The transmission disequilibrium test (TDT) is a useful method to detect linkage between a marker locus and a disease susceptible locus (1–6). For a marker locus with two alleles, the TDT compares the number of heterozygous parents who transmit one allele with the number of heterozygous parents who transmit the other allele to the affected offspring. It is not prone to produce false positive results due to population stratification (7). The TDT has been extended to markers with more than two alleles (2, 8). Recently, the TDT has been extended to sibships with at least one affected and one unaffected individual (9). The new test is referred as the S-TDT. For an allele of interest at a marker locus, the S-TDT essentially compares the frequency of that allele among affected individuals with the frequency of the allele among unaffected individuals. This procedure makes it possible to apply the S-TDT to diseases with late age at onset, such as non-insulin-dependent diabetes mellitus, cardiovascular diseases, Alzheimer’s disease, and other diseases related to aging.

For some disorders, such as alcoholism, often only one parent of affected individuals (most likely the mother) is available. If both affected and unaffected individuals are present within a sibship, we might use the S-TDT. However, in families in which all the offspring are affected and only one parent is available, neither the TDT nor the S-TDT can be used. Curtis and Sham (10) pointed out a problem when genotype data are available for only one parent. Suppose there are only two alleles “N” and “M” at the marker locus. If the offspring of a heterozygous parent (“NM”) is of the genotype “NN” (or “MM”), we can infer that the parent transmits N (or M) to the offspring. If the offspring is NM, we do not know whether the parent transmits N or M and this parent-offspring pair must be discarded. This method can induce biases depending on the allele frequencies of N and M even if M and N are equally transmitted. Therefore, information from families with one available parent should be discarded for biallelic markers using the above approach. This certainly can waste information. The objective of this paper is to propose...
two statistics whereby such families can be used in association or linkage studies without introducing bias under the null hypothesis of no association or no linkage. The first statistic, \( T_1 \), is applicable when either assumption A1: males and females with the same genotype at the marker locus have the same mating preference, or assumption A2: father and mother are missing with the same probability \( \frac{1}{2} \) given one of them is missing holds. When neither assumption holds, we give another statistic, \( T_2 \). We show in the Results section that \( T_2 \) is generally less powerful than \( T_1 \) when either assumption hold.

In some studies, there might be families in which both parental genotypes are available, others in which only one parental genotype is available, and still others in which neither parental genotype is available but genotypes of unaffected sibs are available. We propose a method by which data from studies involving those different groups of families can be combined.

**METHODS**

In our recent effort to find an accurate, noniterative method of estimating risk ratio for case-parental control design studies (11), we found an estimator for the risk ratio between individuals with genotype NM and those with genotype NN (\( \lambda_{NM} \)) and for the risk ratio between individuals with genotype NM and those with genotype MM (\( \lambda_{NM} \)) when only one parent is available. Throughout the paper, we assumed that the probability that a mother's genotype availability is independent of her genotype, with a similar assumption for a father. The estimator is approximate unbiased for the corresponding risk ratio as long as one of the following two assumptions holds:

**Assumption A1:** Males and females with the same genotype at the marker locus have the same mating preference;

**Assumption A2:** Father and mother in each nuclear family are missing with the same probability \( \frac{1}{2} \) given one of them is missing.

The estimator is approximate unbiased even if population stratification and nonrandom mating exist in the population under study. The estimator does not depend on the allele frequencies of alleles M and N. It can be described as follows. For simplicity, let us denote NN = 0, NM = 1, and MM = 2, where the number denotes the number of M alleles a genotype has. Let \( A_{ij} \), \( i, j = 0, 1, 2 \) be the number of case subjects with genotype \( i \) whose one available parent has genotype \( j \). With data summarized as in table 1, the estimator is given by

\[
\hat{\lambda}_{ij} = (A_{i1} + A_{i2} - A_{i0})/(2A_{i2}),
\]

\[
\hat{\lambda}_i = (A_{i1} + A_{i0} - A_{i2})/(2A_{i0}).
\]

