

Pedigree Disequilibrium Tests for Multilocus Haplotypes

Frank Dudbridge*

MRC Human Genome Mapping Project Resource Centre, Cambridge, UK

Association tests of multilocus haplotypes are of interest both in linkage disequilibrium mapping and in candidate gene studies. For case-parent trios, I discuss the extension of existing multilocus methods to include ambiguous haplotypes in tests of models which distinguish between the *cis* and *trans* phase. A likelihood-ratio test is proposed, using the expectation-maximization (E-M) algorithm to account for haplotype ambiguities. Assumptions about the population structure are required, but realistic situations, including population stratification, which violate the assumptions lead to conservative tests. I describe a permutation procedure for the null hypothesis of interest, which controls for violation of the assumptions. For general pedigrees, I describe extensions of the pedigree disequilibrium test to include uncertain haplotypes. The summary statistics are replaced by their expected values over prior distributions of haplotype frequencies. If prior distributions are not available, a valid test is possible by using the E-M algorithm to estimate the null distribution of haplotype frequencies. Similar methods are available for quantitative traits. Exact permutation tests are difficult to construct in small samples, but an approximate procedure is appropriate in large samples, and can be used to account for dependencies between tests of multiple haplotypes and loci. *Genet Epidemiol* 25:115–121, 2003. © 2003 Wiley-Liss, Inc.

Key words: TDT; PDT; association tests; family-based controls; haplotype analysis

*Correspondence to: Frank Dudbridge, MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge CB10 1SB, UK.

E-mail: f.dudbridge@hgmp.mrc.ac.uk

Received for publication 16 December 2002; Revision accepted 10 March 2003

Published online in Wiley InterScience (www.interscience.wiley.com)

DOI: 10.1002/gepi.10252

INTRODUCTION

Association tests of multilocus haplotypes are of interest both in linkage disequilibrium mapping and in candidate gene studies. Well-preserved ancestral haplotypes, or multiple interacting loci, can lead to higher power from haplotype testing [Bader, 2001]. Furthermore, when multiple candidate loci are closely linked, multilocus methods are necessary to distinguish the loci with a primary association from those which are merely in disequilibrium with them [Bitti et al., 2001; Cordell and Clayton, 2002].

In studies of unrelated subjects, several similar methods have emerged which use the expectation-maximization (E-M) algorithm to estimate haplotype frequencies from unphased genotype data [Zhao et al., 2000; Schaid et al., 2002; Zaykin et al., 2002]. Methods for family-based studies are less well-developed, owing in part to the variety of proposed methods, and the desirable property of robustness to population structure. For case-parent trios, a versatile model was described by Cordell and Clayton [2002], in which various hypotheses of within- and between-locus effects

can be tested. However, the method as described cannot include uncertain haplotypes in tests which distinguish between *cis* and *trans* phase interactions, including association tests of individual haplotypes. In contrast, the method of Clayton [1999] could include uncertain haplotypes, but the only test implemented was for individual haplotype association. Here, I discuss the combination of these two models to include all families in all multilocus tests. The proposed approach requires some assumptions about the study population; to achieve robustness to violation of the assumptions, I give a permutation procedure which simulates the null hypothesis of interest.

For general pedigrees, the pedigree disequilibrium test (PDT) was proposed as a general test of association which is robust to linkage [Martin et al., 2001]. Robust tests for quantitative traits based on similar principles were proposed by Monks and Kaplan [2000] and Zhang et al. [2001]. Here, I show that the summary statistics can be replaced by expectations over prior distributions of haplotype frequencies. When prior information is unavailable, I show that valid tests are possible

by using maximum likelihood estimates under the null hypothesis. I also discuss permutation tests to account for multiple tests of nonindependent haplotypes and loci.

