is a standard and convenient chi-square approximation to a binomial test, as we stated explicitly in [2]. Of course [2], the binomial distribution provides an exact test, if desired.

References

- 1 Wittkowski KM, Liu X: A statistically valid alternative to the TDT. *Hum Hered* 2002;54:157–164.
- 2 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.
- 3 Sham PC, Zhao JH, Waldman I, Curtis, D: Should ambiguous trios for the TDT be discarded? *Ann Hum Genet* 2000;64:575–576.
- 4 Sokal RR, Rohlf FJ: *Biometry*. San Francisco, WH Freeman, 1969.

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@sas.upenn.edu

Hum Hered 2004;58:60–61
DOI: 10.1159/000081459

Beyond the TDT: Rejoinder to Ewens and Spielman

Knut M. Wittkowski, Xin Liu

The Rockefeller University, General Clinical Research Center,
New York, N.Y., USA

In their response [1] to our 2002 paper [2], EWENS and SPIELMAN (referred to as ES below, in keeping with their diction) point out a typographical error. The McN numerator, of course, is the difference of n_{PP} and n_{QQ} , rather than the sum. When referring to the convergence of the TDT's distributions to the limiting χ^2 distribution under the null hypothesis (H_0), also on p. 159, we now see that the term 'large sample approximation' would have been more appropriate than the term 'estimate'. Finally, as two other readers have pointed out to us, the numerator of d contains p_{QQ} , rather than p_{PP} , and in table 1 d should have read $r_d = 1 + (r - 1)d$. We apologize for these oversights and appreciate the opportunity for a clarification. Regarding the other points raised by ES, however, we stand firm behind our results.

While ES contend that 'claims about validity of [a] test refer – by definition – to statistical properties [...] under the null hypothesis' only, we follow Pearson in that 'a review of the consequences of alternative decisions must be introduced' [3] to ensure validity and, thus, consider tests to be invalid, when the assumptions on which they rely are violated [4]. Accordingly, in acknowledging the contributions of Neyman and Pearson to the 'testing theory' ES claim to have used in developing the TDT, Lehmann [5] defines the goal toward hypothesis testing strives as 'the selection of a decision function δ which minimizes the resulting risk' $R(\theta, \delta)$ for all hypothetical parameters θ .

To support their assertions that the TDT 'depends entirely on a binomial distribution [...] and not, as claimed by WL, on a multinomial distribution' and that 'the multinomial distribution arises as an artifact of the WL approach, and results from their classification

of families into three groups', ES refer to a paper, which explicitly states that 'the random variables n_{PP} , n_{PQ} , n_{QQ} [our notation] have a multinomial distribution' under H_0 [6]. Even when $n_{PP} = n_{QQ}$, the TDT's distribution may not reduce to the binomial, e.g., when heterozygosity confers an advantage

$$n_{PQ} > 2 \sqrt{n_{PP} n_{QQ}},$$

such as in sickle cell anemia against malaria [7]. Thus, the TDT, in general, is 'based on the multinomial distribution', while the SMN's distribution is always based on the binomial distribution.

To ensure that appropriate assumptions are made when comparing important features of the tests, we modeled the inheritance process directly. We assumed random mixing and, as do ES, independently transmitted alleles. We then allowed the risk of developing the phenotype to depend on filial genotype, relative risk, and degree of dominance. For the case of children born to $PQ \sim PQ$ parents, figure 1 compares the distributions of the TDT (McNemar test counting alleles calculated from the observed children) and the McN (McNemar test counting children) under H_0 . As we have demonstrated, the TDT converges to the McN when the sample size increases, but the TDT's exact distribution is more discrete than the McN's. As a consequence, the TDT is often liberal (anti-conservative) when compared to the common asymptotic χ^2 distribution and the TDT's exact power is more likely to 'jump' erratically than the SMN's (fig. 3).

The paper [6] quoted by ES then arrives at the TDT by switching focus from families to alleles, as did ES earlier [8]: 'The difference in transmission between alleles A and B is $(2n_{PP} - 2n_{QQ})$.' Because the environmental and genetic factors determining which child develops the phenotype are identical for the two alleles transmitted to the same child, switching from 'a sample of affected children' [8] to $2n$ alleles yields observations that are not independently distributed. While this dependency decreases with sample size, having to rely on a large sample approximation is the reason for the TDT's undesirable properties for samples of size $n \leq 500$ (fig. 3). As in the case referred to by ES in their reply to Laird et al. [9], 'the difference is substantial when the data consists of only a small number of families.' Moreover, as we have demonstrated, this problem cannot be resolved by merely using the binomial instead of the χ^2 distribution.

Finally, ES claim that, in contrast to the SMN, the denominator of the TDT is 'known before the experiment'. When screening the genome, the number of children with at least one heterozygous parent varies by locus. Even if the number of such children were known, the number of identifiable transmitted alleles (two if both parents are heterozygous, one otherwise) is not 'known before the experiment'.

In conclusion, we appreciate the additional insight ES provided into the heuristics that led them to propose an algorithm that has 'precisely the form of a McNemar statistic' and even its limiting χ^2 distribution under H_0 , while being invalid for finite sample sizes in the situations for which it was proposed.

References

- 1 Ewens WJ, Spielman RS: The TDT is a statistically valid test: Comments on Wittkowski and Liu. *Human Hered* 2004;58:59–60.
- 2 Wittkowski KM, Liu X: A statistically valid alternative to the TDT. *Human Hered* 2002;54(3):157–164.
- 3 Anscombe FJ: The validity of comparative experiments. *J R Stat Soc (Ser A)* 1948;111(3):181–211.
- 4 Cronbach LJ, Meehl PE: Construct validity in psychological tests. *Psychol Bull* 1955;52(4):281–302.
- 5 Lehmann EL: *Testing statistical hypothesis*. New York, Wiley, 1959.

- 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