Table 1. Case-parental control design when only one parental genotype is available

<table>
<thead>
<tr>
<th>Case genotype</th>
<th>Parental genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN(0)</td>
<td>( A_{00} )</td>
</tr>
<tr>
<td>NM(1)</td>
<td>( A_{10} )</td>
</tr>
<tr>
<td>MM(2)</td>
<td>0</td>
</tr>
</tbody>
</table>

If the disease is not associated with the marker locus, then \( \lambda_{ij} = \lambda_{i} = 1 \). If the disease is associated with the marker locus with \( M \) the susceptible allele, then \( \lambda_{i} > 1 \) and/or \( \lambda_{i} < 1 \). If the disease is associated with the marker locus with \( N \), the susceptible allele, then \( \lambda_{i} < 1 \) or/and \( \lambda_{i} > 1 \). Thus, we compare the sum of the numerator of \( \hat{\lambda}_i \) and the denominator of \( \hat{\lambda}_{ij} \), with the sum of denominator of \( \hat{\lambda}_i \) and the numerator of \( \hat{\lambda}_{ij} \). After simple algebraic manipulations, we found we should compare \( b_1 = A_{01} + A_{12} \) with \( c_1 = A_{10} + A_{21} \). If the disease is associated with the marker locus, then \( b_1 - c_1 \) is different from zero. If, for each nuclear family, only the genotypes of one affected child and one available parent are considered, the variance of \( b_1 - c_1 \) can be estimated by \( V = b_1 + c_1 \). The following statistic can be used to test the association of the disease with the marker locus:

\[
T_1 = \frac{A_{01} + A_{12} - A_{10} - A_{21}}{\sqrt{V}} = \frac{b_1 - c_1}{\sqrt{V}}.
\]

\( T_1 \) has an approximate standard normal distribution when \( b_1 + c_1 \) is large under the null hypothesis. From the value of \( T_1 \), we obtain the \( p \) value using normal approximation.

When \( b_1 + c_1 \) is not large, the normal approximation is not accurate. When this occurs, the exact method described below can be used to find the \( p \) value. Under the null hypothesis of no association, \( b_1 \) and \( c_1 \) should be the same. Conditional on \( b_1 + c_1 \), \( b_1 \) is a binomial random variable, Bin\((b_1 + c_1, \frac{1}{2})\). Then we can calculate the exact \( p \) value using binomial distribution.

For families with multiple affected offspring or multiple generations, the variance of \( b_1 - c_1 \) cannot be estimated by \( b_1 + c_1 \) because of the dependency among the genotypes of family members. To estimate the variance of \( b_1 - c_1 \), we consider each family separately. Suppose in the \( i \)th family, we calculate the corresponding \( b_i \) and \( c_i \) as above. Because the expectation of \( b_i - c_i \) is 0 under the null hypothesis, the variance of \( b_i - c_i \) can be estimated by \( (b_i - c_i)^2 \). The variance of the numerator in \( T_i \) can be estimated by

\[
V_i = \Sigma(b_i - c_i)^2.
\]
where the summation is over all the families. Note that the estimation of the variance of \( b_i - c_i \) is the same as given above when only one affected offspring and one parent are considered in each family. The test statistic will still be \( T_1 \). We refer to the above test as the 1-TDT.

**The 1-TDT when both assumptions A1 and A2 are violated**

In justifying \( T_1 \), we assumed that assumption A1 or A2 holds. If neither assumption A1 nor assumption A2 holds, \( T_1 \) may no longer have a standard normal distribution when no association exists between the disease and the marker locus. To overcome this problem, we propose another statistic which can be used to test the association between the disease and the marker locus even if the assumption is violated. This new test is expected to be less powerful than \( T_1 \) if either assumption A1 or assumption A2 holds. We support this claim using simulations in the Results section.