TRANSMISSION/DISEQUILIBRIUM TEST

METHODS

Here I discuss the application of the model by Clayton [1999] to the more sophisticated tests described by Cordell and Clayton [2002]. From Clayton [1999], the full likelihood contribution of a case-parent trio is

$$\Pr(G_f, G_m, G_c|D) = \Pr(G_f, G_m|D) \cdot \Pr(G_c|G_f, G_m, D)$$

where D is the ascertainment event, and G_f , G_m , and G_c are the sets of phased multilocus genotypes (i.e., paired haplotypes) consistent with the observed unphased genotypes of the father, mother, and child, respectively. When the haplotypes are unambiguous, the first term, the parental component, depends on both the phased genotype frequencies and their relative risks, but the second term, the conditional likelihood contribution, depends only on the relative risks. Inference based on the conditional likelihood is therefore robust to the population structure, which is an important feature of the TDT [Spielman et al., 1993; Sham and Curtis, 1995].

When several haplotypes are consistent with the genotype data, the parental and the conditional likelihood terms both involve the phased genotype frequencies, which become nuisance parameters in tests of association. Score tests may still be derived for the conditional likelihood [Dempster et al., 1977; Clayton, 1999]; however, when comparing the evidence for nested models of haplotype risk, the likelihood-ratio test is often preferred. In this case, conditional inference can be problematic, because the number of estimatable nuisance parameters may differ between the null and alternative hypotheses, and it is more convenient to work with the full likelihood of the trio. Summing over phase resolutions, and assuming Hardy-Weinberg equilibrium (HWE) in the population, random mating in the parents, and simplex ascertainment, the likelihood contribution is

$$\sum_{g_f \in G_f, g_m \in G_m, g_c \in G_c} \Pr(u_{g_f, g_m, g_c}) \Pr(g_c|D)$$

where u_{g_f, g_m, g_c} is the “untransmitted” genotype formed by the two haplotypes of g_f and g_m which

are not transmitted to the child. Under HWE, the genotype frequencies are products of haplotype frequencies, so this likelihood is equivalent to one which regards the child haplotypes as cases and the untransmitted haplotypes as controls. This likelihood can then be maximized under both null and alternative hypotheses, using the E-M algorithm in the same way as for an unmatched case-control sample. This approach incorporates uncertain haplotypes into several previously proposed and similar methods [Terwilliger and Ott, 1992; Schaid and Sommer, 1993; Thomson, 1995; Morris et al., 1997].

Similar to an unmatched case-control sample, models incorporating main effects and interactions of multiple loci may be fitted using unconditional logistic regression, allowing all the models of Cordell and Clayton [2002] to be applied to all families. However, this approach depends on the assumptions stated above, and it is important to consider the situations in which they may not hold. A well-known situation is population stratification, which introduces deviation from HWE. Population membership is a confounder, which leads to a conservative test in an unmatched analysis of a matched design [Breslow and Day, 1980, p. 271]. This is in contrast to unmatched case-control studies, where stratification can lead to both conservative and liberal tests.

Missing parental genotypes may also result in deviation from HWE. When only one parent is available, some families should be excluded from analysis when that parent is heterozygous [Curtis and Sham, 1995]. This leads to an increase in the proportion of homozygous parents analyzed, which biases the odds for transmission towards 1 and again results in a conservative test [Sasieni, 1997].

Misspecification of the ascertainment should retain the correct type-1 error rate, but will lead to either an increase or a decrease in power, depending on the details of the misspecification [Thomson, 1995]. Although simplex ascertainment is assumed in constructing the test, this is not a critical assumption, which is reassuring since many family-based samples are ascertained for linkage studies.

When the null hypothesis is that no haplotypes are associated, a permutation procedure can control for violation of these assumptions. This is possible because we can generate the distribution of the full-likelihood statistic under the null hypothesis of the TDT, by permuting only

on the conditional part of the likelihood. That is, for each parent in each trio, we exchange the “transmitted” and “untransmitted” status of its two haplotypes, with probability $\frac{1}{2}$ [Zhao et al., 1999], thus retaining the matched design of the study.

