Am. J. Hum. Genet. 65:229–235, 1999

Methods for Detection of Parent-of-Origin Effects in Genetic Studies of Case-Parents Triads

Clarice R. Weinberg

Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC

Summary

When affected probands and their biological parents are genotyped at a candidate gene or a marker, the resulting case-parents–triad data enable powerful tests for linkage in the presence of association. When linkage disequilibrium has been detected in such a study, the investigator may wish to look further for possible parent-of-origin effects. If, for example, the transmission/disequilibrium test restricted to fathers is statistically significant, whereas that restricted to mothers is not, the investigator might interpret this as evidence for nonexpression of the maternally derived disease gene—that is, imprinting. This report reviews existing methods for detection of parent-of-origin effects, showing that each can be invalid under certain scenarios. Two new methods are proposed, based on application of likelihood-based inference after stratification on both the parental mating type and the inherited number of copies of the allele under study. If there are no maternal genetic effects expressed prenatally during gestation, the parental-asymmetry test is powerful and provides valid estimation of a parent-of-origin parameter. For diseases for which there could be maternal effects on risk, the parent-of-origin likelihood-ratio test provides a robust alternative. Simulations based on an admixed population demonstrate good operating characteristics for these procedures, under diverse scenarios.

Introduction

Studies based on the genotyping of affected probands and their parents provide a powerful approach for detection of either a direct effect of a candidate gene or linkage disequilibrium between a marker and a presumptive nearby gene involved in disease etiology (Falk and Rubinstein 1987; Self et al. 1991; Spielman et al. 1993; Spielman and Ewens 1996). The design is inherently robust against spurious associations that can sometimes arise in a genetically admixed population.

For a diallelic marker, the data from cases and their parents can validly be analyzed by either the transmission/disequilibrium test (TDT) (Spielman et al. 1993) or likelihood-based methods (Schaid and Sommer 1993; Self et al. 1991; Weinberg et al. 1998). The likelihood-ratio test (LRT) can outperform the TDT under either a dominant model or a recessive model (Schaid 1999; Weinberg et al. 1998), can be extended to allow for possible prenatal maternal genetic effects mediated through the maternal phenotype (Weinberg et al. 1998; Wilcox et al. 1998), and can efficiently exploit information from triads that are incomplete because of a missing parent (Weinberg 1999).

Once a gene has been shown to be related to risk of disease by these methods, the investigator might wish to search further, for possible evidence of parent-of-origin effects. For example, if the evidence for transmission distortion to affected offspring is stronger for mothers than for fathers, then this difference would suggest that the maternally derived allele is more fully expressed than is the paternally derived allele, perhaps because of differentially active "epigenetic" expression-regulator mechanisms, such as methylation.

A number of statistical techniques have been used for detection of parent-of-origin effects, on the basis of case-parents triads. The purpose of this study is first to critique the methods that are now in use or that have been proposed. I then describe two new approaches to detection of parent-of-origin effects, and I characterize their power and other properties, via simulations.

Background

A log-linear model has been developed in previous reports (Weinberg et al. 1998, 1999; Wilcox et al. 1998). In brief, when a case-parents triad is genotyped and jointly classified according to the number of copies of a particular allele carried by the mother, father, and child

Received March 11, 1999; accepted for publication May 4, 1999; electronically published June 8, 1999.

Address for correspondence and reprints: Dr. Clarice R. Weinberg, Biostatistics Branch, MD A3-03, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC 27709. E-mail: weinberg@niehs.nih.gov

0002-9297/99/6501-0029$02.00

(to be denoted hereafter as "M," "P," and "C," respectively), there are 15 possible outcomes. The family-specific outcomes (i.e., the cell into which a particular triad is classified) are independent, provided that each family contributes only one case. The counts based on classification of the triads studied can therefore be thought of as distributed according to a 15-cell multinomial.

