimates of haplotype frequencies (Fallin and Schork 2000). The impact of HWD on the statistical properties of the likelihood-ratio statistic is not well known.

An alternative method that allows for unknown linkage phase was provided by Weir (1979) and Weir and Cockerham (1989) and discussed in the book by Weir (1996). They explicitly incorporate the ambiguity of the double heterozygote by using a composite measure of LD. The composite test measures the association of alleles from different loci on the same haplotype (intragametic LD) as well as on different haplotypes (intergametic LD). The advantages of this approach are that HWD at either locus is incorporated into the test statistic and the statistic is rapidly computed. Weir (1979) showed that this composite statistic provides the correct type I error rate when testing LD whether or not there is departure from HWE at either locus.

The first purpose of this report is to demonstrate the impact of HWD on the statistical properties of the likelihood-ratio statistic. An advantage of the likelihood-ratio method is that it allows for more than two alleles at either locus and provides a global test for LD among any of the pairs of alleles from the loci. The second purpose of this report is to extend the method of Weir and Cockerham to a global test of LD that allows for more than two alleles at either of the loci.

[1] *Address for correspondence:* Harwick 775, Section of Biostatistics, Mayo Clinic/Foundation, Rochester, MN 55905.
E-mail: schaid@mayo.edu

METHODS

To provide the necessary background, some of the developments of Weir and Cockerham (1989) are briefly reviewed (see also Weir 1996, pp. 94 and 125). Suppose that locus A has J possible alleles, $A_1, A_2, \ldots, A_J$, and locus B has K possible alleles, $B_1, B_2, \ldots, B_K$. Assuming that alleles are codominant, the probabilities of the marker phenotypes at the A locus can be expressed in terms of allele frequencies (p_{A_j}) and coefficients for HWD, $D_{A_{ij}}$,

$$P(A_i, A_i) = p^2_{A_i} + D_{A_{ii}},$$

$$P(A_i, A_j) = 2p_{A_i}p_{A_j} - 2D_{A_{ij}},$$

where

$$D_{A_{ii}} = \sum_{j, j \neq i} D_{A_{ij}}.$$

When $D_{A_{ij}} > 0$, there are fewer A_i, A_j heterozygotes than predicted by HWE proportions, and when $D_{A_{ij}} < 0$, there are more A_i, A_j heterozygotes than predicted. Similar probabilities can be written for the phenotypes of the B locus (with subscript A replaced by subscript B). The HWD coefficients can be estimated by the allele frequencies and the relative frequencies of the phenotype categories. Let $\hat{f}_{A_iA_j}$ denote the observed relative frequency of phenotype A_i, A_j ($\hat{f}_{A_iA_j}$ = {number of subjects with phenotype A_i, A_j}/N, where N is the total number of subjects). Then, the HWD coefficient for alleles A_i and A_j is

$$\hat{D}_{A_{ij}} = (2\hat{p}_{A_i}\hat{p}_{A_j} - \hat{f}_{A_iA_j})/2.$$

Linkage disequilibrium when phase is unknown: When haplotypes are directly observed, linkage disequilibrium is measured by the intragametic LD,

$$D_{A_jB_k} = P(A_jB_k \text{ on same haplotype}) - p_{A_j}p_{B_k}.$$

One could also measure the nonrandom association of alleles A_j and B_k from different haplotypes, called the intergametic LD:

$$D_{A_j/B_k} = P(A_jB_k \text{ on different haplotypes}) - p_{A_j}p_{B_k}.$$

When linkage phase is unknown, the underlying pair of haplotypes is ambiguous for the double-heterozygous phenotypes, and so one cannot directly measure the intragametic LD. To surmount this issue, Weir and Cockerham proposed a composite measure of LD, the sum of the intra- and intergametic disequilibria:

$$\begin{aligned}\Delta_{A_jB_k} &= D_{A_jB_k} + D_{A_j/B_k}\\ &= P(A_jB_k \text{ on same or different haplotypes}) - 2p_{A_j}p_{B_k}.\end{aligned}$$

When there are only two alleles per locus, there is one composite LD, say $\Delta_{A_1B_1}$, and an estimator is