Here this procedure is extended to uncertain haplotypes by permuting the transmission status of the latent parental haplotypes, before computation of the likelihood. That is, under the null hypothesis, there are four equally likely child genotypes, consisting of the observed child itself, one formed from the untransmitted haplotypes, and two formed from one transmitted and one untransmitted haplotype. One of these configurations is chosen at random before calculating the likelihood contribution. This is equivalent to choosing one of four possible likelihood contributions, each of which is a sum over the same haplotype resolutions, but with the case/control status varied between the two haplotypes of each parent. Each randomization can be fixed across all markers in a study, to give a significance across multiple haplotypes and loci [Markianos et al., 2001].

This approach is applicable when the null hypothesis is that all haplotypes have equal relative risk. When comparing the evidence for different multilocus models, the null hypothesis imposes different constraints on the haplotype relative risks. In some situations, maximum-likelihood estimates can be used [Li, 2001], but their variance will be large in small samples, and may lead to incorrect results. In a large sample, however, the asymptotic distribution does not depend on the relative risks, being (usually) a chi-square distribution with degrees of freedom determined by the complexity of the models compared. Asymptotically, then, the permutation procedure based on all risks being equal will generate the correct distribution. Therefore, to perform tests of more sophisticated hypotheses, including uncertain haplotypes but controlling for population structure, we may use the permutation procedure for the TDT described above, if the sample is sufficiently large. This conveniently avoids the problem of permuting according to haplotype relative risks when the haplotypes themselves are uncertain.

SIMULATIONS

To assess the asymptotic type-1 error of the full-likelihood test, simulated data for 500 case-parent

trios were generated. Three-locus haplotypes of SNPs were generated in the parents under HWE, and offspring haplotypes were chosen according to Mendelian segregation. Three distributions of haplotype frequencies were used: a uniform distribution, and two in which haplotypes were grouped in complementary pairs with equal frequencies (see Table I). Homogeneous and stratified populations were simulated. Results are given in Table I. For HWE populations, error rates are close to the nominal level, while in stratified populations, the test is conservative, as expected.

The permutation procedure was assessed in combination with the conditional ETDT [Koeleman et al., 2000]. The null hypothesis constrains the relative risks to be equal for haplotypes which are identical at a subset of loci. Here the first SNP was the conditioning locus, and haplotypes with allele 2 had twice the relative risk of those with allele 1. Offspring haplotypes were determined according to a gamete competition model [Sinsheimer et al., 2001]. Results are given in Table II. In all cases, error rates were close to the nominal level. This is consistent with the proposition that, in a large sample, the permutation procedure assuming equal relative risks will generate the correct null distribution, and furthermore is robust to deviation from HWE caused by stratification.

TABLE I. Type-1 error of full-likelihood test for case-parent trios^a

	$\alpha=0.1$ (0.0042) ^b	$\alpha=0.05$ (0.0030)	$\alpha=0.01$ (0.0014)
HWE population 1 ^c	0.0986	0.0558	0.0136
HWE population 2 ^d	0.1036	0.0534	0.0118
Stratified 3:1 ^e	0.0732	0.0356	0.0064
Stratified 1:1	0.0686	0.0334	0.0058

^aResults are given for 5,000 replicates of 500 case-parent trios, using global test for haplotypes of three diallelic markers.

^bParentheses give standard errors for α -level.

^cUniform distribution of haplotype frequencies. Loci are in linkage equilibrium, probability of a heterozygous intercross is $1/8$, and expected proportion of trios with uncertain haplotypes is $1-(7/8)^3=0.33$.

^dHaplotypes 1-1-1 and 2-2-2 have frequency 0.3, 1-2-2 and 2-1-1 have frequency 0.15, and all others have frequency 0.025. Loci are in disequilibrium, and expected proportion of trios with uncertain haplotypes was empirically estimated as 0.26.

^eIn stratified simulations, a third population is simulated in which haplotypes 1-1-2 and 2-2-1 have frequency 0.3, 1-2-1 and 2-1-2 have frequency 0.15, and all others have frequency 0.025. This population is then combined with population 2 in the exact proportion indicated.