Consider first a candidate gene for which two inherited copies of a variant allele increase the child's risk by a factor of R_2, whereas a single copy increases risk by a factor R_1. The risk could also depend on the number of copies carried by the mother, through prenatal effects, and the maternally mediated relative risks will be denoted as S_1 and S_2. To achieve robustness against population stratification, stratification by parental mating type is imposed, as proposed by Schaid and Sommer (1993). The mating type is defined by the parental combination, M,P, and symmetry is assumed across parents within each mating type; that is, the proportion of parents in the population who are (M = m,P = p) is assumed to equal the proportion of parents who are (M = p,P = m).

If there are parent-of-origin effects, then a slight modification of the model is required. Let R_p denote the relative risk associated with a single copy inherited from the father, and suppose that a single copy inherited from the mother confers a relative risk of $I_M R_p$. The parameter I_M will be 1 if and only if there are no parent-of-origin effects; I_M will be >1 if a maternally derived copy is associated with a greater increase in risk than a paternally derived copy and will be <1 if a maternally derived copy is associated with a smaller increase in risk than a paternally derived copy. Table 1 shows the expected counts in the 15 cells under such a model, which corresponds to a log-linear Poisson model. This model is similar to that of table 4 in the report by Weinberg et al. (1998) but follows a different and more parsimonious parameterization.

If the gene under study is not related to disease risk, either directly or through linkage disequilibrium or maternal effects, then R_1, R_2, I_M, S_1, and S_2 are all 1. Specific hypotheses about effects of the inherited gene—for example, $R_1 = R_2 = 1$—can be tested by comparing the fits of suitably nested models (with maximum-likelihood fits obtained by standard software packages such as SAS) and performing the corresponding likelihood-ratio χ^2 (LRT). In particular, the LRT for $R_1 = R_2 = 1$ tests the same hypothesis as does the TDT.

Although forming a valid basis for statistical testing, the model of table 1 is not strictly correct for a marker under alternatives to the null hypothesis. This problem arises because the risk may depend on parental genotypes even after statistical conditioning on the inherited genotype, because of possible recombinations during formation of the gametes. Nevertheless, the model is correct under the null hypothesis, and thus testing is valid, whether the gene is regarded as a candidate gene or as a marker.

Table 1

Frequencies in Case-Parents Triads

No. of Variant Alleles	Parental Origin (M and/or P)	Mating Type	Theoretical Frequency[a]
222	MP	1	$S_2R_2\mu_1$
212	MP	2	$S_2R_2\mu_2$
122	MP	2	$S_1R_2\mu_2$
211	M	2	$I_MS_2R_p\mu_2$
121	P	2	$S_1R_p\mu_2$
201	M	3	$I_MS_2R_p\mu_3$
021	P	3	$R_p\mu_3$
112	MP	4	$S_1R_2\mu_4$
111	M or P	4	$S_1(1 + I_M)R_p\mu_4$
110		4	$S_1\mu_4$
101	M	5	$I_MS_1R_p\mu_5$
011	P	5	$R_p\mu_5$
100		5	$S_1\mu_5$
010		5	μ_5
000		6	μ_6

[a] R_2, S_1, S_2 are relative risks for C = 2, M = 1, M = 2, compared with no copies in the child or mother. R_p is the relative risk for a single paternally inherited copy (relative to no copies in the child or mother), and I_MR_p is the relative risk for a single maternally inherited copy. $I_M = 1$ if and only if there is no parent-of-origin effect.

By contrast, the TDT considers only transmissions from parents who are heterozygous and is computationally simpler to perform. Let N_{MPC} denote the number of families that fall into genotype category MPC. Then the number of heterozygous parents who have transmitted the allele to their affected child is given by $T = N_{212} + N_{122} + 2N_{112} + N_{111} + N_{011} + N_{101}$, whereas the number of heterozygous parents who have not done so is given by $NT = N_{211} + N_{121} + 2N_{110} + N_{111} + N_{010} + N_{100}$. If we assume Mendelian inheritance of the gene, then, under the null hypothesis that the allele under study is either not in linkage with any disease-related gene or not associated with the disease, T should be binomially distributed on $T + NT$, with parameter .5. In particular, if the gene is unrelated to the disease within all subpopulations, then transmission should be random—that is, Mendelian—regardless of whether the offspring has the disease. The TDT tests this null hypothesis, with a statistic whose distribution is approximately χ^2 (1 df) under the null hypothesis: $(T - NT)^2/(T + NT)$.