$$\hat{\Delta}_{A_1B_1} = (n_{A_1B_1}/N) - 2\hat{p}_{A_1}\hat{p}_{B_1},$$

where

$$n_{A_1B_1} = 2X_{A_1,A_1,B_1,B_1} + X_{A_1,A_1,B_1,B_2} + X_{A_1,A_2,B_1,B_1} + (1/2)X_{A_1,A_2,B_1,B_2},$$

X is a count of the number of subjects with the phenotype indicated by its subscript, and $\hat{p}_{A_1}$, $\hat{p}_{B_1}$ are estimates of allele frequencies. The factor ½ in front of the X for double heterozygotes should not be interpreted as assuming equally likely phases of the double heterozygotes, because the advantage of the composite statistic is that this is not assumed. Rather, the coefficients in front of each X count the number of times that A_1 and B_1 occur on either the same haplotype or different haplotypes, in accordance with the definition of the composite statistic based on $P(A_jB_k$ on the same or different haplotypes). For example, the phenotype (A_1, A_1, B_1, B_1) must have the underlying haplotype pair $A_1 - B_1$ and $A_1 - B_1$, so there are two occurrences of A_1 and B_1 on the same haplotype and on different haplotypes. The phenotype (A_1, A_1, B_1, B_2) must have the underlying haplotype pair $A_1 - B_1$ and $A_1 - B_2$, so there is one occurrence of A_1 and B_1 on the same haplotype and on different haplotypes. The phenotype (A_1, A_2, B_1, B_2) can have two pairs of haplotypes, either $A_1 - B_1$ and $A_2 - B_2$ (in which case the count is ½ because A_1 and B_1 occur on the same haplotype but not on different haplotypes) or $A_1 - B_2$ and $A_2 - B_1$ (in which case the count is ½ because A_1 and B_1 occur on different haplotypes but not on the same haplotypes).

When there are only two alleles per locus, there are eight phenotype categories, and the counts of these categories can be represented by the vector X. This emphasizes that $n_{A_1B_1}$ is a linear combination of the elements of the X vector, and so too are $\hat{p}_{A_1}$ and $\hat{p}_{B_1}$. Hence, $\hat{\Delta}_{A_1B_1}$ is a function of linear combinations of observed multinomial frequencies. This fact makes it straightforward to derive an estimator for the variance of $\hat{\Delta}_{A_1B_1}$, and the chi-square statistic to test the null hypothesis of no LD is $S = \hat{\Delta}^2_{A_1B_1}/\mathrm{Var}(\hat{\Delta}_{A_1B_1})$.

When there are more than two alleles at either locus, all possible pairs of LD coefficients can be estimated. For J alleles at locus A and K alleles at locus B, there are a total of $(J-1)(K-1)$ composite coefficients. To extend the work of Weir and Cockerham, we first use the vector X to denote phenotype counts for all possible distinguishable two-locus phenotypes. The sum of the elements of this vector is N, the total number of subjects. Suppose that L is the length of vector X; then, each composite LD can be written as a function of linear combinations of terms from the vector X. To see this, we first define counting vectors, α, β, and γ, each of length L. The vectors α and β are used to count alleles for loci A and B, respectively. A subscript on these vectors indicates the type of allele that is counted. For example, α_j counts alleles of type A_j. The ith element of α_j is denoted $\alpha_{j,i}$, which has a value of 1, 0.5, or 0, according to whether the *i*th phenotype category has 2, 1, or 0 alleles of type A_j. The vector β_k counts alleles of type B_k in a similar manner. Allele frequencies can be estimated by these count vectors, such as $\hat{p}_{A_j} = \alpha_j' X/N$ and $\hat{p}_{B_k} = \beta_k' X/N$. The count vector γ is used to count how often specific alleles from loci *A* and *B* occur together. For alleles A_j and B_k, the count vector is defined as follows:

$$\gamma_{jk} = \begin{cases} 2 & \text{if } A_j,\ A_j,\ B_k,\ B_k \\ 1 & \text{if } A_j,\ A_i,\ B_k,\ B_k \text{ or } A_j,\ A_j,\ B_k,\ B_l, \text{ where } i \neq j,\ l \neq k \\ 0.5 & \text{if } A_j,\ A_i,\ B_k,\ B_l, \text{ where } i \neq j,\ l \neq k \\ 0 & \text{otherwise.} \end{cases}$$

The double heterozygotes receive a factor of 0.5, because these subjects contribute differently to the intragametic and intergametic components of disequilibria [see further details in Weir (1996, p. 122)]. With the defined count vectors, an estimate of the composite LD can be expressed as

$$\begin{aligned} \hat{\Delta}_{A_jB_k} &= (n_{A_jB_k}/N) - 2\hat{p}_{A_j}\hat{p}_{B_k} \\ &= \gamma_{jk}' X/N - 2(\alpha_j' X/N)(\beta_k' X/N). \end{aligned}$$

Variances and covariance: When more than two alleles exist at either locus, there is more than one composite LD coefficient. These coefficients are correlated, because they depend on the multinomial count vector *X* and because the same alleles can overlap between different coefficients. To derive the covariance matrix of the LD coefficients, we use Fisher's formula, which is a special case for a Taylor series approximation for functions that depend on the relative frequencies of the multinomial categories, X_i/N in our case. For a more complete description of Fisher's formula, see Bailey (1961, p. 285). The covariance of the functions T_1 and T_2 (*e.g.*, $T_1 = \hat{\Delta}_{A_jB_k}$ and $T_2 = \hat{\Delta}_{A_lB_m}$) can be derived from

$$\mathrm{Cov}(T_1,\ T_2) \approx N\sum_{i=1}^{L}\left(\frac{\partial T_1}{\partial X_{\mathrm{i}}}\right)\left(\frac{\partial T_2}{\partial X_i}\right)Q_i - N\left(\frac{\partial T_1}{\partial N}\right)\left(\frac{\partial T_2}{\partial N}\right). \quad (1)$$

