*Am. J. Hum. Genet. 62:181–188, 1998*

# A Test Statistic to Detect Errors in Sib-Pair Relationships

Margaret Gelder Ehm[1] and Michael Wagner[2]

[1]Bioinformatics Group and [2]Department of Human Genetics, Glaxo Wellcome, Inc., Research Triangle Park, NC

## Summary

**Several authors have proposed algorithms to detect Mendelian errors in human genetic linkage data. Most currently available methods use likelihood-based methods on multiplex family data to identify typing or pedigree errors. These algorithms cannot be applied in many sib-pair collections, because of lack of parental-genotype information. Nonetheless, misspecifying the relationships between individuals has serious consequences for sib-pair linkage studies: false relationships bias the statistics designed to identify linkage with disease phenotypes. To test the hypothesis that two individuals are sibs, we propose a test statistic based on the summation, over a large number of genetic markers, of the number of alleles shared identical by state by a pair of individuals, for each marker. The test statistic has an approximately normal distribution under the null hypothesis, and extreme negative values correspond to nonsib pairs. Power and significance studies show that the test statistic calculated by use of 50 unlinked markers has 96% power to detect half-sibs and has 100% power to detect unrelated individuals as not full-sib pairs, with a 5% false-positive rate. Furthermore, extreme positive values of the test statistic identify sibs as MZ twins.**

## Introduction

Errors can enter linkage data sets during every step of a genetic-mapping project: pedigree ascertainment, sample collection, sample processing, genotyping, and analysis. Errors may be identified by reviewing the output from analysis programs such as UNKNOWN (Ott 1991) or by manually inspecting the pedigree and genotype data for a limited set of markers—a tedious and error-prone process. Because of the complex and multifactorial nature of many of the diseases now being studied, population samples are larger than ever, exacerbating the problem of error detection. Therefore, specific, accurate, and automated methods of error detection are essential.

There are two types of genetic linkage–data errors to consider: pedigree error and typing error. Pedigree errors are systematic and affect all genotypes for an individual. They generally involve misidentification of individuals and relationships (Ott 1991). Examples include nonpaternity, unidentified adoption, and sample mix-ups. Sporadic errors or typing errors include all other types of errors, such as misreading of gels, data-entry error, and mutations (Buetow 1991). The present paper will concentrate on the detection of individuals causing pedigree errors.

Pedigree error is often detectable, since the incorrect relationship will likely show genotypes not conforming to Mendelian inheritance. Boehnke and Guo (1992), Ott (1993), and Stringham and Boehnke (1996) have developed tests for the identification of genotypes causing Mendelian inconsistencies. These tests use likelihood-based methods on multiplex family data to identify typing or pedigree errors based on genotypes that are likely to be incorrect. The algorithms cannot be applied in many sib-pair collections, because of the lack of parental genotype information.

The effects of errors in genetic data can be serious. Errors will often inflate distances in genetic maps and confound the ordering of polymorphic loci. They can reduce the power to locate disease genes and can bias the results of linkage-analysis statistics, which can lead to incorrect localizations for disease genes. The most common pedigree error is the assumption that sibs are full sibs when in fact they are half-sibs. This type of error decreases the observed number of shared alleles, compared with the number of shared alleles expected when the individuals are assumed to be full sibs. The resulting affected-sib-pair test statistic will be biased, making linkage more difficult to detect.

We propose the SibError algorithm, which uses only sibship data, to identify both the existence of systematic errors and the individual responsible for these errors. The method distinguishes between full-sib pairs and non–full-sib pairs such as half-sibs and unrelated individuals (Ehm and Wagner 1996).

Received February 21, 1997; accepted for publication October 23, 1997; electronically published January 9, 1998.

Address for correspondence and reprints: Dr. Margaret Gelder Ehm, Bioinformatics Group, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, NC 27709. E-mail: mge37216@glaxowellcome.com


0002-9297/98/6201-0026$02.00

**Table 1**

**Conditional Expectations, Mating Types, and Probabilities**

| Mating Type | $E(X_i \mid M_i = t_i)$ | $E(X_i^2 \mid M_i = t_i)$ | $P(M_i = t_i)$ |
|---|---|---|---|
| aa × aa | 1 | 1 | $\sum_m p_m^4$ |
| aa × bb | 1 | 1 | $(\sum_m p_m^2)^2 - \sum_m p_m^4$ |
| aa × ab | $\frac{3}{4}$ | $\frac{5}{8}$ | $\sum_m 4p_m^3(1 - p_m)$ |
| aa × bc | $\frac{3}{4}$ | $\frac{5}{8}$ | $\sum_m 2p_m^2[1 - \sum_n p_n^2 - 2p_m(1 - p_m)]$ |
| ab × ab | $\frac{5}{8}$ | $\frac{1}{2}$ | $\sum_{m<n} (2p_m p_n)^2$ |
| ab × ac | $\frac{9}{16}$ | $\frac{13}{32}$ | $\sum_{m<n} 4p_m p_n(p_m + p_n)(1 - p_m - p_n)$ |
| ab × cd | $\frac{1}{2}$ | $\frac{3}{8}$ | $\sum_{m<n} 2p_m p_n[1 - 2p_m p_n - 2(p_m + p_n)(1 - p_m - p_n) - \sum_k p_k^2]$ |

SOURCE.—Lange (1986).

