

- 6 Sham PC, Zhao JH, Waldman I, Curtis D: Should ambiguous trios for the TDT be discarded? *Ann Hum Genet* 2000;64:575–576.
- 7 Aidoo M, Terlouw DJ, Kolczak MS, et al: Protective effects of the sickle cell gene against malaria morbidity and mortality. *Lancet* 2002;359(9314):1311–1312.
- 8 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;52(3):506–516.
- 9 Laird NM, Blacker D, Wilcox M: The sib transmission/disequilibrium test is a Mantel-Haenszel test [letter]. *Am J Hum Genet* 1998;52(3):506–516.

Knut M. Wittkowski
 The Rockefeller University, General Clinical Research Center
 1230 York Ave Box 322, New York, NY 10021 (USA)
 Tel./Fax +1 212 327 7175, E-Mail kmw@rockefeller.edu

Hum Hered 2004;58:61–62
 DOI: 10.1159/000081460

Reply to Wittkowski and Liu (Beyond the TDT: Rejoinder to Ewens and Spielman)

Warren J. Ewens^a, Richard S. Spielman^b

Departments of ^aBiology and ^bGenetics, The University of Pennsylvania, Philadelphia, Pa., USA

In their rejoinder [1] to our letter [2], Wittkowski and Liu ('WL' below) do not address the points on which we criticize their paper [3]. Instead, they introduce new matters (e.g., heterozygote advantage and selection of a decision function), and invoke papers by other authors, whom they quote out of context, in response to our criticisms. We see no reason to change our 'Comments,' but once more wish to clarify the main point at issue.

In order to do this we focus on the central claim of WL, who say (for example at end of the third paragraph of [1]): 'Thus, the TDT, in general is based on the multinomial distribution...' From this they claim that the calculations leading to the TDT statistic are not derived from the binomial distribution, and that as a consequence the denominator of the TDT is not correct. They conclude from this that the TDT is 'invalid in general' [3] and specifically 'for finite sample sizes' [1].

This is not a debate simply about terminology; the claims, if correct, would mean that the true Type I error (and P values) associated with the TDT are not the stated ones. However, the claims are erroneous, as we now demonstrate by reference to a familiar example and well-established statistical theory. We discussed the denominator of the TDT previously [2], and return to this aspect below. Here we begin with the more general claim by WL, quoted above, that the TDT is not 'based on' the binomial distribution.

This claim is strange, since in [2] and the original TDT paper [4] the variance term in the TDT statistic was explicitly derived from the binomial distribution. Here we illustrate the central role of the binomial distribution by a simple argument; we use the equivalence, not just analogy, between the genetic 'transmissions' tested by the TDT and the tossing of a coin. The null hypothesis of equal probability of transmission of either of two alleles by the heterozygous parent (TDT) is equivalent to the null hypothesis of equal probability of a head or a tail on a coin toss. Under the null hypothesis that marker and disease loci are unlinked, the transmissions of marker alleles,

even those from the two parents of a child, are independent, as are the results of different tosses of a coin. The coin toss provides the familiar paradigm for the properties of the binomial distribution, and the equivalence of the coin and the genetic situation then implies that the binomial distribution applies also for the TDT. We confirm this algebraically below.

The appropriate statistical analysis is therefore identical, whether the example comes from genetics or directly from the elementary statistics of coin-tossing. The statistical test for either can be carried out in various ways, of which two are the exact binomial test, and (more conveniently but approximately) by using a chi-square statistic. There is unanimous agreement about the use of the binomial distribution to describe the 'fair coin' experiment. Similarly, there is unanimous agreement about the variance of the binomial, and about the derivation of the correct denominator of the chi-square statistic. Since the genetic (TDT) case and statistical (coin tossing) cases are identical, this unanimity must extend to the TDT also.

WL justify their use of a statistic different from the TDT statistic by their claim in [3] that 'because the two alleles of a child are observed together, one does not have a sample of independently observed alleles.' However, the fact that the two alleles are observed together is irrelevant: what is important is that they are transmitted independently under the null hypothesis. Their argument is the analogue, in the coin case, of saying that the results of two coin tosses observed at the same time are not independent. The results of all tosses are independent, as are the transmissions of the alleles under the null hypothesis. Were their claim to be true, all current statistical theory concerning the binomial distribution would have to be abandoned.

For an algebraic confirmation of the points raised above, consider the case, which WL focus on, of m families, in which both parents are genotype PQ at the marker locus, and with one affected child in each family. There are then $2m (= n)$ transmissions of marker alleles to affected children, and under the null hypothesis of no linkage, the probability that P is transmitted is $p = 1/2$. The appropriate coin analogy is the test of whether a coin is fair, given that b heads arise from $2m = n$ tosses of the coin. Under this null hypothesis, the number of heads to appear has a binomial distribution with mean $np = 2m(1/2) = m$ and variance $npq = 2m(1/2)(1/2) = m/2$. The standard chi-square statistic testing the null hypothesis is then

$$(b - m)^2/[m/2]. \quad (1)$$

If we write $c = 2m - b$ as the number of tails observed, the statistic (1) becomes

$$(b - c)^2/(b + c). \quad (2)$$

Of course either (1) or (2) can be used (equivalently) to test the null hypothesis that the coin is indeed fair. In (2), the denominator term $b + c (= 2m = n)$ is the variance of the difference $b - c$ when the null hypothesis is true.

The analogy with the genetics case referred to above and discussed in [3] is the following. When the null hypothesis – that disease and marker loci are unlinked – is true, the $2m$ transmissions of marker alleles from the $2m$ parents are independent, as are the results of the coin tosses, and further, the probability that the allele P is transmitted to an affected child is $1/2$. The number b of transmissions of P thus has a binomial distribution identical to that of the number of heads on $2m$ tosses of a fair coin, and the statistic (2) can be used to test the null hypothesis that disease and marker loci are indeed unlinked. But (2) is the TDT statistic, and the above makes clear that

it is based completely on the binomial distribution, and not on multinomial calculations as claimed by WL.

The appropriateness of the TDT denominator term has been confirmed by various authors, including Sham et al. [5], whom Wittkowski and Liu quote, as though the reference supported their argument. The quote is, however, taken out of context. WL do not acknowledge that Sham et al. arrive at the same denominator as we do for the TDT, $2M$ in their terminology [5], $2m = b + c$ above. Thus Sham et al. [5] *confirm*, rather than *contradict*, our argument and variance calculation.

The remaining claims about the TDT made by WL [1] are all derived from their erroneous assertions about the relationship of the TDT statistic to the binomial distribution and its variance; we therefore do not address them here. The users of the TDT can be assured that it provides a valid test.

References

- 1 Wittkowski KM, Liu X: Beyond the TDT: Rejoinder to Ewens and Spielman. *Hum Hered* 2004;58:60–61.
- 2 Ewens WJ, Spielman RS: The TDT is a statistically valid test: Comments on Wittkowski and Liu. *Hum Hered* 2004;58:59–60.
- 3 Wittkowski KM, Liu X: A statistically valid alternative to the TDT. *Hum Hered* 2002;54:157–164.
- 4 Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: The insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 1993;52:506–516.
- 5 Sham PC, Zhao JH, Waldman I, Curtis D: Should ambiguous trios for the TDT be discarded? *Ann Hum Genet* 2000;64:575–576.

Prof. Warren J. Ewens, Department of Biology
University of Pennsylvania, Philadelphia, PA 19104-6018 (USA)
Tel. +1 215 898 7109, Fax +1 215 898 8780
E-Mail wewens@as.upenn.edu