After taking derivatives, the terms X_i are replaced by their expected values, NQ_i, where Q_i is the probability of the *i*th phenotype category. These derivatives for our situation can be expressed as

$$\frac{\partial\hat{\Delta}_{A_jB_k}}{\partial X_i} = \frac{\gamma_{jk,i}}{N} - \frac{2\{(\alpha_j' Q)\beta_{k,i} + (\beta_k' Q)\alpha_{j,i}\}}{N},$$

$$\frac{\partial\hat{\Delta}_{A_jB_k}}{\partial N} = -\frac{\gamma_{jk}' Q}{N} + \frac{4(\alpha_j' Q)(\beta_k' Q)}{N}.$$

Substituting these derivatives into expression (1) provides a way to estimate the covariance matrix for all the LD coefficients. To test the null hypothesis of no composite LD and no higher-order disequilibria, we compute the covariance matrix by using the vector of probabilities, *Q*, computed under the null hypothesis of no linkage disequilibrium and assuming that all disequilibrium parameters between loci (intra- and intergametic disequilibria and higher-order terms) are zero, but we allow for HWD by including appropriate disequilibria coefficients. Under these assumptions, the probabilities of the marker phenotypes at two loci are

$$\begin{aligned} P(A_j,\ A_j,\ B_k,\ B_k) &= (p_{A_j}^2 + D_{A_{jj}})(p_{B_k}^2 + D_{B_{kk}}), \\ P(A_j,\ A_j,\ B_k,\ B_m) &= (p_{A_j}^2 + D_{A_{jj}})(2p_{B_k}p_{B_m} - 2D_{B_{km}}), \\ P(A_j,\ A_l,\ B_k,\ B_k) &= (2p_{A_j}p_{A_l} - 2D_{A_{jl}})(p_{B_k}^2 + D_{B_{kk}}), \\ P(A_j,\ A_l,\ B_k,\ B_m) &= (2p_{A_j}p_{A_l} - 2D_{A_{jl}})(2p_{B_k}p_{B_m} - 2D_{B_{km}}). \quad (2) \end{aligned}$$

Parameter estimates for allele frequencies and HWD coefficients are substituted into expression (2) to estimate the *Q* vector under the null hypothesis.

Testing: To test the null hypothesis that all of the composite LD parameters are zero and that there are no higher-order disequilibria, we use a global chi-square statistic,

$$S = \hat{\Delta}' V^{-1}\hat{\Delta},$$

where $\hat{\Delta}$ is the vector of estimates of all LD coefficients, and V^{-1} is a generalized inverse of the covariance matrix. For large samples, *S* has a chi-square distribution. If all phenotype categories are observed, *V* is of full rank, where d.f. $= (J - 1)(K - 1)$. We use a generalized inverse of *V*, however, in case it is not of full rank; if this occurs, the degrees of freedom are the rank of the matrix *V*. The covariance matrix *V* may be less than full rank when there are sparse data, particularly when there are many alleles at some loci, of which some are rare. An advantage of this general approach is that if the global statistic is found to be significant, the individual coefficients can be tested according to

$$S_i = \frac{\hat{\Delta}_i^2}{V_{ii}},$$

where S_i has an approximate chi-square distribution with 1 d.f. These pair-specific tests are a by-product of the computations of the global test. Although one could ignore the global test and simply compute all possible tests for individual coefficients, one would need to correct for the multiple testing. This approach, of choosing the smallest *P* value and correcting by Bonferroni methods, might be most powerful if there were only one pair of alleles from the two loci in strong LD. However, if the amount of LD is of similar magnitude across multiple pairs of alleles, then the global test is likely to have greater power than testing individual coefficients.

Simulations: To evaluate the type I error rates and power of the composite chi-square and likelihood-ratio statistics, simulations were performed. The composite chi-square statistic was computed two ways: first by allowing for HWD as illustrated in expression (2) and second assuming HWE (*i.e.*, forcing $D_{A_{12}}$ and $D_{B_{12}}$ coefficients equal to zero). Although our motivation is not to require HWE, we evaluated the statistical properties of the composite test with assumed HWE for two reasons. First, we wish to evaluate whether the composite test loses power when in fact data are simulated under the assumption of HWE. Second, it may be tempting to first test for HWE before testing LD; if there is no statistical departure from HWE, then we assume HWE when using the composite test for LD. This practice might be valid if there were significant gains in power by assuming HWE whenever appropriate.