## Methods

The proposed test statistic is based on the summation, over a large number of markers, of the identity-by-state allele sharing for a pair of sibs. If this sum is significantly less than the summation of the expected values of allele sharing calculated by use of the population allele frequencies, then the two individuals are presumed not to be full sibs.

Let $s_1$ and $s_2$ each represent individuals genotyped for $n$ markers. To test the hypothesis $H_0$ ($s_1$ and $s_2$ are full sibs) versus $H_1$ ($s_1$ and $s_2$ are not full sibs), let

$$X_i = \begin{cases} 1 & \text{if } s_1 \text{ and } s_2 \text{ share 2 alleles IBS at locus } i \\ \frac{1}{2} & \text{if } s_1 \text{ and } s_2 \text{ share 1 allele IBS at locus } i \\ 0 & \text{if } s_1 \text{ and } s_2 \text{ share 0 alleles IBS at locus } i \end{cases}.$$

Define $Y = \sum_{i=1}^{n} X_i$. Then

$$E(Y) = \sum_{i=1}^{n} E(X_i) \ , \tag{1}$$

and

$$\mathrm{Var}(Y) = \sum_{i=1}^{n} E(X_i^2) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} E(X_i X_j) - \left[\sum_{i=1}^{n} E(X_i)\right]^2 . \tag{2}$$

The marker concordances, the $X_i$s, are not identically distributed and not independent unless the markers are unlinked. The distribution of each $X_i$ depends on the allele-frequency distribution for the $i$th marker. To compute the expected marker concordances, $E(X_i)$, $E(X_i^2)$, and $E(X_i X_j)$, we condition on parental mating type (Lange 1986). Then

$$E(X_i) = \sum_{t_i} E(X_i \mid M_i = t_i) P(M_i = t_i) \ , \tag{3}$$

$$E(X_i^2) = \sum_{t_i} E(X_i^2 \mid M_i = t_i) P(M_i = t_i) \ , \tag{4}$$

and

$$E(X_i X_j) = \sum_{t_i}\sum_{t_j} E(X_i X_j \mid M_i = t_i, M_j = t_j) \, P(M_i = t_i, M_j = t_j) \tag{5}$$

are computed by summing over all possible mating types, $M_i$, for marker $i$ for one locus and over all pairs of mating types, $M_i$ and $M_j$, for two loci. We assume that the loci are in linkage equilibrium, so that $P(M_i = t_i, M_j = t_j) = P(M_i = t_i)P(M_j = t_j)$. Table 1 lists (1) the seven possible mating types for locus $i$, for the parents of a sib pair as specified by Lange (1986); (2) the conditional expectations of $X_i$ and $X_i^2$, given each mating type; and (3) the probabilities associated with each mating type, $P(M_i = t_i)$. Note that $p_m$ is the frequency for allele $m$ at a given locus. Lange (1986) shows that the conditional expectations and probabilities of each mating type can be easily verified. Table 2 lists, for all possible pairs of mating types for loci $i$ and $j$ for the parents of a sib pair, the conditional expectations of $X_i X_j$, given each mating type, for loci $i$ and $j$. $E(Y)$ is calculated by combining equations (1) and (3), and $\mathrm{Var}(Y)$ is calculated by combining equations (2), (3), (4), and (5).

Under $H_0$ and the assumption that the $X_i$s are independent, $Z = [Y - E(Y)]/\sqrt{\mathrm{Var}(Y)}$ has an approximate normal distribution. The null hypothesis should be rejected if the observed $z$ statistic is less than the $\alpha$th percentile of the standard normal distribution ($-z_\alpha$). Note that this test is one-sided, since increased sharing between $s_1$ and $s_2$ may be unlikely, but is not evidence for the alternative hypothesis. Also note that, when $s_1$ and $s_2$ are MZ twins, $Y = n$.

