REPORT

Monte Carlo Pedigree Disequilibrium Test for Markers on the X Chromosome

Jie Ding, Shili Lin, and Yang Liu

Because of the need for fine mapping of disease loci and the availability of dense single-nucleotide–polymorphism markers, many forms of association tests have been developed. Most of them are applicable only to triads, whereas some are amenable to nuclear families (sibships). Although there are a number of methods that can deal with extended families (e.g., the pedigree disequilibrium test [PDT]), most of them cannot accommodate incomplete data. Furthermore, despite a large body of literature on association mapping, only a very limited number of publications are applicable to X-chromosomal markers. In this report, we first extend the PDT to markers on the X chromosome for testing linkage disequilibrium in the presence of linkage. This method is applicable to any pedigree structure and is termed "X-chromosomal pedigree disequilibrium test" (XPDT). We then further extend the XPDT to accommodate pedigrees with missing genotypes in some of the individuals, especially founders. Monte Carlo (MC) samples of the missing genotypes are generated and used to calculate the XMCPDT (X-chromosomal MC PDT) statistic, which is defined as the conditional expectation of the XPDT statistic given the incomplete (observed) data. This MC version of the XPDT remains a valid test for association under linkage with the assumption that the pedigrees and their associated affection patterns are drawn randomly from a population of pedigrees with at least one affected offspring. This set of methods was compared with existing approaches through simulation, and substantial power gains were observed in all settings considered, with type I error rates closely tracking their nominal values.

Family-based tests for association have been used to map disease genes for more than a decade. The original transmission/disequilibrium test (TDT) was proposed to detect linkage disequilibrium (LD) in family triads.[1] This test and its many variants require the genotypes of the parents for calculation of the statistics. For data with missing genotypes among parents, sibling TDT (STDT) was proposed to use the genotypes of phenotypically discordant sibships.[2] Sometimes, it is possible to reconstruct parental genotypes from the genotypes of offspring. In this situation, reconstruction-combined TDT (RCTDT) was proposed to use reconstructed genotypes to increase the test power.[3] However, RCTDT applies only to independent nuclear families. For extended pedigrees, the pedigree disequilibrium test (PDT) is among a handful of methods that have been proposed to date.[4–6] PDT uses both family triads and discordant sib pairs (DSPs) from an extended pedigree, but, when there are many missing genotypes, its power may be low. All the tests discussed above are applicable only to autosomal markers, except STDT and RCTDT, which have been extended to X-chromosomal markers.[7,8] However, note that both of these tests are amenable to nuclear families only.

We extend the PDT to X-chromosomal markers, taking into account the fact that there are different numbers of alleles in males and females. When there are missing genotypes in a pedigree, known or estimated marker-allele frequencies are used to simulate missing genotypes conditional on the observed ones from their relatives. The test statistic, defined below, is based on these Monte Carlo (MC) samples, which contain complete genotypic data.

For each pedigree, a statistic D is calculated from all family triads and DSPs in the pedigree. Suppose the marker of interest has two alleles, denoted as "M1" and "M2," and the genotypes of all individuals in the pedigree are known. For each family triad consisting of one affected child and two parents, we define

$$X_T = (\#\text{ M1 transmitted from mother}) - (\#\text{ M1 not transmitted from mother}) .$$

Note that this statistic is nonzero (values of 1 or −1) only if the mother is heterozygous at the marker locus. A DSP consists of one affected sibling and one unaffected sibling. Since males and females have different numbers of X chromosomes, only DSPs of the same sex are used. For such a DSP, we have

$$X_S = (\#\text{ M1 in affected sib}) - (\#\text{ M1 in unaffected sib}) .$$

Again, note that this statistic is nonzero (taking value of 1 or −1) only if the two siblings have different genotypes at the marker locus.