For simulations under the null hypothesis of no LD, the distribution of two-locus phenotypes was simulated using expression (2) assuming two alleles per locus, with allele frequencies p_{A_1} and p_{B_1} equal to either 0.2 or 0.5. The amount of departure from HWE was simulated according to the fraction of its extreme values. For locus A, the fraction of HWD is $f_{\mathrm{HWD},A} = -1$ or $+1$ according to whether $D_{A_{12}}$ is equal to its minimum or maximum value [minimum value $= \max(-p_{A_1}^2, -(1 - p_{A_1})^2)$; maximum value $= p_{A_1}(1 - p_{A_1})$]. A similar parameter, $f_{\mathrm{HWD},B}$, was used for locus B. We simulated data according to a grid of values of $f_{\mathrm{HWD},A}$ and $f_{\mathrm{HWD},B}$, each having values of -0.8, -0.2, 0, $+0.2$, and $+0.8$.

We also performed simulations under the null hypothesis of no LD for three alleles per locus. In this case, there are three types of heterozygotes and hence three D coefficients for HWD at each locus. The patterns of HWD can be complex, as the range of each D coefficient depends on allele frequencies and the other D coefficients. To simplify our evaluations, we assumed equal allele frequencies at each locus ($p_{A_i} = p_{B_i} = 1/3$), and we assumed that only alleles 1 and 2 at each locus departed from HWE, so that there is only one D coefficient for each locus (*i.e.*, only $D_{A_{12}}$ and $D_{B_{12}}$ were nonzero). The composite- and likelihood-ratio statistics have 4 d.f. when there are three alleles per locus.

To evaluate power, we assumed two alleles per locus, so that there is only one LD parameter. Because the likelihood-ratio statistic is biased when there is HWD, all simulations for power were computed assuming HWE, to assure that the power of the various statistics was evaluated at approximately the same type I error rates. The amount of LD was simulated according to the fraction of its extreme values, with $f_{\mathrm{LD}} = -1$ when $D_{A_1B_1} = \max(-p_{A_1}p_{B_1}, -p_{A_2}p_{B_2})$, and $f_{\mathrm{LD}} = +1$ when $D_{A_1B_1} = \min(p_{A_2}p_{B_1}, p_{A_1}p_{B_2})$; the parameter f_{LD} is equivalent to the familiar normalized $D'_{A_1B_1}$.

All simulations were based on 50 unrelated subjects and 1000 simulated data sets. Simulations and statistical analyses were conducted with S-PLUS software (Insightful). The code to compute the composite test is available upon request by sending an e-mail to schaid@mayo.edu.

RESULTS

Type I error rates: The estimated type I error rates, with an expected nominal rate of 0.05, are illustrated in Figure 1A for when allele frequencies are $p_{A_1} = p_{B_1} = 0.2$. Figure 1 illustrates that the composite chi-square statistic generally achieves the expected nominal error rate of 0.05 over all 25 simulated combinations of values for $f_{\mathrm{HWD},A}$ and $f_{\mathrm{HWD},B}$. For 1000 simulations, the 95% confidence interval for the simulated type I error rate is 0.036–0.064. For the data in Figure 1, the type I error rate for the composite statistic ranged from 0.038 to 0.068, and only 1 of 25 values exceeded the upper 95% confidence limit. In contrast, the composite statistic that assumed HWE (Figure 1B) was either overly conservative when there was negative HWD at either locus or anticonservative when there was positive HWD at either locus, and the joint effects of HWD at both loci tended to accentuate these trends. The type I error rate for the composite test with assumed HWE ranged from 0.017 to 0.263, with 18 of 25 values falling outside the 95% confidence interval (C.I.). The likelihood-ratio statistic tended to be liberal when the HWD at both loci was in the same direction (Figure 1C). The type I error rate for the likelihood-ratio statistic ranged from 0.042 to 0.2141, with 10 of 25 values falling outside the 95% C.I.

The trends in Figure 2, for when allele frequencies are $p_{A_1} = p_{B_1} = 0.5$, tend to follow similar patterns as those in Figure 1. Contrasting Figures 1 and 2 emphasizes that the impact of HWD on the type I error rate depends not only on the values of $f_{\mathrm{HWD},A}$ and $f_{\mathrm{HWD},B}$, but also on the allele frequencies. The composite statistic maintains the appropriate error rate of 0.05 (range of simulated values 0.034–0.068, with 2 of 25 falling outside the 95% C.I.), but the composite statistic with assumed HWE can be grossly conservative or liberal (range of simulated values 0.000–0.274, with 21 of 25 falling outside the 95% C.I.). The likelihood-ratio statistic can also have large departures from the nominal 0.05 error rate (range of simulated values 0.000–0.783, with 11 of 25 falling outside the 95% C.I.), with the largest departure occurring when both loci have extremely large negative values of $f_{\mathrm{HWD},A}$ and $f_{\mathrm{HWD},B}$, which implies an excessive number of heterozygotes at both loci. This situation is the worst for maximizing the likelihood, because double heterozygotes are ambiguous for linkage phase. In the extreme, with no homozygotes at either locus, the likelihood method fails because there are no unambiguous haplotypes to help estimate the relative frequencies of the different linkage phases among the double heterozygotes. But, an extreme excess number of homozygotes (both $f_{\mathrm{HWD},A}$ and $f_{\mathrm{HWD},B}$ having values of 0.8) also led to an inflated type I error rate for the likelihood-ratio statistic (error rates of 0.14 and 0.13 for allele frequencies of 0.2 and 0.5, respectively). Simulations for a nominal error rate of 0.01 demonstrated similar patterns as those illustrated in Figures 1 and 2 (results not shown). Furthermore, simulations with unequal allele frequencies (*i.e.*, $p_{A_1} = 0.2$, $p_{B_1} = 0.5$, and $p_{A_1} = 0.5$, $p_{B_1} = 0.2$) also showed trends similar to those illustrated in Figures 1 and 2 (results not shown).