**Table 2**

**Conditional Expectations of Joint Mating Types, for Loci *i* and *j*: $E(X_iX_j \mid M_i = t_i, M_j = t_j)$**

| Mating Type for Locus *i* | Conditional Expectations for Loci *i* and *j*, Given Joint Mating Type | | | | | | |
|---|---|---|---|---|---|---|---|
| | aa × aa | aa × bb | aa × ab | aa × bc | ab × ab | ab × ac | ab × cd |
| aa × aa | 1 | | | | | | |
| aa × bb | 1 | 1 | | | | | |
| aa × ab | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{5}{8} - \frac{\theta}{4} + \frac{\theta^2}{4}$ | | | | |
| aa × bc | $\frac{3}{4}$ | $\frac{3}{4}$ | $\frac{5}{8} - \frac{\theta}{4} + \frac{\theta^2}{4}$ | $\frac{5}{8} - \frac{\theta}{4} + \frac{\theta^2}{4}$ | | | |
| ab × ab | $\frac{5}{8}$ | $\frac{5}{8}$ | $\frac{1}{2} - \frac{\theta}{8} + \frac{\theta^2}{8}$ | $\frac{1}{2} - \frac{\theta}{8} + \frac{\theta^2}{8}$ | $\frac{1}{2} - \frac{\theta}{2} + \frac{3\theta^2}{4} - \frac{\theta^3}{2} + \frac{\theta^4}{4}$ | | |
| ab × ac | $\frac{9}{16}$ | $\frac{9}{16}$ | $\frac{15}{32} - \frac{3\theta}{16} + \frac{3\theta^2}{16}$ | $\frac{15}{32} - \frac{3\theta}{16} + \frac{3\theta^2}{16}$ | $\frac{7}{16} - \frac{3\theta}{8} + \frac{\theta^2}{2} - \frac{\theta^3}{4} + \frac{\theta^4}{8}$ | $\frac{13}{32} - \frac{3\theta}{8} + \frac{7\theta^2}{16} - \frac{\theta^3}{8} + \frac{\theta^4}{16}$ | |
| ab × cd | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{7}{16} - \frac{\theta}{4} + \frac{\theta^2}{4}$ | $\frac{7}{16} - \frac{\theta}{4} + \frac{\theta^2}{4}$ | $\frac{3}{8} - \frac{\theta}{4} + \frac{\theta^2}{4}$ | $\frac{3}{8} - \frac{3\theta}{8} + \frac{3\theta^2}{8}$ | $\frac{1}{8} - \frac{\theta}{2} + \frac{\theta^2}{2}$ |

## Results

We now evaluate the significance and power of the test, given sets of 12–300 unlinked and linked markers, apply the method to a sample of Mexican Americans typed on a set of markers, and discuss the issue of multiple testing.

### *Significance and Power Studies*

The significance, or false-positive rate, summarizes how often the test statistic will falsely reject the null hypothesis that the individuals tested are indeed full sibs. The power measures how often the null hypothesis is rejected correctly when the individuals tested are not full sibs.

To estimate the significance, we generated 10,000 pedigrees each consisting of two parents and two sibs all typed on a set of markers with given recombination fractions between the markers (Ott 1989). The test statistic was calculated and the *P* value was determined by use of the standard normal distribution for each of the 10,000 tests. The proportion of *P* values less than the nominal $\alpha$ was used as an estimate of the false-positive rate, or observed $\hat{\alpha}$.

To estimate the power of the test, we generated 10,000 pedigrees under two models: (1) half-sibs and (2) unrelated individuals. Under the half-sib model, each pedigree consists of three parents and two children, in which the children share a mother and have different fathers. Under the unrelated-individual model, there are two independent nuclear families each with two parents and one child, and the two children are tested as though they were a sib pair. The test statistic was calculated and the *P* value was determined by use of the standard normal distribution for each of the 10,000 tests. The null hypothesis that the sib pair consists of two full sibs was rejected if the *P* value was less than the nominal $\alpha$. The frequency with which the null hypothesis was rejected for the 10,000 tests was an estimate of the power $(1 - \hat{\beta})$.

Table 3 lists the nominal $\alpha$, observed $\hat{\alpha}$, and power $(1 - \hat{\beta})$, both under the assumption that the pair are half-sibs and under the assumption that the pair are unrelated individuals, for a set of 50 unlinked ($\theta = .5$) markers. An allele range of 2–19 alleles, with a mean of 9 alleles and with varying frequencies, was assumed. The observed $\hat{\alpha}$ matches the nominal $\alpha$ very closely. The power is .962 for half-sibs and 1.00 for unrelated individuals, when $\alpha = .05$.

To illustrate the minimum number of markers needed for application of the test, we estimated the power necessary to detect half-sibs and unrelated individuals, for 50, 40, 30, and 20 unlinked markers. Table 4 lists nominal $\alpha$ and the power $(1 - \hat{\beta})$ to detect half-sibs and unrelated individuals. The observed $\hat{\alpha}$ is not listed for each sample of markers, but estimates correspond well to nominal $\alpha$. These power estimates indicate that the power to detect half-sibs is reduced when <50 markers are used. The power to detect unrelated individuals is reduced when <30 markers are used.

