Letters to the Editor

Am. J. Hum. Genet. 71:1235–1236, 2002

Using All Alleles in the Multiallelic Versions of the SDT and Combined SDT/TDT

To the Editor:

Horvath and Laird's sibling disequilibrium test (SDT) provides a nonparametric approach to testing genetic markers for both linkage and association with a disease (1998). The advantage over its parametric alternatives is its validity as a test of association when using sibships containing more than one affected sibling and/or more than one unaffected sibling. Horvath and Laird introduced an SDT for multiallelic markers and a biallelic combined SDT/transmission/disequilibrium test (TDT) when some parental genotypic information is available. Curtis et al. (1999) later developed a multiallelic combined SDT/TDT. The multiallelic versions of these tests are designed for situations in which there is no a priori knowledge of which allele at a marker might have an effect on disease status; otherwise, a biallelic test can be performed on the allele of interest versus all other alleles collapsed into one. A problem with the multiallelic extensions is that the statistic varies depending on which allele is omitted from the analysis. We present an alternative multiallelic SDT (mSDT) that takes into account all the allelic information and is consistent with the biallelic approach. This method can also be applied to the combined SDT/TDT.

In calculating the multiallelic versions of both the SDT and combined SDT/TDT, the statistics d^j, $j = 1, \ldots, m$ for a marker with m alleles are used. In the SDT, $d^j = \sum_i d_i^j$, where d_i^j represents the difference between the average number of times allele j occurs in an affected sibling and the average number of times it occurs in an unaffected sibling within sibship i (Horvath and Laird 1998); for the combined SDT/TDT, d^j is the difference between the number of times allele j is transmitted and the number of times it is not transmitted from a heterozygous parent to an affected child (Sham 1997). As discussed in Stuart (1955), a quadratic form of the d^j can be used to create a statistic with an asymptotic χ^2 distribution. It is noted that since $\sum_{j=1}^{m} d^j = 0$, the df for the distribution are $m - 1$. Furthermore, since using all m columns of the variance-covariance matrix creates a singularity, and, thus, the matrix is uninvertible, the natural solution is to eliminate one d^j and the corresponding row and column in the variance-covariance matrix to make it full rank. The invariance of the χ^2 statistic according to which variate (d^j) is omitted from the statistic is demonstrated by Stuart (1955).

To create a nonparametric test, $S_i^j = \text{sgn}(d_i^j)$ is used in place of d_i^j, where $\text{sgn}(d) = -1,0,1$ for $d <, =, > 0$, respectively. Though the sum of the quantities d_i^j, $j = 1, \ldots, m$, is 0 for each sibship $i = 1, \ldots, N$ and $S_i^1 = -S_i^2$ in the biallelic case, for more than two alleles, the sum over j of the S_i^j is not similarly linearly constrained within a sibship. In fact, the S_i^j can sum over j to either -1, 0, or 1. Despite this fact, multiallelic extensions to the SDT and combined SDT/TDT are formed by arbitrarily dropping one of the $S^j = \sum_{i=1}^{N} S_i^j$ from the analysis. The resulting $\chi^2_{(m-1)}$ test statistic is no longer invariant to which allele's information has been omitted, since there is no linear dependency among the values of S^j; information is being discarded unnecessarily. Furthermore, the variance-covariance matrix $\mathbf{W}$ for $\mathbf{S} = (S^1, \ldots, S^m)$ is nonsingular (exceptions are discussed below) before any of the m alleles are omitted. Thus, when all m alleles are used, a valid test statistic can still be created as $\mathbf{S}'\mathbf{W}^{-1}\mathbf{S}$, which has an asymptotic $\chi^2_{(m)}$ distribution (Hettmansperger 1984; Randles 1989).

There are, as mentioned, situations in which $\mathbf{W}$ will not be full rank. Among these are:

1. the biallelic case, in which the S^j are constrained, since there is a perfect negative correlation between S_i^1 and S_i^2 for all i ($\sum_{j=1,2} S_i^j = 0$ for all i);

2. the existence of at least one allele j, such that $S_i^j = 0$ for all N sibships, so that this allele will have a row and column of 0s in $\mathbf{W}$, creating a singularity; and

3. $\sum_{j=1}^{m} S_i^j = C$, the same constant, for all N sibships.

For these situations, we recommend the use of the Moore-Penrose generalized inverse (g-inverse) of the variance-covariance matrix $\mathbf{W}$, $\mathbf{W}^-$. This is a unique generalized inverse of $\mathbf{W}$ that satisfies the following conditions (Rao and Mitra 1971; Searle 1971): $\mathbf{W}\mathbf{W}^-$ and $\mathbf{W}^-\mathbf{W}$ are symmetric; $\mathbf{W}^-\mathbf{W}\mathbf{W}^- = \mathbf{W}^-$; and $\mathbf{W}\mathbf{W}^-\mathbf{W} = \mathbf{W}$. It is worth noting that the last two scenarios listed for a singular variance-covariance matrix are possible with the original SDT statistic, even after one allele has been omitted from the analysis, in which case the statistic cannot be calculated, since $\mathbf{W}$ is uninvertible.