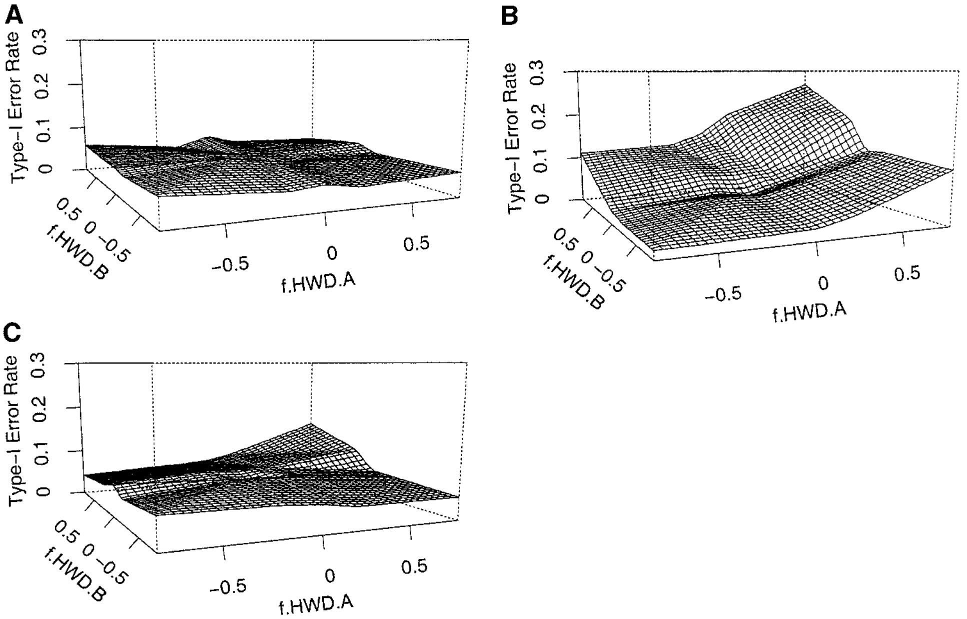

Figure 1.—Type I error rates based on simulations without LD, but allowing HWD to vary at each locus, in terms of $f_{\mathrm{HWD},A}$ and $f_{\mathrm{HWD},B}$, the fraction of HWD relative to their extreme values. Two alleles per locus were simulated, with allele frequencies $p_{A_1} = p_{B_1} = 0.2$. The types of statistics were: (A) the composite statistic, (B) the composite statistic that assumed HWE, and (C) the likelihood-ratio statistic.

Simulation results for three alleles per locus, with alleles 1 and 2 at each locus departing from HWE and yet no LD between the loci, are presented in Figure 3. Similar to the case of two alleles per locus, the composite statistic maintains the appropriate error rate of 0.05 (range of simulated values 0.032–0.063, with 2 of 25 falling outside the 95% C.I.; see Figure 3A); the composite statistic with assumed HWE can be grossly conservative or liberal (range of simulated values 0.006–0.187, with 18 of 25 falling outside the 95% C.I.; see Figure 3B); and the likelihood-ratio statistic can have large departures from the nominal 0.05 error rate (range of simulated values 0.044–0.416, with 17 of 25 falling outside the 95% C.I.; see Figure 3C). Again, the largest departure occurred when both loci had extremely large negative values of $f_{\mathrm{HWD},A}$ and $f_{\mathrm{HWD},B}$—an excessive number of heterozygotes at both loci.

Power: The power of the three statistics is presented in Figure 4. These simulations assumed HWE, so that all tests could be compared with the same approximate type I error rate. Figure 4 illustrates that all three statistics have similar power, although there is a small power advantage of the likelihood-ratio statistic when $p_{A_1} = p_{B_1} = 0.2$ and there is negative LD between the loci (see