To test the method with data that are likely to be available for a given set of families, we investigated the power and significance rates for (1) a full-genome scan,

**Table 3**

**Power and Significance for 50 Unlinked Markers**

| | | $1 - \hat{\beta}$ | |
|---|---|---|---|
| Nominal $\alpha$ | Observed $\hat{\alpha}$ | Half-Sibs | Unrelated Individuals |
| .05 | .0547 | .962 | 1.00 |
| .01 | .0108 | .849 | 1.00 |
| .005 | .0057 | .788 | .999 |
| .001 | .0015 | .628 | .997 |

**Table 4**

**Power to Detect Half-Sibs and Unrelated Individuals, Given Sets of 20–50 Markers**

| | $1-\hat{\beta}$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Half-Sibs | | | | Unrelated Individuals | | | |
| NOMINAL $\alpha$ | 50 Markers | 40 Markers | 30 Markers | 20 Markers | 50 Markers | 40 Markers | 30 Markers | 20 Markers |
| .05 | .962 | .932 | .831 | .692 | 1.00 | 1.00 | .998 | .983 |
| .01 | .849 | .745 | .626 | .404 | 1.00 | .998 | .990 | .914 |
| .005 | .788 | .650 | .506 | .267 | .999 | .997 | .980 | .842 |
| .001 | .628 | .437 | .275 | .159 | .997 | .987 | .929 | .734 |

(2) a low-density full-genome scan, and (3) partial-genome scan. Note that each of these situations included linked markers and thus violated the assumption that the $X_i$s are independent.

To investigate the significance and power that are characteristic of a low-density genome scan, a set of 100 markers with variable allele frequencies placed ~35 cM apart ($\theta = .30$) on 22 autosomal chromosomes was used. Table 5 lists the nominal $\alpha$, observed $\hat{\alpha}$, and power ($1-\hat{\beta}$) to detect half-sibs and unrelated individuals. Observed $\hat{\alpha}$ is similar to nominal $\alpha$. The power to detect half-sibs and unrelated individuals is high: .995 and 1.00 for $\alpha = .05$.

Table 6 lists the nominal $\alpha$, the observed $\hat{\alpha}$, and the power ($1-\hat{\beta}$) to detect half-sibs and unrelated individuals, for markers characteristic of a full-genome scan. The set consists of 300 markers that had variable allele frequencies and that were placed ~10 cM apart ($\theta = .10$) on 22 autosomal chromosomes. Observed $\hat{\alpha}$ is slightly higher than nominal $\alpha$ (.0562 vs. .05), because of the dependence of the $X_i$s. The power to detect half-sibs and unrelated individuals as nonsib pairs is very high (1.00) when $\alpha = .05$.

Given the positive results discussed above for data generated during a low-density or full-genome scan, we investigated the significance and power by using both a set of 25 markers spaced 10 cM apart and a set of 12 markers spaced 21 cM apart ($\theta = .10$ and $\theta = .20$, respectively). These sets were characteristic of data generated for one chromosome. Table 7 lists the nominal $\alpha$, observed $\hat{\alpha}$, and power ($1-\hat{\beta}$) to detect half-sibs and unrelated individual pairs as non–full-sib pairs. Again the significance levels are slightly inflated (.075 vs. .05, for 25 markers), because of the dependence of the markers, but the power to detect unrelated pairs is high (.995 for 25 markers, when $\hat{\alpha} = .075$; and .814 for 12 markers, when $\hat{\alpha} = .0548$). The power to detect half-sib pairs is acceptable (.741) for 25 markers but is low (.351) for 12 markers.

The scenarios described above for linked markers assume that estimates of the map distances are error free–that is, that the marker distances used to generate the simulated sib-pair data are the same as those used to calculate the SibError test statistic. In practice, map-distance estimates are affected by pedigree and typing errors (Ehm et al. 1996). To illustrate the effect that error in map-distance estimates has on the significance and power of the SibError test statistic, a set of 50 markers with variable allele frequencies placed ~10 cM apart ($\theta = .10$) on two autosomal chromosomes was simulated. When values of $\theta$ equal to .05, .10 (no error), and .15 were used in calculating the SibError statistic, and when a nominal $\alpha = .05$ was assumed, the observed values of $\hat{\alpha}$ were .0753, .0971, and .0971, respectively; and the corresponding power to detect half-sibs, for $\theta$ equal to .05, .10, and .15, was .803, .856, and .856, respectively. Observed $\hat{\alpha}$ and power ($1-\hat{\beta}$) to detect half-sibs under the assumption that the map distances were underestimated were less than those under the assumption that there were no errors in the map distances. When the map distances were inflated, there was no effect on the significance and power. In practice, errors in map-distance estimates usually result in inflated map distances (Ehm et al. 1996), thus having a minimal effect on the characteristics of the SibError test statistic.