Existing Methods for Detection of Parent-of-Origin Effects

When a TDT or likelihood-based analysis has revealed significant evidence for linkage/association for a particular allele, and when the investigator wishes to explore

parent-of-origin effects, a natural approach is to stratify the transmission/nontransmission counts according to whether the mother or the father is the source. (The N_{111} cell usually needs to be omitted, because the source of the inherited copy is ambiguous.) The total number of usable paternal transmissions would be $T_p = N_{212} + N_{112} + N_{011}$, whereas the number of usable paternal nontransmissions would be $NT_p = N_{211} + N_{110} + N_{010}$; the corresponding counts for the mother would be $T_M = N_{122} + N_{112} + N_{101}$ and $NT_M = N_{121} + N_{110} + N_{100}$. Using the resulting 2 × 2 table of counts, one can test whether the apparent distortion is more pronounced for one parent than for the other. The test for differential distortion in transmission is usually based on either Fisher's exact test or, for larger samples, the χ^2 (1 df) for testing the equality of two proportions. This general approach will be denoted as the "TDT_{MvsF}" approach.

One problem with the TDT_{MvsF} is that, when both parents are heterozygous for the variant allele and there is linkage/association, then the transmissions from the two parents to an affected offspring are not statistically independent. To see why this is so, consider a scenario in which there are no parent-of-origin effects (i.e., the null hypothesis for this test) for a candidate gene. For simplicity, assume that there are also no maternal effects. Then, for parents who are both heterozygous, the probability of maternal transmission given paternal transmission can be shown to be $R_2/(R_2 + R_1)$, whereas the probability of maternal transmission given *no* paternal transmission is $R_1/(1 + R_1)$. The dependency between the parental transmissions implies that Fisher's exact test and the χ^2 for equality of proportions are both invalid.

To avoid this problem, Weinberg et al. (1998) have considered what they call the "transmission asymmetry test" (TAT), which forms the same 2 × 2 table but omits data from parents who are both heterozygous (the N_{110} and N_{112} counts). This data reduction ensures independence of the parental transmissions, because each family now contributes, at most, one heterozygous parent to the analysis. The resulting test was found to have very low power compared with the alternative, likelihood-based approach (Weinberg et al. 1998).

Another problem is that both the TDT_{MvsF} and the TAT can be invalid (with a type I–error rate that differs from the nominal level) if there are maternal effects. On the basis of the expected counts given in table 1, even when $I_M = 1$ (the null hypothesis for parent-of-origin effects), for the TAT the paternal ratio of expected counts for T_p:NT_p would be $(S_2R_2\mu_2 + R_1\mu_5)$:$(S_2R_1\mu_2 + \mu_5)$, whereas the corresponding maternal ratio would be $(R_2\mu_2 + R_1\mu_5)$:$(R_1\mu_2 + \mu_5)$. These ratios will not necessarily be equal under the null hypothesis, because maternal effects can introduce differential weighting of the maternal and paternal transmissions. For example, if there is a maternal effect with $S_2 > 1$ and if $R_1 < R_2$, then the preferential transmission to an affected offspring will appear to be higher maternally than paternally, giving spurious evidence for a parent-of-origin effect. A similar distortion would be seen with the TDT_{MvsF}. For this reason, methods based on different frequencies of transmission from the mothers and fathers in case-parents analyses can be invalid, unless the investigator can be confident that there are no maternally mediated genetic effects.