To motivate the new test, we consider another approximate unbiased estimator for the risk ratio presented in Sun et al. (11) for use even if neither assumption A1 nor assumption A2 holds. This estimator is given by replacing \( A_{ij} = \sum_{i,j} A_{ij} \) in equation 1 with

\[
A_{ij}' = M \times P_{ij} + P \times M_{ij},
\]

where \( P \) and \( M \) are the numbers of families with father and mother available, respectively. \( P_{ij} (M_{ij}) \) is the number of case subjects with genotype \( i \) and the father (mother) has genotype \( j \). Using the same idea as above, we can show that under the null hypothesis that there is no association between the disease and the marker locus

\[
E(M(pb_1 - pc_1) + P(mb_1 - mc_1)) = 0
\]

where

\[
pb_1 = P_{01} + P_{12}, \quad pc_1 = P_{10} + P_{21}, \quad mb_1 = M_{01} + M_{12}
\]

and

\[
mc_1 = M_{10} + M_{21}.
\]

Thus, we propose the following statistic:

\[
T_2 = \frac{M(pb_1 - pc_1) + P(mb_1 - mc_1)}{\sqrt{M^2 \times pV + p^2 \times mV}},
\]

where \( pV \) and \( mV \) can be estimated similarly as \( V \) in the above section by considering only families with the father and the mother available, respectively.

**The test as a test for linkage**

We derive the 1-TDT as a test for association between a marker locus and the disease. In this section, we show that, like the TDT, the 1-TDT is also a test for linkage under linkage disequilibrium.

For simplicity of exposition, we only prove this claim under a simple model of Ott (12), although the results hold for more general models. This model is based on the assumptions that the disease is recessive at the disease locus and that there are random mating and Hardy-Weinberg equilibrium among members of the parental generation. Following Ott (12), let \( p \) be the allele frequency of the disease allele \( D \) at the disease locus; \( q \) be the allele frequency of \( N \) at the marker locus; \( \delta = p(DN) - pq \) be the association between the disease locus and the marker locus; and \( \theta \) be the recombination fraction between the disease locus and the marker locus.

Under the above model, Ott (12) obtained the joint probabilities for the transmitted and the nontransmitted alleles as in table 2 (table 2 in Ott (12)), where

\[
r = (q + \delta/p)q, \quad s = (q + \delta/p)(1 - q) - \theta \delta/p,
\]

\[
t = (1 - q - \delta/p)q + \theta \delta/p,
\]

and

\[
u = (1 - q - \delta/p)(1 - q).
\]

From table 2, one can calculate the cell probabilities corresponding to the data in table 1. The result is given in table 3.

Suppose that there are \( n \) case subjects with only one parent available. From tables 1 and 3, we calculate the expected value of \( b_1 - c_1 \) as

\[
E(b_1 - c_1) = E(A_{01} + A_{12} - A_{10} - A_{21}) =
\]

\[
n((s + u)(r + s) - (r + t)(t + u)) = n(s - t) = n(1 - 2\theta)\delta/p.
\]

The quantity in the above equation is zero if and only if \( \theta = \frac{1}{2} \) or \( \delta = 0 \). Thus, the test can be used as a test for linkage under linkage disequilibrium.
families. If it is believed that either assumption A1 or A2 is suspected, we can use \( T_1 \). Then \( S_1 = M \times p_b + P \times m_b \) has mean \( M_1 = (M \times (p_b + p_c) + P \times (m_b + m_c))/2 \) and variance \( V_1/4 = (M^2 \times p^2 + P^2 \times m^2)/4 \). The notation is the same as above.

The S-TDT can be applied to the third group of families. Let \( A \) be the mean and \( V \) be the variance of the number, \( Y \), of allele N among affected individuals calculated as follows. In a sibship of \( a \) affected and \( u \) unaffected sibs, let \( r \) be the number of sibs having genotype NN and \( s \) be the number of sibs having genotype NM. Then the mean and variance of the number of N alleles among affected individuals under the null hypothesis are \((2r + s)a/t\) and \( au(4r(t - r - s) + s(t - s))/(t - 1)\). The overall mean \( A \) and variance \( V \) are obtained by summing over all the families.