### *Example*

We applied the SibError algorithm to genotype data generated on individuals in a nuclear family collected as part of the American Diabetes Association GENNID Study. The family consists of two parents (individuals 425 and 426) and nine offspring. SibError results for four of the offspring (sibs 427–430) are listed in table 8. Like many of the families studied for complex dis-

**Table 5**

**Power and Significance for Low-Density Genome Scan**

| | | $1-\hat{\beta}$ | |
|---|---|---|---|
| NOMINAL $\alpha$ | OBSERVED $\hat{\alpha}$ | Half-Sibs | Unrelated Individuals |
| .05 | .0448 | .995 | 1.00 |
| .01 | .0084 | .958 | 1.00 |
| .005 | .0048 | .916 | 1.00 |
| .001 | .0014 | .753 | 1.00 |

**Table 6**

**Power and Significance for Full-Genome Scan**

| | | $1-\hat{\beta}$ | |
|---|---|---|---|
| NOMINAL $\alpha$ | OBSERVED $\hat{\alpha}$ | Half-Sibs | Unrelated Individuals |
| .05 | .0562 | 1.00 | 1.00 |
| .01 | .0132 | 1.00 | 1.00 |
| .005 | .0061 | .999 | 1.00 |
| .001 | .0014 | .999 | 1.00 |

eases, the parents are unavailable for genotyping. All four offspring have been typed for 50 unlinked markers spaced throughout the autosomal genome, although, because of laboratory error, not all typings are recorded. Several markers show more than four alleles for the sibship. Note that $n$ is the number of markers with typings for both sibs, $Y$ is the observed sharing, $E(Y)$ is the expected sharing, $\mathrm{Var}(Y)$ is the variance, and $[Y - E(Y)]/\sqrt{\mathrm{Var}(Y)}$ is the test statistic. Each test that involves sib 430 results in a significant test statistic (at the .05 level), indicating that sib 430 is not a full sib of sibs 427–429. Note that testing all sib pairs in a sibship constitutes multiple testing, which should be accounted for.

### *Multiple Testing*

As illustrated in the example, the methodology can lead to multiple tests for a sib in a large sibship. The hypothesis test is designed to determine if a pair of individuals are full sibs. Since many families consist of more than two sibs, fully testing a family involves the calculation of the test statistic for each possible sib pair. All $s(s-1)_2$ tests on all of the possible sib pairs are not independent but are conditionally independent, given that they involve a particular sib (Hodge 1984). In the previous example, there are four sibs. All tests involving sib 427 (i.e., sib pairs 427 and 428, 427 and 429, and 427 and 430) are independent. Multiple tests can result in an increased sibship false-positive rate. However, the multiple tests performed on each sib provide additional information that may increase the power to detect a problem sib. For example, in a sibship of three sibs, if one individual is not a full sib of the other two, then we should expect a significant result when that individual is tested against each of the other two sibs. However, if only one of the two test results is significant, we are in a quandary: is the significant result a false positive, or is the nonsignificant result a false negative? In a sibship of four sibs, three tests are performed on each individual. If all three are significant, we can be very confident that the individual is not a full sib of the other three. Even if only two of the three tests are significant, we can still be confident that the individual involved is not a full sib of the others, since it is more likely that we have one false-negative result rather than two false-positive results. Larger sibships allow more opportunities to detect and identify a problem sib. We discuss below the issues of significance and power in large sibships and suggest a rejection scheme that results in increased power and reduced false-positive rates in most sibships.

To accurately estimate the false-positive rate and power associated with testing all sib pairs within a family, we designed a simulation study. To estimate the sibship significance, we generated 10,000 pedigrees each consisting of two parents and 4–12 sibs all typed for 50 unlinked markers (Ott 1989). The test statistic was calculated for each sib pair within the sibship and the $P$ value was determined by use of the standard normal distribution. An individual was identified as not a full sib of the remaining members of the sibship if $s-2$ of the $s-1$ tests involving that individual had $P$ values less than the nominal value. (Requiring all $s-1$ tests to be significant resulted in low power for large sibships, because of the large number of opportunities for a false negative to occur.) Note that this scheme has power to detect errors in pedigrees in which there are at least $s-2$ opportunities for significant tests (with false positives being ignored) for each sib. The proportion of times that an individual was identified as erroneous is an estimate of the sibship false-positive rate, or observed $\hat{\alpha}$.