An appealing alternative is to fit the complete model corresponding to table 1 and apply likelihood-based testing to the question (Weinberg et al. 1998). One should be able to fit a background model that incorporates the possibility that there are both maternal effects and effects of the number of inherited copies and then to fit an extended model that also incorporates possible effects of the parent of origin, by allowing values other than 1 for the I_M parameter. The comparison of the two models should theoretically yield a 1 df χ^2 statistic that is not subject to the validity problems described above.

There is, however, a subtlety not fully appreciated in earlier studies (Weinberg et al. 1998). If the gene under study is in linkage disequilibrium with a different disease-susceptibility gene, then the formulation of table 1 is not strictly correct, for reasons mentioned already. This means that a background model that includes R_1 and R_2, as in table 1, may not be quite correctly specified; and it follows that the LRT for imprinting may not be quite valid if the gene under study is a marker rather than a candidate gene.

For easy reference, the issues that plague each of the three methods are summarized in table 2. Clearly, the challenge is to develop a valid test for parent-of-origin effects that is valid in the presence of maternal effects and that remains valid if the gene under study is a marker that may be in linkage disequilibrium with a disease-susceptibility gene.

Proposed Method

The proposed method is based on consideration of the three mating types in which the mother and father carry unequally many copies of the variant allele, with further stratification on the number of inherited copies of the allele, C. This second level of conditioning (on C) effectively removes any effects related jointly to the inherited number of copies and the parental-allele counts M,P. The probability that M > P is then expressible as is shown in table 3 and depends only on I_M, S_1, and S_2. For example, in the mating type in which one parent is homozygous and the other is heterozygous and the child has inherited two copies of the allele (rows two and three in table 1), the probability that M > P is obtained by dividing the corresponding theoretical frequencies, yielding $S_2/(S_1 + S_2)$. (For strata in which the affected child

Table 2

Assumptions Required for Existing Tests to Be Valid

Method of Testing for Parent-of-Origin Effects[a]	Assumption(s) Required for Validity
TDT_{MvsF}	Never valid, unless there is no linkage/association for the gene under study, in which case, testing for parent-of-origin effects is of little interest; moreover, a rejection of the null hypothesis would logically imply invalidity of the test on which the rejection was based
TAT	Valid in the absence of maternal effects, provided that inheritance of allele is Mendelian in the population and that there is parental symmetry within mating types in the population studied
LRT based on full log-linear model of table 1	Valid if inheritance of allele is Mendelian, if there is parental symmetry within mating types in the population studied, and if the gene under study is not in linkage disequilibrium with another disease-susceptibility gene

[a] All are 1 df χ^2 tests.

has inherited at least one copy of the allele, M > P corresponds to maternal transmission.) The corresponding odds—that is, Pr[M > P]/Pr[P > M]—is given in the last column of table 3 and takes a simple linear form, under a logarithmic transformation. Notice that all dependence on R_p and R_2 (and possible gametic recombination) has been removed by simultaneously conditioning on the parental mating type and C.

The corresponding logistic model is as follows:

$$\ln\left(\frac{\Pr[\mathrm{M}>\mathrm{P}|\text{mating type,C}]}{\Pr[\mathrm{M}<\mathrm{P}|\text{mating type,C}]}\right) = \alpha I_{(\mathrm{C}=1)} + \beta I_{(\mathrm{M+P}>1)} + \gamma[I_{(\mathrm{M+P}=1)} - I_{(\mathrm{M+P}>2)}] \ , \quad (1)$$

where $I_{(\text{comparison statement})} = 1$ when the comparison statement holds and is 0 otherwise. The parameters in model (1) have the following interpretation: e^{α} is I_M, e^{β} is S_2,and e^{γ} is S_1. The test for parent-of-origin effects can then be based on testing the null hypothesis that $\alpha = 0$—for example, by means of an LRT. Model (1) is an unconditional logistic-regression formulation, and standard software can be used. The user is cautioned, however, that the model must be specified to exclude any intercept parameter. The corresponding LRT for testing $\alpha = 0$ will hereafter be referred to as the "parent-of-origin LRT" (PO-LRT). The model can also be used for estimation of the parent-of-origin parameter, I_M, and a confidence interval (CI) is easily constructed that will, for example, cover the true parameter 95% of the time, provided that the sample size is adequate. One can also include allowance for a possible interaction between a parent-of-origin effect and an exposure, by extending model (1) appropriately.