For each pedigree, the X_T and X_S statistics from all family triads and DSPs are summed up to give the statistic D. Under the null hypothesis of no association between the

From the Department of Statistics (J.D.; S.L.) and Division of Cancer Immunology, Department of Pathology (Y.L.), The Ohio State University, Columbus
Received May 22, 2006; accepted for publication July 6, 2006; electronically published August 1, 2006.
Address for correspondence and reprints: Dr. Shili Lin, Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210-1247. E-mail: shili@stat.ohio-state.edu
Am. J. Hum. Genet. 2006;79:567–573. 0002-9297/2006/7903-0021$15.00

Table 1. Allele Frequencies, Haplotype Frequencies, and Penetrances under Nine Different Settings

	Frequency										
	Marker Allele[a]		Disease Allele		Haplotype[b]				Penetrances[c]		
Setting	1	2	1	2	D1B1	D2B1	D1B2	D2B2	f_{D1D1}	f_{D1D2}	f_{D2D2}
1	.2	.8	.3	.7	.2	.0	.1	.7	.390	.325	.260
2	.4	.6	.3	.7	.3	.1	.0	.6	.390	.325	.260
3	.3	.7	.3	.7	.3	.0	.0	.7	.390	.325	.260
4	.2	.8	.3	.7	.2	.0	.1	.7	.440	.340	.240
5	.4	.6	.3	.7	.3	.1	.0	.6	.440	.340	.240
6	.3	.7	.3	.7	.3	.0	.0	.7	.440	.340	.240
7	.2	.8	.3	.7	.2	.0	.1	.7	.580	.380	.180
8	.4	.6	.3	.7	.3	.1	.0	.6	.580	.380	.180
9	.3	.7	.3	.7	.3	.0	.0	.7	.580	.380	.180

[a] The frequencies specified apply to both markers B and C.

[b] The disease locus is in linkage equilibrium with marker C. DiBj = a haplotype with allele i at the disease locus D and allele j at the marker locus B; $i,j = 1,2$. Three levels of LD between marker B and the disease locus were investigated: settings 3, 6, and 9 ({3,6,9}) depict complete LD, {2,5,8} specifies that D2 can be associated with both B1 and B2 with different marker and disease frequencies, and {1,4,7} indicates that it is D1 that can be associated with both B1 and B2, also with different marker and disease frequencies.

[c] We studied three disease models—{1–3}, {4–6}, and {7–9}—where f_{DiDj} is the penetrance with genotype ij at the disease locus; $ij = 11,12,22$. They are ordered from least genetic to most genetic, but even the most genetic one still accounts for only 8% of the total variance of the trait.

marker locus and the disease locus, each X_T or X_S has a mean of 0, so their sum D also has a mean of 0. Under the assumption that all the n pedigrees in the data set are independent and that D_i is the statistic from the ith pedigree, we have

$$E\left(\sum_{i=1}^{n} D_i\right) = 0$$

and

$$Var\left(\sum_{i=1}^{n} D_i\right) = \sum_{i=1}^{n} Var(D_i) = \sum_{i=1}^{n} E(D_i^2) = E\left(\sum_{i=1}^{n} D_i^2\right) .$$

Then, the overall statistic,

$$T = \frac{\sum_{i=1}^{n} D_i}{\sqrt{\sum_{i=1}^{n} D_i^2}} ,$$

follows a standard normal distribution asymptotically. The test based on this statistic is referred to hereafter as "XPDT" (X-chromosomal PDT).

In the case that genotypes of some individuals in a pedigree are missing, we propose the use of the following test statistic:

$$D_{\mathrm{MC}} = E[D \mid G_{\mathrm{o}}] = E[D(G_{\mathrm{m}},G_{\mathrm{o}},A) \mid G_{\mathrm{o}}] , \qquad (1)$$

where G_{o} and G_{m} denote the observed and missing genotypes, respectively, and A is the collection of observed phenotypes (disease affection statuses). Furthermore, D, the test statistic defined above, is based on the second equality of complete genotype data and is written more fully in equation (1). Since the conditional expectation cannot be written in a closed form, we estimate the test statistic through an MC simulation method by drawing independent samples $G_{\mathrm{m}k}, k = 1, \ldots, K$ from $P(G_{\mathrm{m}} \mid G_{\mathrm{o}})$. That is,

$$D_{\mathrm{MC}} \approx \frac{1}{K}\sum_{k=1}^{K} D(G_{\mathrm{m}k},G_{\mathrm{o}},A) .$$

For multiple pedigrees, the T_{MC} statistic is calculated, as before, using the D_{MC} statistics from all the pedigrees. The test based on this T_{MC} statistic is called "XMCPDT" (X-chromosomal MC PDT). Note that when multiple linked markers are available, generation of the MC samples of missing genotypes can be done using all the markers together, to reduce variability.