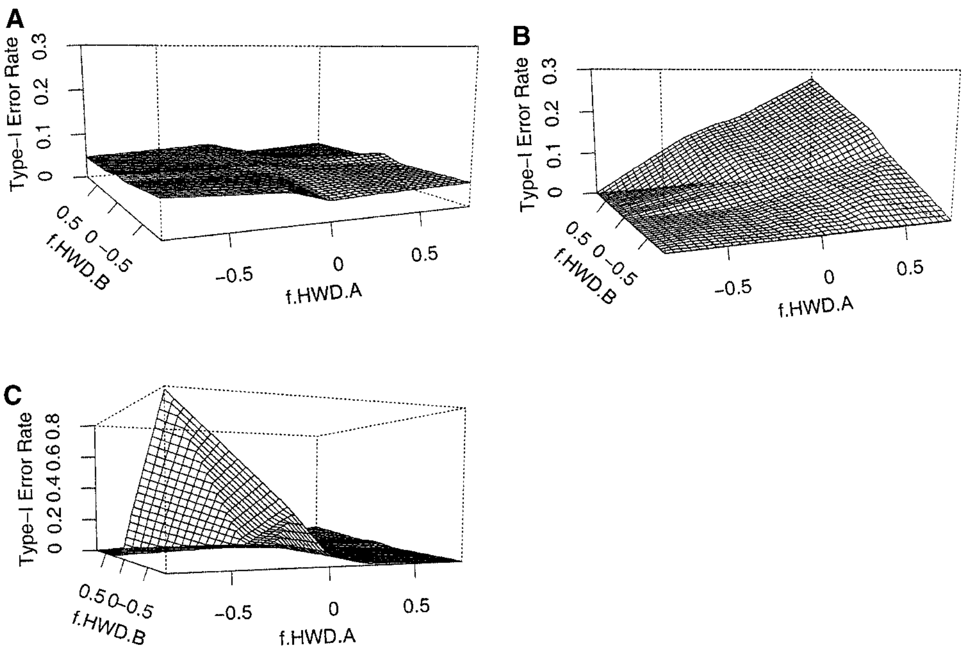

Figure 2.—Type I error rates based on simulations without LD, but allowing HWD to vary at each locus, in terms of $f_{\mathrm{HWD},A}$ and $f_{\mathrm{HWD},B}$, the fraction of HWD relative to their extreme values. Two alleles per locus were simulated, with allele frequencies $p_{A_1} = p_{B_1} = 0.5$. The types of statistics were: (A) the composite statistic, (B) the composite statistic that assumed HWE, and (C) the likelihood-ratio statistic.

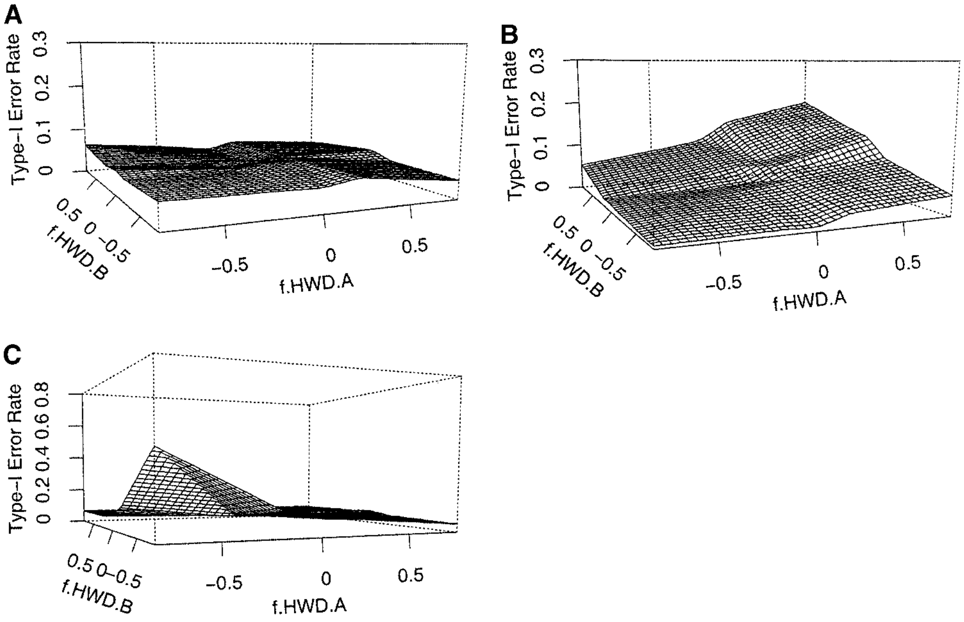

FIGURE 3.—Type I error rates for three alleles per locus (equal allele frequencies). Simulations allowed HWD to vary at each locus, in terms of $f_{\mathrm{HWD},A}$ and $f_{\mathrm{HWD},B}$ for alleles 1 and 2 at each locus. The types of statistics were: (A) the composite statistic, (B) the composite statistic that assumed HWE, and (C) the likelihood-ratio statistic.

Figure 4, left side). Surprisingly, there was no power difference between the composite test that allowed for HWD and that which assumed HWE. These results suggest that there is no advantage, in terms of power, to assume HWE for the composite statistic and that there can be a significant disadvantage in terms of robustness of the type I error rate to departures from HWE.