Table 3

Frequencies for Testing for Parent-of-Origin Effects

Stratum Defined by Parental Mating Type and C (MPC)	Conditional (on Stratum) Probability That M > P	Odds That M > P
$212 \cup 122$	$\frac{S_1}{S_1+S_2}$	$\frac{S_2}{S_1}$
$211 \cup 121$	$\frac{I_M S_2}{I_M S_2+S_1}$	$\frac{I_M S_2}{S_1}$
$201 \cup 021$	$\frac{I_M S_2}{I_M S_2+1}$	$I_M S_2$
$101 \cup 011$	$\frac{I_M S_1}{I_M S_1+1}$	$I_M S_1$
$100 \cup 010$	$\frac{S_1}{1+S_1}$	S_1

Model (1) as presented in table 3 can also be used to estimate S_1 and S_2 in the presence of possible effects of the inherited gene. This estimation avoids possible validity problems with the likelihood-based adjusted estimation described above, which, for reasons discussed above, will only be strictly valid for a candidate gene (Weinberg et al. 1998; Wilcox et al. 1998) and not for a marker.

If the investigator is willing to assume that there are no maternal effects—that is, that $S_1 = 1$ and $S_2 = 1$—then the first and last row of table 3 are omitted as noninformative, and the middle three rows then each have odds I_M for maternal transmission. In this situation, the test of $I_M = 1$ reduces to testing whether the probability of maternal transmission for those three strata is .5, and a TDT-like statistic can be used. The test would be based on the binomial, probability .5 distribution of the count $A = N_{211} + N_{201} + N_{101}$ from the total $A + N_{121} + N_{021} + N_{011}$. If we let $B = N_{121} + N_{021} + N_{011}$, then the test statistic takes the simple form $(A - B)^2/(A + B)$, which should have a χ^2 distribution with 1 df, under the null hypothesis that there are no parent-of-origin effects. Notice that, in contrast to the TAT, this test uses only heterozygous cases and includes counts for transmissions from homozygous parents. This procedure will be termed the "parental-asymmetry test" (PAT).

In some situations—for example, when there is a sufficiently dense marker scan—the investigator may be able to deduce the parent of origin for triads in which all three individuals are heterozygous, by using the haplotype data. Such transmission information can then be

exploited by including a sixth row in table 3, where what would be modeled for that row is whether the mother is the source of the inherited allele. Maternal transmission for the (1,1,1) stratum would carry an odds of I_M, with no dependence on either S_1 or S_2. Thus, when parental source can be determined, data from the (1,1,1) stratum can be used regardless of whether the investigator does or does not wish to allow for possible maternal effects, by use of either the PO-LRT or the PAT, respectively.

Simulations

I simulated data from an admixed population, using the same structure as had been used in a previous study (Weinberg et al. 1998). In brief, a 20% subpopulation had an allele prevalence of .3 and a background risk of .05 (i.e., risk in those who had no copies and who had a mother with no copies), whereas the complementary 80% subpopulation had an allele prevalence of .1 and a background risk of .01. Each of the two subpopulations was assumed to be in Hardy-Weinberg equilibrium, although the mixed population as a whole was not.

Case-parents–triad data were simulated under scenarios with and without a maternal contribution to risk and with and without parent-of-origin effects. The simulations assumed that the triads in which all three had one copy of the variant allele were indeterminate as to the parent of origin and that they therefore had to be excluded from analyses. The characteristics of the TAT were compared with those of the PO-LRT, and estimation based on model (1) was assessed by (*a*) comparison of the average estimated I_M versus the known true value and (*b*) assessment of the proportion of simulated studies for which the nominal 95% CI covered the true parameter, I_M. The simulated data included various numbers of case-parents triads (range 100–400), although fewer than the listed *N* would have been informative in the analyses performed. Each scenario was simulated 1,000 times, to allow fairly precise estimation of the type I–error rates, the powers, and the CI coverage rates. If the true coverage rate is 95%, then the standard error for the percent of intervals that cover, on the basis of 1,000 simulations, is ~0.7%, implying that observed coverage rates of 93.6%–96.4% are quite consistent with the nominal 95%.