As in Spielman and Ewens (9), we use statistic \( W \), the sum of \( b \), \( S \), and \( Y \). Under the null hypothesis, \( W \) has mean \( A_{com} \) and variance \( V_{com} \):

\[
A_{com} = (b + c)/2 + M_1/2 + A.
\]
\[
V_{com} = (b + c)/4 + V_1/4 + V.
\]

The test of significance can be obtained using

\[
z = (W - A_{com})/\sqrt{V_{com}},
\]

which has an approximate normal distribution under the null hypothesis.

RESULTS

In this section, we first validate the above theory and compare the powers of the S-TDT, \( T_1 \), and \( T_2 \) using simulations for different demographic histories. We then apply the 1-TDT and the combined test to a simulated data set from Genetic Analysis Workshop (GAW) 9 and a real data set on insulin-dependent diabetes mellitus (IDDM) from GAW 5.

The type I error rates and powers of the S-TDT and the 1-TDT tests

We studied the type I error rates of the S-TDT and the 1-TDT tests under the null hypothesis of no linkage between the marker locus and the disease for different demographic histories. We also compared the powers of the various tests under the alternative hypothesis that there is linkage between the marker locus and the disease. Three simple models were considered: population

-Am J Epidemiol Vol. 150, No. 1, 1999-
stratification, assortive mating, and differential gender mating preferences. For each model, we first generated the haplotypes at the marker locus and the disease locus for the parents according to the demographic history. The number of children in each family was assumed to have a specific distribution. Here we assumed that each family had three children. The haplotype of the offspring in each family was generated for a given recombination fraction, $\theta$. The phenotype of each offspring was determined by the genotype of the offspring at the disease locus. In our simulations, we assumed that the disease was recessive at the disease locus with alleles $d$ and $D$, where $d$ and $D$ were the normal and disease alleles, respectively. Only individuals with genotype DD at the disease locus were affected. In each nuclear family, we assumed the father was missing with probability 0.8. We sampled 100 families with at least one affected offspring and calculated the statistics for the S-TDT, $T_1$, and $T_2$. For $T_1$ and $T_2$, we also considered using only one affected offspring and all the affected offspring, respectively. We ran the simulation 10,000 times. The type I error was set to be 0.05. The power of each test statistic was the fraction of times that the absolute value of the test statistic is at least 1.96.

In the population stratification model, we assumed that the population under study consisted of two subpopulations. With probability 0.3 a family belonged to the first subpopulation and with probability 0.7 a family belonged to the second subpopulation. The frequencies of haplotypes $N_d$, $N_D$, $M_d$, and $M_D$ in the first population were set to 0.3, 0.2, 0.2, and 0.3, respectively. The frequencies of haplotypes $N_d$, $N_D$, $M_d$, and $M_D$ in the second population was set to 0.4, 0.1, 0.1, and 0.4, respectively. In the assortive mating model and differential mating preferences model, we assumed that the sample was from a population with haplotype frequencies for $N_d$, $N_D$, $M_d$, and $M_D$ to be 0.4, 0.1, 0.1, and 0.4, respectively. In the assortive mating model, we assumed that 80 percent of the families were formed by random mating and 20 percent of the families were formed through assortive mating with parents of the same phenotype. In the differential gender mating preference model, we assumed that unaffected males can mate randomly with females while affected males can only mate with affected females.

In table 4, models A, B, and C give the power estimates for different values of $\theta$, the recombination fraction between the marker locus and the disease locus. In models A and B, where males and females with the same genotype had the same mating preference, the simulated type I error rates for all the tests were close to 0.05 under the null hypothesis ($\theta = 0.5$). When $\theta < 0.5$, the S-TDT is more powerful than the 1-TDT tests proposed here. This conclusion is not general and depends on the distribution of the number of offspring in each family. If we restrict the tests only to families with both