Given a fixed pedigree affection pattern A and a missing genotype pattern, the expectation of the D_{MC} statistic is

$$E(D_{\mathrm{MC}} \mid A) = \sum_{G_{\mathrm{o}} \in \mathcal{G}_{\mathrm{o}}} \sum_{G_{\mathrm{m}} \in \mathcal{G}_{\mathrm{m}}} D(G_{\mathrm{m}},G_{\mathrm{o}},A) P(G_{\mathrm{m}} \mid G_{\mathrm{o}}) P(G_{\mathrm{o}} \mid A) ,$$

where $\mathcal{G}_{\mathrm{o}}$ (or $\mathcal{G}_{\mathrm{m}}$) is the set of all possible genotypes for individuals with known (or missing) genotypes. When linkage exists among disease locus and marker loci, $P(G_{\mathrm{m}} \mid G_{\mathrm{o}})$

Table 2. Type I Error Rates Estimated from 2,000 Replicates

	Type I Error Rate											
	Complete Data		Incomplete Data									
			OSUMS Missing Pattern[a]					All Founders Missing[b]				
Setting	XS	XP	XS	XRC	XP	XMC_T	XMC_E	XS	XRC	XP	XMC_T	XMC_E
1	.035	.045	.026	.033	.040	.038	.038	.020	.023	.036	.038	.029
2	.044	.055	.029	.034	.056	.052	.054	.030	.032	.052	.052	.050
3	.038	.048	.029	.040	.053	.043	.044	.025	.028	.035	.044	.035
4	.040	.047	.026	.034	.048	.045	.043	.018	.021	.035	.041	.035
5	.041	.054	.034	.042	.051	.061	.058	.026	.029	.041	.044	.038
6	.035	.045	.033	.036	.045	.039	.038	.025	.030	.041	.047	.043
7	.037	.052	.034	.036	.046	.045	.047	.025	.028	.038	.043	.035
8	.040	.047	.044	.040	.049	.049	.051	.035	.040	.041	.050	.042
9	.041	.052	.033	.041	.058	.053	.051	.025	.036	.042	.046	.044

NOTE.—Estimates are based on either complete or incomplete data, with use of various methods. XS = XSTDT; XP = XPDT; XRC = XRCTDT; XMC_T = XMCPDT with use of the true (exact) allele frequencies; XMC_E = XMCPDT with use of estimated allele frequencies.

[a] Simulated genotypes of individuals with unobserved data in the OSUMS data set were removed before analysis.

[b] All founder genotypes were removed before analysis.

is different from $P(G_m|G_o,A)$, and so $E(D_{MC}|A)$ may not be the same as

$$E(D|A) = \sum_{G_o \in \mathcal{G}_o} \sum_{G_m \in \mathcal{G}_m} D(G_m,G_o,A)P(G_m|G_o,A)P(G_o|A) ,$$

which is equal to 0 under the null hypothesis of no association, as discussed above. The value of $E(D_{MC}|A)$ under the null hypothesis depends on many factors, including the pedigree disease (affection) pattern A, the pedigree structure, and the underlying disease model.

If we assume that all the pedigrees are drawn from a certain underlying population and treat A as random, then the expectation of D_{MC} becomes

$$E(D_{MC}) = \sum_{A \in \mathcal{A}} \sum_{G_o \in \mathcal{G}_o} \sum_{G_m \in \mathcal{G}_m} D(G_m,G_o,A)P(G_m|G_o)P(G_o|A)P(A) ,$$

where $\mathcal{A}$ is the set of all possible disease patterns for this pedigree. It can be shown that this expectation is equal to 0 when there is no association among the disease and marker loci, regardless of whether linkage is present (see appendix A). If minimal ascertainment criteria are used in the sampling process (i.e., the only requirement is at least one affected individual among the offspring in each pedigree), then XMCPDT remains a valid test for association under linkage.

To evaluate type I errors and powers of XPDT and XMCPDT, a simulation study was performed on the basis of the pedigree structures and the missing-data patterns of the Ohio State University Multiple Sclerosis (OSUMS) data set,[9] which investigated a SNP marker on chromosome 6 for its association with the risk and progression of MS. Although there are 101 pedigrees in the data set, only a subset of 81 was used in our simulation, after removal of pedigrees that have no genotyped offspring. The total number of individuals used in our simulation is 386, 102 of whom have missing genotypes. Note that we are using only the structures, the observed phenotypes (affection statuses), and the genotype-missing patterns of these pedigrees, not their genotype data, since such data are available on an autosome but not on the X chromosome.