The first set of simulations assessed behavior of the three χ^2 tests in scenarios in which there were no parent-of-origin effects, some of which included maternal effects and some of which did not. The scenarios were all null with respect to parent-of-origin effects.

The second set of simulations assessed the performance of the tests under alternatives to the null hypothesis. One scenario provided that the allele from the father was completely silenced. This was done by setting the parameter vector (R_1,R_2,S_1,S_2,I_M) to (1,4,1,1,4). In this situation, a single copy from the father has no effect on risk, whereas a single copy from the mother is associated with a quadrupling of risk. The complement, (4,4,1,1,.25)—in which it is the allele from the father that is associated with a quadrupling of risk and the maternal allele is silent—can be shown (on the basis of a mathematical-symmetry argument) to have exactly the same characteristics as (1,4,1,1,4) and does not need to be assessed separately.

Another kind of scenario considered was one in which there is not complete silencing of the allele from either parent but in which there is differential expression. This was done by assignment of the parameter values (2,5,1,1,2)—that is, a single copy from the father confers a doubling of risk, whereas a single copy from the mother confers a quadrupling of risk. A third scenario provides differential expression but no increase in the risk to the child inheriting two copies, rather than one copy, from the mother. This was done by use of parameters (2,6,1,1,3).

Finally, a set of simulations was performed for scenarios in which the only valid testing procedure is PO-LRT, because of the presence of both maternal effects and effects of the inherited gene. These simulations were performed to assess bias, CI coverage, and power for likelihood methods based on model (1). In the presence of maternal effects, it does matter whether it is the maternally derived or paternally derived allele that is silenced, so, for each scenario, the complementary scenario was also simulated.

Results of Simulations

Results of simulations under the null hypothesis—that is, when there are no parent-of-origin effects—are shown in table 4. In the absence of maternal effects (table 4, rows 1 and 2), all three testing procedures had empirical type I–error rates that are consistent with the nominal .05. In the presence of maternal effects, the type I–error rate of the PAT was badly biased to >.05, reflecting the requirement that, for that test to be valid, there be no maternal effects. The TAT becomes invalid too, although less dramatically, as seen in the last five rows. For instance, for 400 triads with the scenario (1,3,1,3,1), the type I–error rate for the TAT is .136, revealing significant bias for the transmission-based test. By contrast, the PO-LRT remains valid even in the presence of maternal effects.

The likelihood given in table 3 also permits estimation, and the estimation of the parent-of-origin parameter, I_M (estimated on the logarithmic scale), was generally unbiased, on the basis of the means and standard errors (data not shown). The CI coverage was consistent with the nominal 95%.

Table 4

Results of Simulations of Scenarios with No Parent-of-Origin Effects

Scenario R_1,R_2,S_1,S_2,I_M; No. of Triads (1,000 Simulations/Scenario)	Empirical Type I Error with[a]			Proportion of 95% CIs That Cover I_M[b]
	TAT	PAT	PO-LRT	
2,2,1,1,1; 100	.048	.047	.052	.961
2,2,1,1,1; 200	.056	.042	.057	.945
1,1,1,3,1; 100	.062	.327	.056	.946
1,1,1,3,1; 200	.050	.596	.054	.949
1,1,3,3,1; 100	.050	.694	.050	.962
1,1,3,3,1; 200	.055	.939	.049	.955
2,2,1,3,1; 100	.064	.398	.068	.942
2,2,1,3,1; 200	.067	.714	.059	.944
1,3,1,3,1; 100	.067	.286	.057	.952
1,3,1,3,1; 200	.081	.490	.048	.952
1,3,1,3,1; 400	.136	.797	.062	.940

[a] All tests are two sided, nominally set at level .05.
[b] CIs are nominally 95%.