TABLE II. Type-1 error of permutation test with CETDT^a

	$\alpha=0.1$ (0.009)	$\alpha=0.05$ (0.007)	$\alpha=0.01$ (0.003)
HWE population 1	0.102	0.053	0.014
HWE population 2	0.097	0.048	0.007
Stratified 3:1	0.087	0.049	0.009
Stratified 1:1	0.079	0.044	0.009

^aResults are given for 1,000 replicates of 200 case-parent trios. In each replicate, 100 random permutations are generated, and the P -value is estimated as recommended by North et al. [2003]. Relative risk of haplotypes with allele 2 at first locus is twice that of those with allele 1. Other parameters are as in Table I.

The permutation procedure was found to give accurate results in samples as small as 10 trios (data not shown), although the asymptotic test becomes liberal. This may be because the procedure is still appropriate for the null hypothesis, up to a misspecification of some nuisance parameters (the haplotype relative risks). It would therefore seem to be more robust to small sample sizes than the asymptotic test; but in practice, one should perform simulations closely modeled on the given data to establish the validity of the procedure.

PEDIGREE DISEQUILIBRIUM TEST

METHODS

The PDT breaks a pedigree into N_T case-parent trios and N_S discordant sib-pairs. For a specific allele, define X_{T_i} as the number of transmissions minus the number of nontransmissions in trio i , and X_{S_i} as the number of copies in the affected sib minus the number of copies in the unaffected sib, in sib-pair i . A measure of association is

$$D = w \left[\sum_{i=1}^{N_T} X_{T_i} + \sum_{i=1}^{N_S} X_{S_i} \right]$$

where w is a real-valued weight. Under the null hypothesis of no association, D has expectation zero in any pedigree, and can be combined across multiple pedigrees into a z -ratio using an empirical variance estimate [Martin et al., 2001].

D can be calculated for a specific haplotype by using just the triads and sib-pairs in which it can be deduced. Because an empirical variance estimate is used, we avoid problems of bias due to phase ambiguities. However, ambiguous haplotypes may be included by redefining D in terms of expected haplotype counts:

$$D = w \left[\sum_{i=1}^{N_T} E_{\Omega_{T_i}}(X_{T_i}) + \sum_{i=1}^{N_S} E_{\Omega_{S_i}}(X_{S_i}) \right]$$

where Ω_{T_i} and Ω_{S_i} are prior probability distributions for the gametic phased genotype frequencies. Then

$$\begin{aligned} E_{H_0}(D) &= w \cdot E_{H_0} \left(\sum E_{\Omega_{T_i}}(X_{T_i}) + \sum E_{\Omega_{S_i}}(X_{S_i}) \right) \\ &= w \left[\sum E_{\Omega_{T_i}}(E_{H_0}(X_{T_i})) + \sum E_{\Omega_{S_i}}(E_{H_0}(X_{S_i})) \right] = 0. \end{aligned}$$

D has expectation zero for any prior distributions, and a different distribution may be used for each trio and sib-pair. $E(X_{T_i})$ is calculated by enumerating all phased configurations consistent with the genotype data of trio i , and weighting the corresponding transmission counts by the probability of the configuration. Using the previous notation,

$$\begin{aligned} E_{\Omega_{T_i}}(X_{T_i}) &= \frac{\sum_{g_f \in G_f, g_m \in G_m, g_c \in G_c} \Pr(g_f, g_m, g_c)(X_{T_i} | g_c, g_f, g_m)}{\sum_{g_f \in G_f, g_m \in G_m, g_c \in G_c} \Pr(g_f, g_m, g_c)}. \end{aligned}$$

A similar expression holds for $E(X_{S_i})$. It is again convenient to assume HWE and random mating in the parents, so that $\Pr(g_f, g_m, g_c) \propto \Pr(u_{g_f, g_m, g_c}) \Pr(g_c)$, which is a product of haplotype frequencies. When there is only one configuration consistent with the genotype data, D is the same as in the original PDT. Thus we may construct a test of association for a specific haplotype which uses all trios and sib-pairs in all pedigrees. A global test for all haplotypes may be obtained from the marginal homogeneity statistic [Spielman and Ewens, 1996].