We simulated two SNP markers—B and C, with the same allele frequencies—on the X chromosome. The disease locus was assumed to be in complete linkage with the two markers and in linkage equilibrium with marker C but in various levels of LD with marker B. In this design, marker C is intended to be used to gauge the type I error rates, whereas marker B can be used to evaluate powers of the competing methods. The marker-allele frequencies, disease-allele frequencies, haplotype frequencies, and penetrances under nine different settings are shown in table 1. The shown penetrances are for female individuals; those for males were set to be the same as the corresponding homozygous females. Note that, in all settings, the disease allele 1 (D1) (the allele of interest) was more likely to be associated with marker B allele 1 (B1).

For each of the nine models considered, 2,000 replicates were simulated. To generate missing genotypes in the data set, either the missing patterns in the OSUMS data set were used or the genotypes of all founders were treated as missing. One hundred MC samples of missing genotypes were generated for each replicate with use of the software SLINK.[10,11] Either the simulation marker-allele frequencies or those estimated from the founders in each replicate were used in the MC sampling. XPDT and XMCPDT were performed using our own software (MCPDT) based on R (R Project for Statistical Computing). For comparison, we also applied XSTDT and XRCTDT (see Michael Knapp's Web site) to the same set of simulated data,

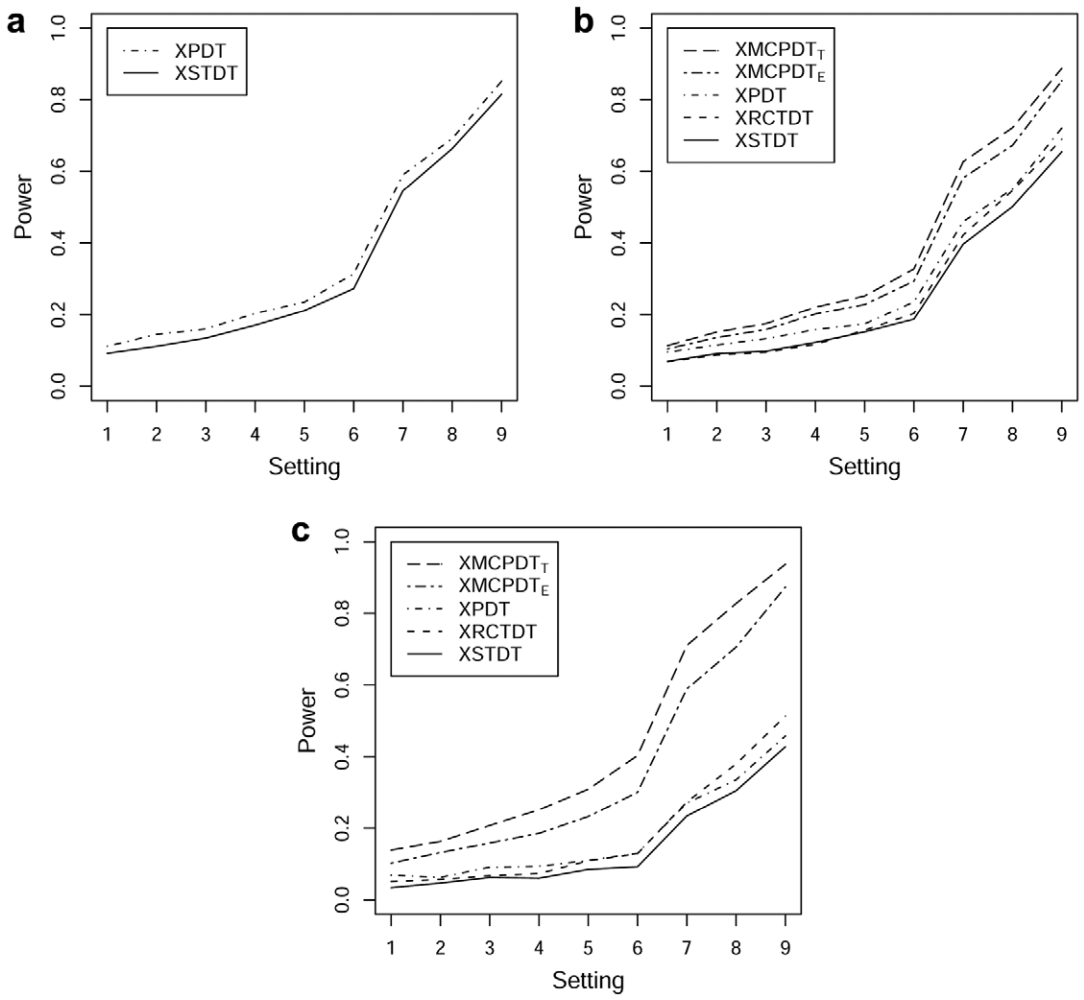

Figure 1. Comparisons of powers from several methods under nine settings for each of the following three scenarios. *a,* Complete data. *b,* Incomplete data, with missing patterns in pedigrees matching those of the OSUMS data set. *c,* Incomplete data, with the assumption that all founders' genotypes are missing. The methods are: XSTDT, XRCTDT, XPDT, XMCPDT$_T$ (XMCPDT with use of the true allele frequencies), and XMCPDT$_E$ (XMCPDT with use of estimated allele frequencies).

in which extended pedigrees were separated into nuclear families and treated as independent. Note that, for data with complete genotypes, XSTDT and XRCTDT are the same. In addition, for XSTDT and XRCTDT, exact *P* values were used, whereas the *P* values were obtained using normal approximation for XPDT and XMCPDT. Nominal levels of significance were set at .05 for all settings.

The type I error rates are shown in table 2. It can be seen that XSTDT and XRCTDT appear to be conservative, since their type I error rates are well below the nominal level. One explanation for the observed conservativeness of these two procedures is that there are more *P* values close to 1 (instead of uniform 0–1), which is especially true when there is a large proportion of missing genotypes (e.g., when the genotypes of all the founders are missing). This may be because correlated nuclear families were treated as if they were independent. On the other hand, the error rates for the various PDT procedures were around the nominal level, with some slightly >.05. Use of the true allele frequencies or the estimated ones in XMCPDT resulted in very similar error rates.