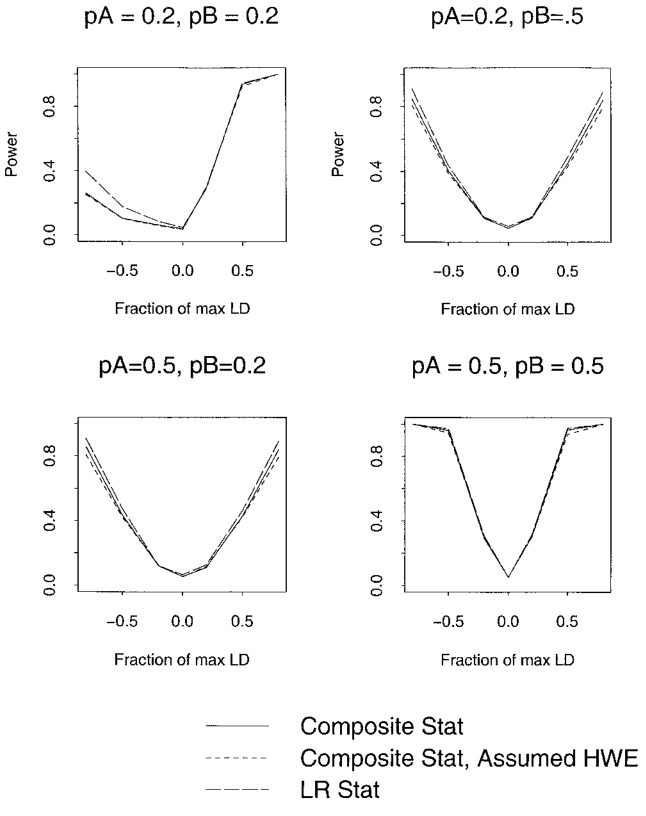

FIGURE 4.—Power for the composite statistic, composite statistic that assumed HWE, and likelihood-ratio statistic, with allele frequencies (p_A and p_B) varied between 0.2 and 0.5. Simulations assumed HWE at both loci.

DISCUSSION

Our simulation results illustrate that when linkage phase is unknown, departures from HWE can have dramatic effects on the commonly used likelihood-ratio statistic for testing LD. Gross departures from HWE, particularly an excess number of heterozygotes, can increase the rate of false-positive conclusions regarding LD. In contrast, the composite statistic provides a robust method to test for LD between loci. This statistic is based on estimates of composite LD and their covariances under the null hypothesis of no LD and no higher-order disequilibria. Our methods are direct extensions of those by Weir and Cockerham, where we derive the covariance between composite measures of LD. An alternative statistic, proposed by WEIR (1979), is based on the goodness-of-fit of the observed phenotype frequencies to their null expected values and is implemented in SAS (2003). For large sample sizes, the Wald-type of statistic that we propose and the goodness-of-fit statistic by Weir are expected to give similar results. For sparse data, due to some rare alleles, we speculate that the goodness-of-fit statistic may not be well approximated by the chi-square distribution, as is often found for other goodness-of-fit statistics. Our approach, based on covariances of composite LD measures, can use the singular values of the covariance matrix to assess the numerical stability of the statistic and reduce the degrees of freedom according to the rank of the covariance matrix, if needed. Further work is needed to compare the small sample properties of our proposed statistic and the goodness-of-fit statistic.

Although it may be tempting to first test for HWE and then decide whether or not to assume HWE in the composite statistic, our simulations suggest that assuming HWE does not provide any power advantage, yet it could inflate the type I error rate. This suggests that the composite statistic should be used for routine testing for LD regardless of whether or not HWE exists at either locus.

Several forces could cause departure from HWE, and a critically important cause could be error in the measurement of genotypes. For this reason, departures from HWE are often used as a crude measure of quality control. This approach, however, does not provide adequate guidelines on when a marker should be excluded from the analysis (*i.e.,* the threshold of statistical significance for concern) or whether particular subjects should be excluded. An alternative approach is to incorporate genotyping errors into methods of analysis, an approach that has been successful in linkage analysis of pedigree data (Sobel *et al.* 2002). Because departures from HWE could be caused by genotyping errors, explicit models of genotyping error could be incorporated into the usual likelihood models for haplotype frequencies, so that departures from HWE would be absorbed into parameters that measure genotyping error rates. More work along this type of modeling may prove beneficial. Although our simulations are limited in terms of the many different patterns of LD that could arise when more than two alleles exist at either locus, the broad range of LD that we explored for the simple case of two alleles per locus suggests that the composite statistic has power similar to that of the likelihood-ratio statistic. It may be possible to construct situations where the likelihood-ratio statistic has greater power, yet the potential inflation of the type I error rate does not seem to warrant routine use of this method.