### Table 4

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>S-TDT</th>
<th>$T_1$</th>
<th>$T_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>One affected</td>
<td>All affected</td>
<td>One affected</td>
</tr>
<tr>
<td>A. Population stratification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>4.92</td>
<td>4.94</td>
<td>4.48</td>
</tr>
<tr>
<td>0.4</td>
<td>32.6</td>
<td>12.8</td>
<td>11.1</td>
</tr>
<tr>
<td>0.3</td>
<td>84.2</td>
<td>37.5</td>
<td>31.6</td>
</tr>
<tr>
<td>0.2</td>
<td>90.9</td>
<td>69.1</td>
<td>59.2</td>
</tr>
<tr>
<td>0.1</td>
<td>100</td>
<td>91.5</td>
<td>83.1</td>
</tr>
<tr>
<td>0.0</td>
<td>100</td>
<td>98.5</td>
<td>94.1</td>
</tr>
<tr>
<td>B. Assortive mating</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>4.79</td>
<td>4.57</td>
<td>4.73</td>
</tr>
<tr>
<td>0.4</td>
<td>41.4</td>
<td>17.2</td>
<td>13.9</td>
</tr>
<tr>
<td>0.3</td>
<td>93.9</td>
<td>50.5</td>
<td>41.6</td>
</tr>
<tr>
<td>0.2</td>
<td>100</td>
<td>85.2</td>
<td>72.9</td>
</tr>
<tr>
<td>0.1</td>
<td>100</td>
<td>97.8</td>
<td>91.3</td>
</tr>
<tr>
<td>0.0</td>
<td>100</td>
<td>100</td>
<td>98.1</td>
</tr>
<tr>
<td>C. Differential gender mating preferences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>4.96</td>
<td>9.88</td>
<td>8.63</td>
</tr>
<tr>
<td>0.4</td>
<td>30.8</td>
<td>9.29</td>
<td>7.42</td>
</tr>
<tr>
<td>0.3</td>
<td>82.4</td>
<td>43.0</td>
<td>28.1</td>
</tr>
<tr>
<td>0.2</td>
<td>99.2</td>
<td>82.9</td>
<td>58.0</td>
</tr>
</tbody>
</table>
affected and unaffected sibs, the S-TDT is usually more powerful than the 1-TDT tests presented here. Thus, in combining different types of families, it might be more powerful to analyze families with both affected and unaffected sibs by the S-TDT instead of the 1-TDT even if one parent is available. For the 1-TDT tests presented here, \( T_i \) is generally more powerful than \( T_s \).

In model C, where males and females with the same genotype had different mating preference, \( T_i \) did not give the correct type I error rate under the null hypothesis \((\theta = 0.5)\). Thus, \( T_i \) cannot be used in this model.

### Simulated data from GAW 9

GAW 9, problem 1, was a simulation of an oligogenic disease involving four disease loci (13, 14). Two hundred families with at least one affected offspring were simulated. Each individual was typed at 360 marker loci. Two of the four disease loci coincided with marker loci D1G31 and D5G23. Strong associations (no recombination) were introduced between the disease and allele M8 of locus D1G31 and allele M7 of locus D5G23.

We used the same simulated data set to assess the 1-TDT by ignoring one of the parents. We considered four methods: a) father and first affected offspring; b) mother and first affected offspring; c) father and all the affected offspring; and d) mother and all the affected offspring. We first evaluated whether the 1-TDT gave the correct type I error rate. We performed the 1-TDT test on the 200 nuclear families and allele 1 of all the markers except markers D1G31 and D5G23 as these two markers were associated with the disease. We counted the number of markers, \( n \), where \( b_i + c_i \) was at least 10, and the number of markers, \( m \), whose absolute \( z \) score was greater than 1.96, the two-sided critical value for significance level 0.05. The simulated type I error rate was given by \( p = m/n \). The values of \( n, m, \) and type I error rate \( p = m/n \) are given in Table 5. Next we tested whether the 1-TDT can detect the association between the disease and allele M8 of D1G31 as well as allele M7 of D5G23. The values of \( b_i, c_i, V_i, \) and the corresponding \( z \) score are given in Table 6. The 1-TDT detected both associations.