To estimate the sibshipwide power of the test, we generated 10,000 pedigrees under two models: (1) one half-sib and (2) two half-sibs. The one–half-sib model's pedigree includes 1 half-sib and 3–12 full sibs, in which the children all share a mother and in which the half-sib has

**Table 7**

**Power and Significance for Partial-Genome Scans**

| | 25 MARKERS AT 10 cM | | | 12 MARKERS AT 20 cM | | |
|---|---|---|---|---|---|---|
| | | $1-\hat{\beta}$ | | | $1-\hat{\beta}$ | |
| NOMINAL $\alpha$ | Observed $\hat{\alpha}$ | Half-Sibs | Unrelated Individuals | Observed $\hat{\alpha}$ | Half-Sibs | Unrelated Individuals |
| .05 | .075 | .741 | .995 | .0548 | .351 | .814 |
| .01 | .0193 | .532 | .976 | .009 | .0995 | .476 |
| .005 | .0104 | .425 | .953 | .009 | .0995 | .476 |
| .001 | .0027 | .231 | .842 | .0024 | .0413 | .304 |

**Table 8**

**SibError Results for Family 218, with Parents 425 and 426**

| Family | Sib Pair | $n$ | $Y$ | E($Y$) | Var($Y$) | $[Y - E(Y)]/\sqrt{\text{Var}(Y)}$ |
|---|---|---|---|---|---|---|
| 218 | 427 and 428 | 46 | 31.500 | 29.465 | 4.630 | .946 |
| 218 | 427 and 429 | 47 | 33.000 | 30.088 | 4.735 | 1.338 |
| 218 | 427 and 430 | 46 | 15.000 | 29.465 | 4.630 | −6.722* |
| 218 | 428 and 429 | 49 | 32.000 | 31.438 | 4.923 | .253 |
| 218 | 428 and 430 | 49 | 17.500 | 31.438 | 4.923 | −6.282* |
| 218 | 429 and 430 | 49 | 16.000 | 31.438 | 4.923 | −6.958* |

*$P = .05$.

a father different from that of the other sibs. The two–half-sibs model's pedigree includes 2 sibs with one father and 2–10 sibs with another father, with all children having the same mother. The sibships were tested as though the members were all sibs. All were typed for 50 unlinked markers. The test statistic was calculated and the $P$ value was determined by use of the standard normal distribution for each sib pair, in all of the 10,000 pedigrees. The null hypothesis that a sib pair consists of two sibs was rejected if the $P$ value was less than the nominal $\alpha$, and an individual was identified as erroneous if $s - 2$ of the tests involving that individual were significant. For the 10,000 tests, the frequency with which other full sibs were identified as not full sibs was an estimate of the power $(1 - \hat{\beta})$. The frequency with which the other full sibs were identified as erroneous is an estimate of the false-positive rate for this type of family. Although other, more complex family structures may occur, we chose these two simple models that encompass many of the problems observed in sibships encountered in real data.

Table 9 lists both the sibship false-positive rate under the assumption that the sibship consists entirely of full sibs and the sibship power and false-positive rate under the assumption that the sibship consists of one half-sib and $s - 1$ full sibs. The single test rates are $\alpha = .05$ and $(1 - \hat{\beta}) = .962$. The sibship false-positive rate (under the assumption that all sibs are full sibs) is less than the single test rate in all cases except when there are four sibs (.068). For the one–half-sib model, the sibship power is >.90 for all sibship sizes, and the false-positive rate is less than the single test rate for all sibship sizes except when there are four sibs (.103). Under the two-half-sibs model, the power to detect a non–full sib is somewhat less (.975 vs. .903, for 5 sibs; and .904 vs. .778, for 12 sibs), and the false-positive rates are similar (data not shown). When there are more than four sibs, the rejection scheme proposed leads to high power to detect half-sibs and to low false-positive rates.

## Discussion

SibError detects pedigree errors by using genotypes derived from sibs typed for $\geqslant 50$ markers. SibError can pinpoint the person who is not a full sib, in sibships larger than two, and can identify MZ twins (under the assumptions that there is no typing error, $Y$ should be equal to $n$). Significance studies show that the test statistic conforms to the normal distribution for unlinked markers but that, because of the violation of the assumption of the independence of the $X_i$s, $\hat{\alpha}$ is slightly inflated for linked markers. Power to detect half-sibs as not being full sibs is good, and power to detect unrelated individuals as not being full sibs is excellent. Fewer unlinked markers may be used, with some loss in power. The test statistic can be applied to a large number of linked markers such as a set collected as part of a low-density or full-genome scan, with outstanding results. It is not appropriate for smaller sets of tightly linked markers, since $\hat{\alpha}$ increases as $\theta$ decreases. The test statistic can be used for entire sibships by consideration of all possible tests. A rejection scheme that rejects full-sib status when all the tests or all but one of the tests involving that sib are significant provides a conservative test with high power.