When using $\mathbf{W}^-$ in place of $\mathbf{W}^{-1}$ in the quadratic form, the test statistic $\mathbf{S}'\mathbf{W}^-\mathbf{S}$ still has an asymptotic χ^2 distribution, now with df equal to the rank of $\mathbf{W}$ (Rao and Mitra 1971). Note that, for the biallelic case, in Horvath and Laird's notation (1998), the mSDT gives $\mathbf{S} = (b - c, c - b)$, and the $\mathbf{W}$ matrix will be of the form

$$\begin{bmatrix} b + c & -(b + c) \\ -(b + c) & b + c \end{bmatrix}.$$

The g-inverse is then calculated as

$$\begin{bmatrix} 1/(4b + 4c) & -1/(4b + 4c) \\ -1/(4b + 4c) & 1/(4b + 4c) \end{bmatrix},$$

which yields a χ^2 statistic of $(b - c)^2/(b + c)$ with 1 df, the same as the usual biallelic statistic.

To summarize our approach, we suggest modifying Horvath and Laird's SDT statistic (1998) and the combined SDT/TDT of Curtis et al. (1999) in the following manner to calculate the statistic for the mSDT:

1. Use all m alleles in the $\mathbf{S}$ vector and $\mathbf{W}$ matrix.
2. Use $\mathbf{W}^-$ in place of $\mathbf{W}^{-1}$ to create the χ^2 statistic (note that these are identical when $\mathbf{W}$ is full rank).
3. Use rank($\mathbf{W}$) as the df for the χ^2 distribution.

We give an example here, using simulated data from GAW9 (Hodge 1995). As in Spielman and Ewens (1998) and Knapp (1999), we focus on multiallelic markers D1G31 and D5G23, which contain actual disease alleles, M8 and M7, respectively. Table 1 shows the results of analyzing the data using the original Horvath and Laird SDT method, in which each allele is dropped in turn. Also shown are the results from analyzing the data using our mSDT approach. Note that each marker has eight alleles, so P values from the SDT are based on a χ^2_7 distribution, whereas the mSDT P values are from a χ^2_8 distribution, since the variance-covariance matrices for both markers are full rank. This example is not intended as any sort of power comparison but merely to illustrate that there is not necessarily a loss of power by introducing an additional df. The other thing to note from this table is the variation of the SDT P values depending on which allele is dropped. Although all test statistics are highly significant for marker D5G23, we can see quite a discrepancy between the SDT statistic for marker D1G31 when dropping allele M8 and any of the other seven SDT statistics. The mSDT approach will always give a unique χ^2 statistic, regardless of whether $\mathbf{W}$ is full rank. This method will be available in a future release of SAS/Genetics™.

WENDY CZIKA[1,2] AND JACK J. BERRY[1]
[1]*SAS Institute, Cary, NC; and* [2]*Department of Statistics, North Carolina State University, Raleigh*

Table 1

SDT and mSDT Statistics for Two Markers Linked and Associated with Disease

Allele Dropped	Statistic for Marker			
	D1G31		D5G23	
	χ^2	P	χ^2	P[a]
M1	23.115255	.001628	52.441075	.000048
M2	23.543802	.001370	52.365979	.000049
M3	23.239746	.001548	52.382481	.000049
M4	23.621073	.001328	51.086058	.000088
M5	23.661028	.001307	52.546616	.000046
M6	23.648748	.001313	53.238694	.000033
M7	23.417311	.001441	45.631132	.001031
M8	14.806102	.038567	51.811979	.000064
mSDT	23.667390	.002605	53.455015	.000088

[a] P values multiplied by 10^4.

References

Curtis D, Miller MB, Sham PC (1999) Combining the sibling disequilibrium test and transmission/disequilibrium test for multiallelic markers. Am J Hum Genet 64:1785–1786

Hettmansperger TP (1984) Statistical inference based on ranks. John Wiley and Sons, New York

Hodge SE (1995) An oligogenic disease displaying weak marker associations: a summary of contributions to problem 1 of GAW9. Genet Epidemiol 12:545–554

Horvath S and Laird NM (1998) A discordant-sibship test for disequilibrium and linkage: no need for parental data. Am J Hum Genet 63:1886–1897

Knapp M (1999) The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/disequilibrium test. Am J Hum Genet 64: 861–870

Randles RH (1989) A distribution-free multivariate sign test based on interdirections. J Am Stat Assoc 84:1045–1050

Rao CR and Mitra SK (1971) Generalized inverse of matrices and its applications. John Wiley and Sons, New York

Searle SR (1971) Linear Models. John Wiley and Sons, New York

Sham P (1997) Transmission/disequilibrium tests for multiallelic loci. Am J Hum Genet 61:774–778

Spielman RS and Ewens WJ (1998) A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 62:450–458

Stuart A (1955) A test of homogeneity of the marginal distributions in a two-way classification. Biometrika 42:412–416

Address for correspondence and reprints: Wendy Czika, R3212 SAS Campus Drive, Cary, NC 27513. E-mail: wendy.czika@sas.com

0002-9297/2002/7105-0025$15.00