### The insulin gene 5' VNTR

Next we applied the 1-TDT to a real data set on IDDM studied in GAW 5 (15, 16). We studied the VNTR locus within the 5' end of the insulin gene as in Spielman et al. (1) and Spielman and Ewens (9). As in Spielman et al. (1), we assigned restriction fragments to allele class 1 if they were smaller than 1kb and to allele class X if they were larger than 1kb.

We used similar methods as for the GAW 9 simulated data set, except that in methods a and b we included parents and the proband if the proband is given, or the parents and the oldest affected individual otherwise. The values of \( b_i, c_i, V_i, \) and the corresponding \( z \) score are given in Table 7.

### DISCUSSION

In this paper, we have proposed a new TDT-like test, the 1-TDT, to detect linkage and association between a candidate marker locus and a disease locus by using genotypes of case subjects and only one parent for the case subjects. The proposed test was applied to both simulated data sets and a real data set on IDDM for which the TDT and the S-TDT have shown linkage between the marker locus and the disease. The 1-TDT found the linkage, although the power of the 1-TDT is smaller than that of the TDT and the S-TDT.

We presented the 1-TDT and the combined test for biallelic markers. In some applications, the locus we

### Table 5. Values of \( n, m, \) and type I error rate \( p = m/n \) for the simulated data of GAW9

<table>
<thead>
<tr>
<th>Subjects included in analysis*</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>309</td>
<td>313</td>
<td>317</td>
<td>323</td>
</tr>
<tr>
<td>( m )</td>
<td>20</td>
<td>11</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>( 100p )</td>
<td>6.5</td>
<td>3.5</td>
<td>4.4</td>
<td>4.6</td>
</tr>
</tbody>
</table>

* a, father and first affected offspring; b, mother and first affected offspring; c, father and all affected offspring; d, mother and all affected offspring.

### Table 6. Values of \( b_i, c_i, V_i, \) and the corresponding \( z \) score for allele M8 of D1G31 and allele M7 of D5G32 for the simulated data set of GAW9

<table>
<thead>
<tr>
<th>Subjects included in analysis*</th>
<th>M8 of D1G31</th>
<th>M7 of D5G32</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>( b_i )</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>( c_i )</td>
<td>4</td>
<td>12</td>
</tr>
<tr>
<td>( V_i )</td>
<td>29</td>
<td>37</td>
</tr>
<tr>
<td>( z )</td>
<td>3.90</td>
<td>2.14</td>
</tr>
</tbody>
</table>

* a, father and first affected offspring; b, mother and first affected offspring; c, father and all affected offspring; d, mother and all affected offspring.
Transmission Disequilibrium Test with One Parent

There are interested in may have multiple alleles and we do not know which allele is the susceptible allele. This situation is the same as with the TDT and the S-TDT. We can use the same \( z_{\text{max}} \) procedure as suggested in Spielman and Ewens (9). In this procedure, a \( z \) score is calculated for each allele versus other alleles, and \( z_{\text{max}} \) is taken as the \( z \) score with the largest absolute value. Bonferroni correction for multiple testing or simulations should be used to declare statistical significance.

We have proposed a method to combine data from different types of families similar to the method of Spielman and Ewens (9). The test statistic proposed above is the summation of \( b \), \( S \), and \( Y \). It might be reasonable to take a weighted sum of \( b \), \( S \), and \( Y \), and then decide the best weight for each term to have the maximum power to detect the linkage if it exists. This will involve both the means and the variances of \( b \), \( S \), and \( Y \) under the alternative hypothesis. The mean and variance of \( Y \) are difficult to obtain because they depend on the probabilities that a heterozygous parent transmits certain alleles to both affected and unaffacted offspring. The mean and variance of \( b \) and \( S \) depend only on the probability that a heterozygous parent transmits one or the other to the affected offspring. Thus, we only consider the first two types of families here. Suppose we use the statistic \( W = b + \omega S \). Under Ott’s model (12), we can prove that the optimal weight \( \omega \) satisfies (see Appendix)

\[
\omega = 1 + \frac{\theta(1 - \theta)\delta^2}{p^2(s + t - (s - t)^2)}
\]

which depends on the allele frequency of allele N, the linkage disequilibrium parameter \( \delta \), and the recombination fraction \( \theta \) between the marker locus and the disease locus. All the parameters are unknown. When we test for tight linkage, that is, \( \theta = 0 \), it is reasonable to take the weight \( \omega = 1 \).