Table 5 shows results for scenarios that included parent-of-origin effects but that did not include maternal effects, the only sort of scenario in which all three tests are valid. The TAT was somewhat more powerful than the PO-LRT, under every such scenario considered. But the PAT was much more powerful than either the TAT or the PO-LRT. Estimation appeared to be slightly biased toward values away from 0 (on the basis of means of the estimates; data not shown), but CI coverage based on model (1) was, if anything, conservative, showing a tendency to cover the true parameter with a frequency slightly >.95.

Table 6 shows results for scenarios that included both parent-of-origin effects and maternal effects. Only the PO-LRT is valid for such scenarios, so results are not given for either the PAT or the TAT. CI coverage was again statistically consistent with the nominal 95%, indicating that estimation works well. In the presence of deleterious maternal effects, silencing of the maternally derived allele was easier to detect than the complementary silencing of the paternally derived allele (i.e., power in rows 4–6 is better than that in rows 1–3). Differential expression was also easier to detect when it was the paternally derived allele, rather than the maternally derived allele, that was overexpressed (rows 7–9 versus rows 10–12).

Discussion

Although investigators already familiar with the TDT will find it natural to look for parent-of-origin effects by comparing the frequency of transmission from heterozygous mothers versus that from heterozygous fathers, the development of a valid test based on differential transmission to offspring is problematic. If triads in which both parents are heterozygous are included (as they are in the TDT), then they can contaminate the comparison, because of statistical dependency between maternal and paternal transmissions. This problem does not affect the TDT itself, because the null hypothesis for the TDT is that the meiotic selection of an allele for transmission is simply random, with probability .5. For parent-of-origin effects, we do not begin with a null hypothesis of randomness but, rather, with a null hypothesis of *equality* of transmission (at some value that may not be .5) for the mother and the father. In this way, it is the existence of a real effect of the candidate gene (or linkage disequilibrium for a marker) that invalidates the TDT_{MvsF}.

Although exclusion of those homozygous/homozygous parents has seemed to be a simple solution to the problem (Weinberg et al. 1998), the apparent transmission rates can be different for mothers compared with fathers, if there are prenatal maternal effects on risk, even when there are no parent-of-origin effects. Thus, the test based on simple stratification of the TDT transmission-count data (TDT_{MvsF}) is always invalid, whereas the reduced-data version, the TAT, will be invalid if there are maternal effects.

A likelihood-based approach, based on simultaneous stratification on both mating type and C, the inherited number of copies of the allele, provides a reliably valid alternative. The within-strata distribution of counts (table 3) depends only on the maternal parameters S_1 and S_2 and on the parent-of-origin parameter I_M. The primary benefit of this reduction (compared with an LRT based the model presented in table 1) is that the testing is valid whether the gene under study is either a candidate gene or a marker that is possibly in linkage disequilibrium with a nearby causative gene.

If the investigator is comfortable in assuming that there are no maternal effects, then the TAT can be presumed to be valid. However, the simulations of table 5 demonstrate that, under that same assumption, the PAT has much better power than the TAT.

Table 5

Results of Simulations of Scenarios with Parent-of-Origin Effects

Scenario R_1,R_2,S_1,S_2,I_M; No. of Triads (1,000 Simulations/Scenario)	Empirical Power with			Proportion of 95% CIs That Cover I_M
	TAT	PAT	PO-LRT	
1,4,1,1,4; 100	.601	.989	.552	.954
1,4,1,1,4; 200	.856	1.000	.831	.954
1,4,1,1,4; 400	.992	1.000	.986	.953
2,5,1,1,2; 100	.217	.615	.190	.953
2,5,1,1,2; 200	.400	.892	.335	.948
2,5,1,1,2; 400	.678	.993	.584	.952
2,6,1,1,3; 100	.464	.954	.348	.964
2,6,1,1,3; 200	.748	1.000	.596	.944
2,6,1,1,3; 400	.960	1.000	.889	.943

Note.—Conditions are as defined in the footnotes to table 4.