The powers are shown in figure 1. Under all models, XPDT had slightly higher powers than did XSTDT (or XRCTDT) for complete data (fig. 1*a*). When there were missing data, XMCPDT had considerably higher powers than did XSTDT, XRCTDT, or XPDT. The increases were larger for the settings where the genotypes for all founders were missing (fig. 1*c*) compared with settings where the OSUMS data-missing patterns were used (fig. 1*b*). Specifically, for the settings with missing-data patterns matching those of the OSUMS data, the relative increases in power over XSTDT had a range of 30%–81%, with an average of 56% across all nine settings and the two different specifications of the allele frequencies (exact or estimated). When all founders' genotypes were removed before the analysis, the average relative increase in power is 207%,

and the range is 104%–336%. Use of estimated allele frequencies in XMCPDT always gave slightly lower powers than did use of true allele frequencies, as one may expect (49% vs. 63% increase for missing patterns following the OSUMS data and 170% vs. 244% increase for setting all founders' genotypes as missing).

In summary, we have proposed a disequilibrium test that is applicable to pedigrees and X-chromosomal markers. This basic test was further extended to accommodate missing genotypes through an MC estimation procedure. The developed methods can easily be adapted for autosomal markers, although extensive analytical analysis and simulation are necessary to assess robustness and power gains, which will be performed in a separate study. One key assumption that we have made is that the pedigrees in a study are assumed to be drawn from a population of (extended) families, each of which has at least one affected offspring. Otherwise, bias may exist, especially when all families have the same structure and affection pattern, which, fortunately, is not the case in a genetic study that collects pedigrees of all shapes and sizes and affection patterns. To gauge the magnitude of such potential bias, we studied a nuclear family with six affected children (three males and three females) under conditions of setting 9 in table 1. We considered all the 15 nonredundant configurations of affection patterns and found that all of the expectations are close to zero. The largest in magnitude is 0.008, which corresponds to the case that all three male but no female children are affected. Thus, if all the families in a study are phenotypically the same, then the test might be biased and could lead to inflated type I error, although note that this is not always the case; for instance, a study with a triad design renders no bias. However, in a genetic study with pedigree data, bias should be negligible, and the proposed test statistic may be safely used. This is demonstrated through our simulation study based on the pedigree data from the OSUMS study.

