Letters to the Editor

Am. J. Hum. Genet. 64:1485, 1999

Genomewide Transmission/Disequilibrium Testing: A Correction

To the Editor:

This response is to address comments made by several investigators regarding the sample sizes required for genomewide transmission/disequilibrium testing (TDT) in my earlier article (Camp 1997). There are two main comments: first, the issue of independence of parental transmissions and, second, the issue of the definition for the random variable (RV) B_i (for detailed explanations, see Risch and Merikangas 1996; Camp 1997).

Fengzhu Sun and Rong Cheng have kindly pointed out that the assumption that I made (Camp 1997) about the independence of parental transmissions is incorrect. This assumption of independence is valid only under the multiplicative mode of inheritance (MOI) (Knapp et al. 1993). Sun and Cheng have shown that, for the number of samples that are required in order to perform a genomewide TDT using singletons, the correct version of the formula given in Camp (1997) is

$$N \geq \frac{\left[z_{1-\alpha}\sqrt{2pq} - z_{\beta}\sqrt{S - 2h_s^2(2\tau_s - 1)^2}\right]^2}{2h_s^2(2\tau_s - 1)^2},$$

where $S = pq[(1 - q^2)f_2 + (1 - 2pq)f_1 + (1 - p^2)f_0]/K$, p is the frequency of the putative disease allele ($q = 1 - p$), h_s is the probability that a parent will be affected, given that she or he has a single affected offspring (SAO), τ_s is the conditional probability that the disease allele is transmitted, given that the parent is heterozygous and has an SAO, and K is the population prevalence of the disease. Correct formulas for h_s, τ_s, and K can be found in Camp (1997).

Both I and Sun and Cheng have recalculated, using the revised formula given above, and have found that the new sample sizes are extremely similar to those shown in table 3 in Camp (1997). The average discrepancy, although not always in the same direction, was ~0.65%. I agree with Sun and Cheng that it is important that the correct formula be available in the literature (hence it is included it here); however, I think that it is equally important and interesting to note that the incorrect assumption of independence of parental transmissions that I made (Camp 1997) has little impact on the sample size calculated (see table 1). A similar conclusion is also true in the case of genomewide TDT with affected sib pairs (not shown).

A second issue—and one with greater impact on sample size—regards the choice of definition for the RV B_i. Recall that in Risch and Merikangas (1996) the RV B_i takes the values $+(h_s)^{-\frac{1}{2}}$, $-(h_s)^{-\frac{1}{2}}$, and 0 for the cases in which the parent is heterozygous and transmits the disease allele, is heterozygous and transmits the normal allele, or is homozygous, respectively; Camp (1997) alternatively used $+(2pq)^{-\frac{1}{2}}$, $-(2pq)^{-\frac{1}{2}}$, and 0. These two definitions for the RV B_i result in two different TDT statistics (in which the subscripts "R&M" and "C" denote "Risch and Merikangas" and "Camp," respectively): $T_{\text{R\&M}} = \{(b - c)^2/[h_s(2N)]\}$ and $T_{\text{C}} = \{(b - c)^2/[2pq(2N)]\}$, where b and c are the standard symbols used, in the TDT statistic, for the number of times that the disease and the normal alleles, respectively, are transmitted from heterozygous parents and where N is the total number of trios collected for study.

The denominator in the true TDT statistic is $(b + c)$—that is, the total number of heterozygous parents (M, say) within the $2N$ total possible parents in the N trios collected. Thus, the two different RVs for B_i are effectively using two different estimates for M: $h_s(2N)$ and $2pq(2N)$. Now, it is true that $E(M) = h_2(2N)$, as used by Risch and Merikangas. However, the calculation of h_s includes information on the genotypic relative risks (GRRs), or γ, which are unknown. Hence the RV specified by Risch and Merikangas (1996) could never be used to actually perform the test, since the values assigned to the RV B_i assume knowledge of the values for the GRRs. This was the rationale for changing the RV B_i in Camp (1997) to one that was not dependent on unknown parameters. Under the null hypothesis ($\gamma = 1$), $h_s = 2pq$; hence, the false-positive rates under both definitions will be as expected under the true TDT statistic. For power, however, the definition as given by Risch and Merikangas (1996) estimates power more accurately (M. M. Iles, personal communication). It is also