In using SibError, it is important that all typings available for each individual are used without prior exclusion of those typings that contribute to Mendelian errors. Because SibError uses information from all markers typed for a sib pair, the removal of marker typings giving obvious Mendelian errors may mask true pedigree errors. Using unlinked markers gives the best power per marker and is preferred, if it is available. If linked mark-

**Table 9**

**Significance and Power to Identify Errors within Large Sibships**

| No. of Sibs | Full-Sib False-Positive Rate | One Half-Sib | |
|---|---|---|---|
| | | Power | False-Positive Rate |
| 4 | .068 | .987 | .103 |
| 5 | .021 | .975 | .018 |
| 6 | .0087 | .963 | .0043 |
| 7 | .0085 | .956 | .0011 |
| 8 | .0034 | .942 | .00051 |
| 9 | .0038 | .931 | .00024 |
| 10 | .0017 | .920 | .000089 |
| 11 | .0014 | .912 | .000010 |
| 12 | .0005 | .904 | .000027 |

ers must be used, then SibError requires map distances that can be either estimated from the data or obtained from public databases. On the basis of the results of the SibError program, the pedigree should be altered and analyzed again. When all pedigree changes have been made, then the data can be checked for typing errors. In order to maintain a false-positive rate $<.05$ and to have high power for most of the pedigrees and marker sets encountered, we recommend the use of a nominal $\alpha = .05$.

Göring and Ott (1995) have described an algorithm, Relative, to compute the likelihood of multilocus (linked and unlinked) genotype data observed for two individuals, given a stated relationship. The posterior probability of the relationship is calculated by use of Bayes's theorem, and the relationship with the highest posterior probability is the estimated relationship. A special case of this algorithm distinguishes full-sib pairs from half-sib and unrelated-individuals pairs. Results for power and significance studies assessing the properties of Relative similar to those of the studies described above are given in their paper. The studies assumed linked and unlinked markers each with an allele-frequency distribution of .32, .3, .2, .1, .05, .02, and .01. A total of 10,000 replicates were used to estimate the significance, and 1,000 were used to estimate the power. In our experience, the allele-frequency distribution has a minimal effect on power and significance estimates when SibError is used (data not shown). Therefore, we compare Göring and Ott's results to ours, which assume a varying allele distribution. Under the assumption that there are 50 unlinked markers and that the false positive rate is .0002, which is obtained by use of Göring and Ott's rejection scheme, the power to detect unrelated individuals by use of Relative is .999, and the power to detect half-sibs is .650. With SibError, the power is .993 for unrelated individuals and .437 for half-sibs. Under the assumption that there are 25 unlinked markers and that the false-positive rate is .0008, the power to detect unrelated individuals is .944 and .844 and the power to detect half-sibs is .267 and .197, for Relative and SibError, respectively. Finally, under the assumption that there are 50 markers in linkage groups of two markers with $\theta = .10$ within each group and with $\theta = .50$ between groups and that the false-positive rate is .0004, the power to detect unrelated individuals is .997 and .736 and the power to detect half-sibs is .977 and .274, for Relative and SibError, respectively.

Boehnke and Cox (1997) have described a likelihood-ratio method (RELPAIR) to infer the true relationship of a supposed sib pair. The method compares the multipoint probability of the marker data, conditional on different genetic relationships, and infers the relationship, given the data, that is most likely. Results of power and significance studies that assess the properties of RELPAIR and that are similar to those studies described above are given in their paper. The studies assumed linked markers each with an allele-frequency distribution of .25, .25, .25, and .25. A total of 10,000 replicates were used to estimate the significance and power. We compare Boehnke and Cox's results to ours, which assume a varying allele distribution. Under the assumption that there are 50 markers spaced 10 cM apart and that the false-positive rate is .086, which is obtained by use of Boehnke and Cox's rejection scheme, the power to detect unrelated individuals is 1.00 for both methods, and the power to detect half-sib pairs is .956 and .856, for RELPAIR and SibError, respectively. Under the assumption that there are 100 markers spaced 10 cM apart and that the false-positive rate is .0191, the power to detect unrelated individuals is 1.00 for both methods, and the power to detect half-sibs is .990 and .903, for RELPAIR and SibError, respectively. Under the assumption that there are 20 markers spaced 20 cM apart and that the false-positive rate is .162, the power to detect unrelated individuals is .995 and .992 and the power to detect half-sibs is .884 and .793, for RELPAIR and SibError, respectively. Under the assumption that there are 100 markers spaced 20 cM apart and that the false-positive rate is .0090, the power to detect unrelated individuals is 1.00 for both methods, and the power to detect half-sib pairs is .993 and .980, for RELPAIR and SibError, respectively. Note that the table given in Boehnke and Cox's paper comparing SibError and RELPAIR used data generated by an earlier version of the SibError algorithm, a version whose performance, when linked markers are used, is inferior to that of the algorithm described in the present paper.