We have shown that if missing-marker genotypes are simulated using correct marker-allele frequencies, a substantial gain in power over several competitive methods can be observed, especially when there is a large proportion of missing genotypes. However, it should be noted that, although PDT and XPDT are robust to population substructuring in the sample, their MC counterparts will be influenced by the population stratification. In such a case, a modification of the test is necessary to take such population stratification explicitly into account. Nevertheless, in the case that the underlying population is relatively homogeneous, XMCPDT is a useful alternative, especially when there is insufficient information to detect association from the observed genotypes alone.

Acknowledgments

This work was supported in part by National Institutes of Health grants HG002657 and NS046696, National Science Foundation grant DMS-0306800, and the Ohio State Biomedical Research and Technology Transfer grant Tech 05-062. The authors thank two anonymous referees for constructive comments, which have led to clearer presentation of the materials.

Appendix A

To show that $E(D_{MC}) = 0$, we first note that, for a single pedigree with complete genotypes, the D statistic can be written as a weighted sum of contributions from each offspring (nonfounder). Suppose that there are N offspring in the pedigree. For offspring j, define

$$X_j = (\#\ \text{M1 transmitted from mother}) - (\#\ \text{M1 not transmitted from mother}) ,$$

without restriction to affected offspring. Equivalently,

$$X_j = 2 \times (\#\ \text{M1 transmitted from mother}) - (\#\ \text{M1 in the mother}) .$$

It is apparent that X_j depends on the genotypes only. Then, for the whole pedigree, we can write

$$D(G_m, G_o, A) = \sum_{j=1}^{N} C_j(A) X_j(G) ,$$

where C_j is a coefficient that depends on only A. Specifically, $C_j(A)$ can be decomposed into the sum of coefficients from a family triad and DSPs. Suppose there are S offspring who are the same-sex siblings of offspring j. If j is affected, his or her contribution of the X_T statistic defined above is the same as X_j. If j is affected and his or her sibling l is unaffected, then, for this DSP, the X_S statistic defined above is

$$\begin{aligned} X_S &= (\#\ \text{M1 in offspring } j) - (\#\ \text{M1 in sibling } l) \\ &= (\#\ \text{M1 in } j \text{ transmitted from mother}) - (\#\ \text{M1 in } l \text{ transmitted from mother}) \\ &= \frac{1}{2}(X_j - X_l) . \end{aligned}$$

So, we have

$$\begin{aligned} C_j(A) = {} & I(\text{Offspring } j \text{ is affected}) \\ & + \sum_{l=1}^{S} \frac{1}{2} I(\text{Offspring } j \text{ is affected and sibling } l \text{ is unaffected}) \\ & - \sum_{l=1}^{S} \frac{1}{2} I(\text{Offspring } j \text{ is unaffected and sibling } l \text{ is affected}) \ . \end{aligned}$$

For D_{MC}, let $D_{\mathrm{MC}j}$ denote the contribution from offspring j. Then,

$$\begin{aligned} E(D_{\mathrm{MC}j}) &= \sum_{A \in \mathcal{A}} \sum_{G_{\mathrm{o}} \in \mathcal{G}_{\mathrm{o}}} \sum_{G_{\mathrm{m}} \in \mathcal{G}_{\mathrm{m}}} D_j(G_{\mathrm{m}}, G_{\mathrm{o}}, A) P(G_{\mathrm{m}} | G_{\mathrm{o}}) P(G_{\mathrm{o}} | A) P(A) \\ &= \sum_{A \in \mathcal{A}} \sum_{G_{\mathrm{o}} \in \mathcal{G}_{\mathrm{o}}} \sum_{G_{\mathrm{m}} \in \mathcal{G}_{\mathrm{m}}} C_j(A) X_j(G_{\mathrm{m}}, G_{\mathrm{o}}) P(G_{\mathrm{m}} | G_{\mathrm{o}}) P(A | G_{\mathrm{o}}) P(G_{\mathrm{o}}) \\ &= \sum_{G_{\mathrm{o}} \in \mathcal{G}_{\mathrm{o}}} \sum_{G_{\mathrm{m}} \in \mathcal{G}_{\mathrm{m}}} X_j(G_{\mathrm{m}}, G_{\mathrm{o}}) P(G_{\mathrm{m}} | G_{\mathrm{o}}) P(G_{\mathrm{o}}) \sum_{A \in \mathcal{A}} C_j(A) P(A | G_{\mathrm{o}}) \ . \end{aligned}$$