Since the calculation is on subunits of pedigrees, it is convenient to consider distributions of gametic haplotypes, rather than of founder haplotypes with recombination. If prior frequency data are unavailable, a practical solution is to estimate a single distribution for the whole data by maximum-likelihood, using the E-M algorithm, assuming the null hypothesis. Although this is not now a prior distribution, the test is seen to remain valid. If, in each trio, the transmitted and nontransmitted haplotypes are exchanged for both parental meioses, all heterozygous intercross loci remain heterozygous, and homozygous intercross loci remain homozygous. Therefore, the same trios have uncertain haplotypes, with the same resolutions, so the maximum-likelihood distribution is the same. Furthermore, this configuration has the same probability under the null hypothesis as the original data. Therefore, the realizations of the null hypothesis can be arranged

into pairs for which the same distribution is estimated, and X_{T_i} are the same except for a change of sign, so that $E_{H_0}(E(X_{T_i})) = 0$. Similarly, if the affection status is exchanged in each discordant sib-pair, then trivially the same distribution is estimated and X_{S_i} is the same, except for a change of sign. Therefore, $E(D) = 0$, and individual and global tests may again be constructed.

For quantitative traits, similar tests were described by Monks and Kaplan [2000] and Zhang et al. [2001], here termed QPDT. In a nuclear family with n offspring, let Y_i be the trait value in offspring i and G_i be the number of copies of the allele of interest. The within-family covariance between trait values and genotype is estimated by

$$R = \frac{1}{n} \sum_i (Y_i - E(Y))(G_i - E(G))$$

where the expectations are appropriately defined. When there is no association, R has expectation zero and can be combined across multiple families into a z-ratio. The method extends to general pedigrees by taking the mean of R over all nuclear families in the pedigree [Zhang et al., 2001].

To include uncertain haplotypes into the QPDT, the covariance can be estimated over all haplotype configurations as well as all siblings. That is, for sibling i , we use

$$E_{\Omega}((Y_i - E(Y))(G_i - E(G))) = \frac{\sum_{g_f \in G_f, g_m \in G_m, g_c \in G_c} \Pr(g_f, g_m, g_c) (Y_i - E(Y))(G_i - E(G) | g_f, g_m, g_c)}{\sum_{g \in G_f, g_m \in G_m, g_c \in G_c} \Pr(g_f, g_m, g_c)}$$

where Ω is a prior probability distribution for gametic phased genotype frequencies. Again, it is convenient to work with gametic haplotype frequencies, and any prior distribution could be used in each family. If a single distribution is estimated from the data, assuming the null hypothesis, then realizations of the null hypothesis can be arranged into pairs in which the same distribution is estimated, and $G_i - E(G)$ differs only by a change of sign, leading again to $E(R) = 0$.

A permutation procedure is desirable, but it is unclear how to proceed when the null hypothesis may include linkage. Monks and Kaplan [2000] proposed randomizing the sign of R , which corresponds to simultaneously exchanging the transmitted and nontransmitted haplotypes for all meioses in a sibship. However, this assumes complete linkage, and excludes the situations where the transmission is exchanged in only one

parent. Both of these factors can result in failure to cover the whole permutation space: in simulations of 10 small pedigrees, the procedure was found to give conservative results (data not shown).

In a large sample, the central limit theorem assures us that the permutation distribution of R need not be the same as the null distribution, provided that it has mean zero. This is indeed the case when the sign of R is randomized. Therefore, the procedure is appropriate in large samples, delivering the asymptotic distribution. It is reasonable to expect that it would preserve the correlations between multiple haplotypes and loci which exist under the true null hypothesis. By applying all tests of multiple haplotypes and loci to each randomized data set, accurate significance levels can be obtained across multiple tests. The same remarks apply to D in the discrete trait PDT.

SIMULATIONS

Three-generation pedigrees were simulated, consisting of a sib-pair family in which each sibling founds a next-generation sib-pair family. The third generation consists of two sets of sib-pairs who are first cousins, and there are 10 pedigree members. A diallelic quantitative-trait locus (QTL) was simulated, and a discrete trait was generated according to a liability-threshold model (see Table III). A three-locus haplotype of SNPs was simulated with complete linkage to the QTL, but no association. Genotypes were generated for all pedigree members.