Table 6

Results of Simulations of Scenarios with Both Maternal and Parent-of-Origin Effects

Scenario R_1,R_2,S_1,S_2,I_M; No. of Triads (1,000 Simulations/Scenario)	Empirical Power with PO-LRT	Proportion of 95% CIs That Cover I_M
1,3,1,3,3; 100	.360	.961
1,3,1,3,3; 200	.606	.960
1,3,1,3,3; 400	.881	.947
3,3,1,3,1/3; 100	.401	.944
3,3,1,3,1/3; 200	.686	.963
3,3,1,3,1/3; 400	.926	.955
2,6,1,3,3; 100	.296	.963
2,6,1,3,3; 200	.555	.953
2,6,1,3,3; 400	.807	.951
6,6,1,3,1/3; 100	.361	.961
6,6,1,3,1/3; 200	.639	.941
6,6,1,3,1/3; 400	.914	.953

NOTE.—Conditions are as defined in the footnotes to table 4.

By contrast, the power of the PO-LRT is disappointing (table 5), so there is a price to pay for its robust validity. One might imagine that it could be improved if a hybrid stepwise procedure were performed: one could first test for a maternal effect based on model (1)—that is, the model as presented in table 3; if there were little evidence for a maternal effect—say $P > .10$, based on the 2 df χ^2 LRT for S_1 and S_2—then one would simply revert to the PAT. However, simulations (data not shown) revealed this to be a dangerous procedure, with an overall type I error rate that was significantly above the nominal .05. Thus, if, a priori, there is any doubt about possible maternal effects, the investigator is advised to use the PO-LRT.

The problem with use of the full log-linear model, on the basis of the expected counts shown in table 1, is that one must test the parent-of-origin hypothesis against a reduced-background model that includes R_1 and R_2, and this background model may not be properly specified if the gene under study is a marker. The same point holds for tests of maternal effects, adjusted for possible effects of the inherited gene (Weinberg et al. 1998; Wilcox et al. 1998). The model (1)–based LRT test of the null hypothesis that there are no maternal effects remains valid and appears to be comparable in power with results given on the basis of the full LRT (Wilcox et al. 1998), provided that one assumes that $I_M = 1$ (simulations not shown).

In summary, if the investigator is confident that there are no maternal effects for the disease under study, then the PAT offers excellent power for testing for parent-of-origin effects. For diseases that could be subject to prenatal maternal effects, the only valid testing procedure is the PO-LRT. Although the power of the PO-LRT (table 5) is somewhat disappointing, if data from heterozygous triads could be included (i.e., if the parental source of the allele could be determined), then the power of this likelihood-based procedure could be improved.

Acknowledgment

The author thanks Drs. Allen Wilcox and Norman Kaplan for their suggestions.

References

Falk CT, Rubinstein P (1987) Haplotype relative risks: an easy, reliable way to construct a proper control sample for risk calculations. Ann Hum Genet 51:227–233

Schaid DJ (1999) Likelihoods and TDT for the case-parents design. Genet Epidemiol 16:250–260

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

Self SG, Longton G, Kopecky KJ, Liang KY (1991) On estimating HLA-disease association with application to a study of aplastic anemia. Biometrics 47:53–61

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

Weinberg CR (1999) Allowing for missing parents in genetic studies of case-parent triads. Am J Hum Genet 64: 1186–1193

Weinberg CR, Wilcox AJ, Lie RT (1998) A log-linear approach to case-parent–triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet 62: 969–978

Wilcox AJ, Weinberg CR, Lie RT (1998) Distinguishing the effects of maternal and offspring genes through studies of "case-parent triads." Am J Epidemiol 148:893–901