Table 1

Sample Sizes Necessary to Gain 80% Power in a Genomewide TDT, for the C and R&M Definitions of RV B_i and with Consideration of Dependence between Parental Transmissions

	Sample Size Required[a]											
	Multiplicative			Additive			Recessive			Dominant		
		R&M			R&M			R&M			R&M	
γ and p	C; No	No	Yes	C; No	No	Yes	C; No	No	Yes	C; No	No	Yes
4.0:												
.01	523	1,097	1,097	549	1,137	1,135	4.3×10^6	4.3×10^6	4.3×10^6	562	1,158	1,155
.10	86	150	150	123	200	197	5,056	5,559	5,643	153	239	234
.50	103	103	103	222	222	220	205	205	208	712	712	698
.80	291	222	222	663	560	558	337	263	264	9,873	9,476	9,388
2.0:												
.01	4,154	5,817	5,817	4,154	5,817	5,817	3.8×10^7	3.9×10^7	3.9×10^7	4,317	6,014	6,008
.10	533	695	695	533	695	695	43,331	44,808	45,065	766	960	954
.50	340	340	340	340	340	340	949	949	959	1,861	1,861	1,840
.80	750	640	640	750	640	640	976	851	855	22,728	22,126	21,995
1.5:												
.01	16,008	19,300	19,300	15,550	18,794	18,801	1.5×10^8	1.5×10^8	1.5×10^8	16,487	19,829	19,821
.10	1,908	2,216	2,216	1,485	1,757	1,762	1.7×10^5	1.7×10^5	1.7×10^5	2,554	2,911	2,901
.50	949	949	949	464	464	466	3,078	3,078	3,099	4,599	4,599	4,568
.80	1,833	1,662	1,662	816	701	703	2,553	2,351	2,359	51,914	51,003	50808

[a] "No" denotes that dependence of parental transmissions were not accounted for; "Yes" denotes that dependence of parental transmissions were accounted for.

worth noting that, when $p = q = .5$, it is also the case that $h_s = 2pq$, and so discrepancies between the two methods are small when p and q are near equifrequent or when γ is near 1.0, and discrepancies are larger for those cases in which p and q are more divergent and when γ is large. I believe that the RV B_i as used by Risch and Merikangas (1996) leads to the correct sample sizes for a given power but that their B_i is inappropriately parameterized for use in a test statistic. The RV given by Camp (1997) was an attempt to gain both a tool for power and a valid RV for use in a TDT test statistic. Table 1 illustrates the sample sizes necessary to gain 80% power in a genomewide TDT using the two different types of RV B_i defined above and for various values for γ and p. For each MOI, the third column illustrates (using B_i as defined by Risch and Merikangas [1996]) the sample sizes when the dependence of parental transmissions are considered, as discussed above. The formula for this column is as follows:

$$N \geqslant \frac{\left[z_{1-\alpha} - z_{\beta}\sqrt{S' - 2h_s(2\tau_s - 1)^2}\right]^2}{2h_s(2\tau_s - 1)^2},$$

where $S' = S/h_2$.

Qualitatively, the results from all the alternatives discussed above are concordant—that is, they all indicate that genomewide TDT analysis could be useful as an alternative to classical affected-sib-pair linkage analysis for localization of genes of small effect in complex disease. Quantitatively, with respect to the RV B_i, results differ for large γ and extreme values for p. It is now left to the reader to determine whether either of these issues is worthy of further investigation.

Nicola J. Camp
Genetic Research, Intermountain Health Care, and Genetic Epidemiology, Department of Medical Informatics, University of Utah, Salt Lake City

References

Camp NJ (1997) Genomewide transmission/disequilibrium testing—consideration of the genotypic relative risks at disease loci. Am J Hum Genet 61:1424–1430

Knapp M, Seuchter SA, Baur MP (1993) The haplotype-relative-risk (HRR) method for analysis of association in nuclear families. Am J Hum Genet 52:1085–1093

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Address for correspondence and reprints: Dr. Nicola J. Camp, Genetic Research, 391 Chipeta Way, Suite C, Salt Lake City, UT 84108. E-mail: nicki@gr.ihc.com

0002-9297/99/6405-0035$02.00