Type-1 error rates were estimated for analyses using only unambiguous haplotypes, all haplotypes using the true distribution as the prior, all haplotypes using the (untrue) uniform prior, and all haplotypes using the null distribution

TABLE III. Type-1 error of haplotype PDT^a

	$\alpha=0.1$ (0.0042)	$\alpha=0.05$ (0.0030)	$\alpha=0.01$ (0.0014)
Unambiguous ^b	0.097	0.0494	0.0094
True prior ^c	0.1016	0.0492	0.0094
Uniform prior	0.0964	0.0472	0.01
Maximum likelihood ^d	0.101	0.0496	0.0092

^aResults are given for 5,000 replicates of 500 three-generation pedigrees. Marker haplotype is completely linked to a diallelic QTL with minor allele frequency 0.3 and additive value of 1 on a standard normal distribution, with no dominance. Liability threshold of 1 determines affection status.

^bOnly unambiguous haplotype transmissions are scored.

^cTrue distribution of haplotype frequencies is as for population 2 in Table I.

^dNull distribution is estimated in each replicate.

TABLE IV. Type-1 error haplotype QPDT^a

	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$
Unambiguous	0.1024	0.0482	0.0096
True prior	0.0976	0.0484	0.0092
Uniform prior	0.0962	0.051	0.0108
Maximum likelihood	0.0968	0.0486	0.009

^aAll parameters are as in Table III.

estimated from the data. Results for the PDT are given in Table III, and for the QPDT in Table IV. In all cases, the error rate was close to the nominal level.

DISCUSSION

This report describes some practical methods for incorporating uncertain haplotype data into the TDT, PDT, and QPDT. The methods are implemented in software which is available from the author (see below). The software also includes programs for case-control and quantitative trait analyses for unrelated subjects, which incorporate some common applications of the generalized linear model in a simplified interface.

The case-parent trio design is well-studied, and methods are now available to test a wide range of models of within- and between-locus interaction. The PDT and QPDT are less well-developed in the sense that only the simple null hypothesis can be tested. More sophisticated tests may require a linear modeling framework similar to that used in trios. This could perhaps be achieved by the use of full-pedigree likelihoods [Sinsheimer et al., 2001], or by extension of the variance-components model [Abecasis et al., 2000]. In all cases, there are challenges remaining in correctly incorporating uncertain haplotype data.

I have not explicitly considered the problem of missing genotypes, but it could be addressed by considering all possible genotypes to be consistent with the data. This approach has the advantage of including all informative subjects in the analysis, and eliminates problems of bias and deviation from HWE due to missing data, although it can significantly increase the computation time.

The permutation procedures described here are only accurate in large samples, but will be useful to account for dependence between tests of multiple haplotypes and loci. Procedures for small samples which apply in all situations are an open problem; they are important because, even in large

collections of pedigrees, some haplotypes will be rare, with correspondingly small cell counts. One solution is to group rare haplotypes into a common pool with sufficiently high frequency. Although this would reduce the power to detect effects in rare haplotypes, an increase in power for common haplotypes is possible by excluding the rare group from the alternative hypothesis.

I have not considered the estimation of effect size. Unbiased estimates of relative risk are available for independent trios [Sham and Curtis, 1995], but for the proposed PDT methods, estimates of effects such as the mean trait value per haplotype will only be unbiased if the true haplotype distribution is used for the prior. The present methods are applicable to exploratory studies, for which secondary estimation of haplotype effects remains an open problem.

ELECTRONIC DATABASE INFORMATION

Software implementing the methods in this report is available at <http://www.hgmp.mrc.ac.uk/~fdudbrid/software/> and <ftp://ftp.hgmp.mrc.ac.uk/pub/linkage/>.

ACKNOWLEDGMENTS

I thank Heather Cordell and David Clayton for many helpful discussions; Bobby Koeleman, Iain Eaves, Francesco Cucca, and John Todd for motivating and testing the software; Mathias Chiano, Lars Berglund, Adrian Mander, and Kourosh Ahmadi for useful comments and suggestions; and the reviewers for constructive suggestions.