In summary, the 1-TDT complements the TDT and the S-TDT in that we can include families in which only one parent is available and all the sibs are affected in the combined test. These types of families cannot be analyzed by the TDT or the S-TDT.

**ACKNOWLEDGMENTS**

GAW is supported by NIH GM31575. This research was supported in part by a grant from the Emory University Research Council and a FIRST award from NIDDK R29DK53392 (F.S.).

We thank the referees for suggestions that improved the presentation of the paper. We also thank Dr. R. Fimmers for providing the IDDM data from GAW 5 and Dr. J. W. MacCluer for providing us with the simulated nuclear family data from GAW 9.

**REFERENCES**


**TABLE 7. Values of \( b \), \( c \), \( V \), and the corresponding \( z \) score for the insulin-dependent diabetes mellitus (IDDM) data set**

<table>
<thead>
<tr>
<th>Subjects included in analysis*</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b )</td>
<td>11</td>
<td>10</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>( c )</td>
<td>21</td>
<td>18</td>
<td>38</td>
<td>35</td>
</tr>
<tr>
<td>( V )</td>
<td>32</td>
<td>28</td>
<td>113</td>
<td>79</td>
</tr>
<tr>
<td>( z )</td>
<td>-1.77</td>
<td>-1.51</td>
<td>-1.32</td>
<td>-1.46</td>
</tr>
</tbody>
</table>

* a, father and proband or first affected offspring; b, mother and proband or first affected offspring; c, father and all affected offspring; d, mother and all affected offspring.

Am J Epidemiol Vol. 150, No. 1, 1999
Here we give the optimum weight, \( \omega \), to achieve the maximum power in detecting the linkage using the statistic \( W = b + \omega S_i \) under Ott's model (12).

Let \( \lambda \) be the fraction of case subjects with both parents available and \( 1 - \lambda \) be the fraction of case subjects with only one parent available. Let \( n \) be the total number of case subjects with at least one parent available. We assume that only one affected individual and one available parent are considered for each family. Then \( S_i = b_i \). We use the notation in table 2. Let \( b, c, b_i, \) and \( c_i \) be defined as in the test. Let \( s_i = (r + s)(s + u) \) and \( t_i = (r + t)(t + u) \). Then it can be shown that

\[
E(b - c + \omega(b_i - c_i)) = n(2\lambda + (1 - \lambda)\omega)(s - t),
\]

and

\[
\text{Var}(b - c + \omega(b_i - c_i)) = n(2\lambda(t + s) + (1 - \lambda)\omega^2(t_1 + s_1) - (2\lambda + (1 - \lambda)\omega^2)(t - s)^2).
\]

Using the two equations above, we can obtain the approximate power of the test using statistic \( W \) as \( 1 - \Phi(d_1\sqrt{n + d_2}) \), where \( \Phi \) is the standard normal cumulative distribution, \( d_1 \) and \( d_2 \) not depending on \( n \).

\[
d_1 = \frac{(2\lambda + (1 - \lambda)\omega)(s - t)}{\sqrt{2\lambda(s + t) + (1 - \lambda)\omega^2(t_1 + s_1) - (2\lambda + (1 - \lambda)\omega^2)(t - s)^2}}.
\]

When \( n \) is large, the dominant term inside \( \Phi \) is the first term. The larger the \( d_1 \), the higher the power of the test. Thus, the optimum weight \( \omega \) is the maximum point of \( d_1 \). Differentiating \( d_1 \) with respect to \( \omega \) and letting it equal to zero, we have

\[
\omega^{-1} = 1 + \frac{\theta(1 - \theta)d^2}{p^2(s + t - (s - t)^2)}.
\]