Relative, RELPAIR, and SibError have several differences. The power to detect unrelated individuals and half-sibs is higher for Relative, given the excessively stringent $\alpha$ levels. The power for RELPAIR is slightly higher in some cases but is similar in others. Relative and RELPAIR allow one to distinguish between half-sib pairs and unrelated pairs, which SibError does not do. However, interpretation of the results from Relative and RELPAIR can be difficult, since no test statistic is defined. The SibError test statistic provides the flexibility to define a false-positive rate that is acceptable for a given situation. Simulation studies would be necessary to do the same with Relative and RELPAIR. Furthermore, the issue of multiple testing is not addressed by these latter two algorithms. For sibships having more than four sibs, the rejection scheme proposed with the SibError test statistic has power similar to or higher than those of Relative and RELPAIR. Relative does not use information on markers showing Mendelian errors for the individuals tested, whereas RELPAIR and SibError do. Therefore, to obtain information on 50 markers for use by Relative, more markers may need to be typed, to

ensure that there are 50 markers showing no Mendelian errors. Furthermore, the evidence for pedigree errors may be diluted by not utilizing markers with Mendelian errors. Computation of the posterior probabilities in Relative requires specification of prior probabilities, which are not generally known. Relative does not specifically identify MZ twins, whereas SibError and RELPAIR do.

Other published methods of error detection (Ott 1993; Stringham and Boehnke 1996) that have been developed to identify genotypes responsible for Mendelian errors consider one marker at a time. These methods were designed for large, extended pedigrees, do not integrate information from multiple markers, and were not intended to identify systematic errors.

Lathrop et al. (1983) proposed a model of pedigree error, obtained maximum-likelihood estimates of error parameters, and calculated posterior probabilities for the possible true relationships in each family, conditional on the putative relationships and marker data and using parameter estimates. The probabilities were then used to distinguish between pedigree and typing error, where Mendelian inconsistencies had been observed. Although the method appears to perform well despite having typings on only seven markers (see the example in the Lathrop et al. paper), it relies on typings from parents and focuses on inconsistencies, neither of which is necessary for the SibError test statistic to work. SibError addresses all of these issues by using a nonparametric approach that has been proved to work well.

### *Code Distribution*

The SibError test statistic has been implemented for two-generation pedigrees, in the C programming language. The input files required are the same as those required for the LINKAGE package. For specific instructions, please contact the first author (M.G.E.) by electronic mail (mge37216@glaxowellcome.com).

## Acknowledgments

The authors thank the Bioinformatics Group and the Departments of Human Genetics and Molecular Genetics at Glaxo Wellcome Inc., Drs. Hakan Sakul and Lon Cardon at Sequana Therapeutics, Dr. Bruce Weir and Dahlia Nielsen at North Carolina State University, and the reviewers, for their helpful comments and suggestions. The example pedigree used was collected as part of the American Diabetes Association's GENNID Study at the Harold Rifkin Family Acquisition Center at the University of Texas Health Science Center, San Antonio (principal investigator, Dr. Ralph A. DeFronzo). This study was supported by the American Diabetes Association. The genotypes used were generated by the Marshfield Medical Research Foundation (Dr. James L. Weber).

## References

Boehnke M, Cox NJ (1997) Accurate inference of relationships in sib-pair linkage studies. Am J Hum Genet 61:423–429

Boehnke M, Guo S-W (1992) Statistical approaches to identify marker typing error in linkage analysis. Am J Hum Genet Suppl 51:A183

Buetow KH (1991) Influence of aberrant observations on high-resolution linkage analysis outcomes. Am J Hum Genet 49: 985–994

Ehm MG, Kimmel M, Cottingham RW Jr (1996) Error detection for genetic data using likelihood methods. Am J Hum Genet 58:225–234

Ehm MG, Wagner M (1996) A test statistic for the affected sib-pair test method. Am J Hum Genet Suppl 59:A217

Göring HHH, Ott J (1995) Verification of sib relationship without knowledge of parental genotypes. Am J Hum Genet Suppl 57:A192

Hodge SE (1984) The information contained in multiple sibling pairs. Genet Epidemiol 1:109–122

Lange K (1986) A test statistic for the affected-sib-set method. Ann Hum Genet 50:283–290

Lathrop GM, Hooper AB, Huntsman JW, Ward RH (1983) Evaluating pedigree data. I. The estimation of pedigree error in the presence of marker mistyping. Am J Hum Genet 25: 241–262

Ott J (1989) Computer-simulation methods in human linkage analysis. Proc Natl Acad Sci USA 86:4175–4178

——— (1991) Analysis of human genetic linkage, rev ed. Johns Hopkins University, Baltimore

——— (1993) Detecting marker inconsistencies in human gene mapping. Hum Hered 43:25–30

Stringham HM, Boehnke M (1996) Identifying marker typing incompatibilities in linkage analysis. Am J Hum Genet 59: 946–950