REFERENCES

- Abecasis GR, Cardon LR, Cookson WO. 2000. A general test of association for quantitative traits in nuclear families. *Am J Hum Genet* 66:279–292.
- Bader JS. 2001. The relative power of SNPs and haplotypes as genetic markers for association tests. *Pharmacogenomics* 2: 11–24.
- Bitti PP, Murgia BS, Ticca A, Ferrai R, Musu L, Piras ML, Puledra E, Campo S, Durando S, Montomoli C, Clayton DG, Mander AP, Bernardinelli L. 2001. Association between the ancestral haplotype HLA A30B18DR3 and multiple sclerosis in central Sardinia. *Genet Epidemiol* 20:271–283.
- Breslow NE, Day NE. 1980. The analysis of case-control studies. Volume 1, statistical methods in cancer research. Lyon: IARC.

- Clayton D. 1999. A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. *Am J Hum Genet* 65:1170–1177.
- Cordell HJ, Clayton DG. 2002. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 70:124–141.
- Curtis D, Sham PC. 1995. A note on the application of the transmission disequilibrium test when a parent is missing. *Am J Hum Genet* 56:811–812.
- Dempster AP, Laird NM, Rubin D. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* 39: 1–38.
- Koeleman BP, Dudbridge F, Cordell HJ, Todd JA. 2000. Adaptation of the extended transmission/disequilibrium test to distinguish disease associations of multiple loci: the conditional extended transmission/disequilibrium test. *Ann Hum Genet* 64: 207–213.
- Li H. 2001. A permutation procedure for the haplotype method for identification of disease-predisposing variants. *Ann Hum Genet* 65:189–196.
- Markianos K, Daly MJ, Kruglyak L. 2001. Efficient multipoint linkage analysis through reduction of inheritance space. *Am J Hum Genet* 68:963–977.
- Martin ER, Bass MP, Kaplan NL. 2001. Correcting for a potential bias in the pedigree disequilibrium test. *Am J Hum Genet* 68:1065–1067.
- Monks SA, Kaplan NL. 2000. Removing the sampling restrictions from family-based tests of association for a quantitative-trait locus. *Am J Hum Genet* 66:576–592.
- Morris AP, Whittaker JC, Curnow RN. 1997. A likelihood ratio test for detecting patterns of disease-marker association. *Ann Hum Genet* 61:335–350.
- 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.
- Sasieni PD. 1997. From genotypes to genes: doubling the sample size. *Biometrics* 53:1253–1261.
- Schaid DJ, Sommer SS. 1993. Genotype relative risks: methods for design and analysis of candidate-gene association studies. *Am J Hum Genet* 53:1114–1126.
- Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. 2002. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70:425–434.
- Sham PC, Curtis D. 1995. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann Hum Genet* 59:323–326.
- Sinsheimer JS, McKenzie CA, Keavney B, Lange K. 2001. SNPs and snails and puppy dogs' tails: analysis of SNP haplotype data using the gamete competition model. *Ann Hum Genet* 65:483–490.
- Spielman RS, Ewens WJ. 1996. The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 59:983–989.
- Spielman RS, McGinnis RE, Ewens WJ. 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52: 506–516.
- Terwilliger JD, Ott J. 1992. A haplotype-based "haplotype relative risk" approach to detecting allelic associations. *Hum Hered* 42:337–346.
- Thomson G. 1995. Mapping disease genes: family-based association studies. *Am J Hum Genet* 57:487–498.
- Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. 2002. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum Hered* 53:79–91.
- Zhang S, Zhang K, Li J, Sun F, Zhao H. 2001. Test of association for quantitative traits in general pedigrees: the quantitative pedigree disequilibrium test. *Genet Epidemiol [Suppl]* 21: 370–375.
- Zhao JH, Sham PC, Curtis D. 1999. A program for the Monte Carlo evaluation of significance of the extended transmission/disequilibrium test. *Am J Hum Genet* 64:1484–1485.
- Zhao JH, Curtis D, Sham PC. 2000. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 50: 133–139.