Note that, with our specification of C_j, we have

$$\begin{aligned} \sum_{A \in \mathcal{A}} C_j(A) P(A | G_{\mathrm{o}}) = {} & P(\text{Offspring } j \text{ is affected} | G_{\mathrm{o}}) \\ & + \sum_{l=1}^{S} \frac{1}{2} P(\text{Offspring } j \text{ is affected and sibling } l \text{ is unaffected} | G_{\mathrm{o}}) \\ & - \sum_{k=1}^{S} \frac{1}{2} P(\text{Offspring } j \text{ is unaffected and sibling } l \text{ is affected} | G_{\mathrm{o}}) \ . \end{aligned}$$

Under the null hypothesis of no association,

$$P(\text{Individual } j \text{ is affected and sibling } l \text{ is unaffected} | G_{\mathrm{o}})$$

is equal to

$$P(\text{Individual } j \text{ is unaffected and sibling } l \text{ is affected} | G_{\mathrm{o}})$$

for all $l, l = 1, \ldots, S$. Thus,

$$\begin{aligned} \sum_{A \in \mathcal{A}} C_j(A) P(A | G_o) &= P(\text{Offspring } j \text{ is affected} | G_{\mathrm{o}}) \\ &= P(\text{Offspring } j \text{ is affected}) \ (= a_j) \ . \end{aligned}$$

The second equality is true under the null hypothesis. Consequently,

$$\begin{aligned} E(D_{\mathrm{MC}j}) &= \sum_{G_{\mathrm{o}} \in \mathcal{G}_{\mathrm{o}}} \sum_{G_{\mathrm{m}} \in \mathcal{G}_{\mathrm{m}}} X_j(G_{\mathrm{m}}, G_{\mathrm{o}}) P(G_{\mathrm{m}} | G_{\mathrm{o}}) P(G_{\mathrm{o}}) a_j \\ &= a_j E(X_j) \\ &= 0 \ , \end{aligned}$$

where $E(X_j) = 0$, because each offspring has an equal chance of getting either one of the mother's alleles. Finally, we have

$$E(D_{\mathrm{MC}}) = \sum_{j=1}^{n} E(D_{\mathrm{MC}j}) = 0 \ .$$

Web Resources

The URLs for data presented herein are as follows:

MC-PDT, http://www.stat.ohio-state.edu/~statgen/SOFTWARE/MC-PDT/
Michael Knapp's Web site, http://www.uni-bonn.de/~umt70e/soft.htm
R Project for Statistical Computing, http://www.r-project.org/

References

1. 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
2. Spielman RS, Ewens WJ (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 62:450–458
3. Knapp M (1999) The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/disequilibrium test. Am J Hum Genet 64: 861–870
4. Clayton D (1999) A generalization of the transmission/disequilibrium test for uncertain-haplotype transmission. Am J Hum Genet 65:1170–1177
5. Martin ER, Monks SA, Warren LL, Kaplan NL (2000) A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67:146–154
6. Horvath S, Xu X, Laird NM (2001) The family based association test method: strategies for studying general genotype-phenotype associations. Eur J Hum Genet 9:301–306
7. Ho GYF, Bailey-Wilson JE (2000) The transmission/disequilibrium test for linkage on the X chromosome. Am J Hum Genet 66:1158–1160
8. Horvath S, Laird NM, Knapp M (2000) The transmission/disequilibrium test and parental-genotype reconstruction for X-chromosomal markers. Am J Hum Genet 66:1161–1167
9. Zhou QM, Rammohan K, Lin SL, Robinson N, Li O, Liu XL, Bai XF, Yin LJ, Scarberry B, Du PS, You M, Guan KL, Zheng P, Liu Y (2003) CD24 is a genetic modifier for risk and progression of multiple sclerosis. Proc Natl Acad Sci USA 100: 15041–15046
10. Ott J (1989) Computer-simulation methods in human linkage analysis. Proc Natl Acad Sci USA 86:4175–4178
11. Weeks DE, Ott J, Lathrop GM (1990) SLINK: a general simulation program for linkage analysis. Am J Hum Genet Suppl 47:A204