Our work has focused entirely on determination of an appropriate way to test for LD, regardless of whether either locus attains HWE. We have not addressed the best way to estimate the amount of LD when there are departures from HWE. Numerous authors have discussed the statistical properties of competing measures of LD when linkage phase of double heterozygotes is known (Hedrick 1987; Devlin and Risch 1995; Zabetian *et al.* 2003), but there is little understanding about measures of LD when linkage phase is unknown and there are departures from HWE. The composite measure offers appeal, but it can be difficult to interpret for several reasons. First, it is an additive measure of the departure of the observed genotype frequency from that expected if there were no LD. This is analogous to the measure D when linkage phase of the double heterozygotes is known (*i.e.,* $D_{AB} = p_{AB} - p_A p_B$). Hence, this type of additive measure will depend on allele frequencies; see Hedrick (1987) for more discussion. Second, the composite measure depends not only on the association of alleles between two loci on the same gamete (the usual D value), but also on the association of alleles between the two loci on different gametes. This latter type of association is typically ignored, but may occur when there are departures from HWE. The composite measure of LD is confounded between LD and HWD. Clearly, more work is required to determine the best measure of LD when the assumption of HWE is violated.

In conclusion, our results suggest that testing for the presence of LD between two loci with unknown linkage phase should be performed by the composite statistic. We have extended the work of Weir and Cockerham to allow for more than two alleles at either of the loci, and so this general composite statistic is a strong competitor to the traditional likelihood-ratio statistic.

This research was supported by United States Public Health Services, National Institutes of Health, contract grant no. GM65450.

LITERATURE CITED

Bailey, N., 1961 *Introduction to the Mathematical Theory of Genetic Linkage.* Oxford University Press, Oxford.

Daly, M., J. Rioux, S. Schaffner, T. Hudson and E. Lander, 2001 High-resolution haplotype structure in the human genome. Nat. Genet. **29:** 229–232.

Devlin, B., and N. Risch, 1995 A comparison of linkage diequilibrium measures for fine-scale mapping. Genomics **29:** 1–12.

Excoffier, L., and M. Slatkin, 1995 Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol. Biol. Evol. **12:** 921–927.

Fallin, D., and N. Schork, 2000 Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. Am. J. Hum. Genet. **67:** 947–959.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy *et al.*, 2002 The structure of haplotype blocks in the human genome. Science **296:** 2225–2229.

Hawley, M. E., and K. K. Kidd, 1995 HAPLO: a program using the EM algorithm to estimate the frequencies of multi-site haplotypes. J. Hered. **86:** 409–411.

Hedrick, P. W., 1987 Gametic disequilibrium measures: proceed with caution. Genetics **117:** 331–341.

Long, J. C., R. C. Williams and M. Urbanek, 1995 An E-M algorithm and testing strategy for multiple-locus haplotypes. Am. J. Hum. Genet. **56:** 799–810.

SAS, 2003 *Genetics User's Guide for SAS*, Release 8.2 (http://support.sas.com/documentation/onlinedoc/genetics/).

Slatkin, M., and L. Excoffier, 1996 Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. Heredity **76:** 377–383.

Sobel, E., J. C. Papp and K. Lange, 2002 Detection and integration of genotyping errors in statistical genetics. Am. J. Hum. Genet. **70:** 496–508.

Weir, B., 1979 Inferences about linkage disequilibrium. Biometrics **35:** 235–254.

Weir, B., 1996 *Genetic Data Analysis II.* Sinauer Associates, Sunderland, MA.

Weir, B. S., and C. C. Cockerham, 1989 Complete characterization of disequilibrium at two loci, pp. 86–110 in *Mathematical Evolutionary Theory*, edited by M. W. Feldman. Princeton University Press, Princeton, NJ.

Zabetian, C. P., S. G. Buxbaum, R. C. Elston, M. D. Kohnke, G. M. Anderson *et al.*, 2003 The structure of linkage disequilibrium at the DBH locus strongly influences the magnitude of association between diallelic markers and plasma dopamine beta-hydroxylase activity. Am. J. Hum. Genet. **72:** 1389–1400.

Communicating editor: R. W. Doerge

APPENDIX

The random pairing of haplotypes implies that the genotypes at each locus are expected to have genotype proportions in HWE. We can show why this occurs for the case of two loci; it is straightforward to extend our arguments to more loci. Let A_jB_k denote a haplotype. If haplotypes are randomly paired, the probability of the pair $(A_jB_k, A_{j'}B_{k'})$ is

$$P(A_jB_k, A_{j'}B_{k'}) = P(A_jB_k)P(A_{j'}B_{k'}).$$

Under this assumption, the probability of the single-locus genotype $A_jA_{j'}$ is

$$\begin{aligned} P(A_jA_{j'}) &= \sum_k\sum_{k'}P(A_jB_k)P(A_{j'}B_{k'}) \\ &= \left(\sum_k P(A_jB_k)\right)\left(\sum_{k'}P(A_{j'}B_{k'})\right) \\ &= P(A_j)P(A_{j'}), \end{aligned}$$

which illustrates that the probability of the single-locus genotype is the product of its allele frequencies and hence fits HWE. Symmetric arguments can be used to show that single-locus genotypes at the B locus are also expected to fit HWE.