

## Letters to the Editor

---

Am. J. Hum. Genet. 63:1563, 1998

### Estimation of Pairwise Relationships in the Presence of Genotyping Errors

To the Editor:

Boehnke and Cox (1997) described a likelihood method for the inference of pairwise relationships from data for a genome scan. This tool is useful for verification of the relationships in a linkage study, to identify pedigree errors, which, if undetected, could reduce the power to detect genes. While they showed that the method is robust for genotyping errors when the relationships full sibs, half sibs, and unrelated are inferred, a single genotyping error will lead to zero likelihood for the relationships MZ twins and parent/offspring. A simple modification of the method eliminates this problem and thus allows the accurate inference of the relationships MZ twins and parent/offspring, even in the presence of genotyping errors.

The crucial calculation in the method of Boehnke and Cox (1997), with regard to genotyping errors, is  $P(X_k | I_k = i)$ , the probability that two individuals have genotypes  $X_k = (X_{k1}, X_{k2})$  at locus  $k$ , given that they share  $i$  alleles identical by descent at the locus. To simplify the notation, let  $p_i(x) = P(X_k = x | I_k = i)$ . When two individuals share no alleles identical by state at a locus,  $p_1(x) = p_2(x) = 0$ , which results in a zero likelihood for relationships such as MZ twins and parent/offspring. If we allow  $p_i(x) > 0$  for all possible  $x$ , this problem is eliminated.

In the calculation of the likelihood for a putative relationship, we propose to replace the values for  $p_i(x)$  used by Boehnke and Cox (1997) with the following:

$$p_0^*(x) = p_0(x) ,$$

$$p_1^*(x) = (1 - \epsilon)p_1(x) + \epsilon p_0(x) ,$$

$$p_2^*(x) = (1 - \epsilon)p_2(x) + \epsilon p_0(x) ,$$

where  $\epsilon$  denotes twice the approximate genotyping error rate.

To test this idea, we performed a computer simulation. By use of the 366 autosomal markers in Weber screening

set version 9 (Yuan et al. 1997), 10,000 relative pairs were simulated for each of five relationships: MZ twins, parent/offspring, full sibs, half sibs, and unrelated. The sex-averaged map locations were taken from the study by Broman et al. (1998); allele frequencies were estimated by use of eight of the CEPH families. The intermarker spacings were  $9.4 \pm 3.6$  cM; the marker heterozygosities were  $.78 \pm .06$ . Genotypes were simulated by use of an error rate of 2%. For each relative pair, the likelihood for the five relationships were calculated by use of  $\epsilon = .02, .04, \text{ and } .08$ .

For the values of  $\epsilon$  considered, all five relationships were classified correctly in all 10,000 replicates. Thus, our proposal is successful in extending Boehnke and Cox's (1997) method to the inference of MZ twins and parent/offspring pairs, in the presence of genotyping errors.

KARL W. BROMAN AND JAMES L. WEBER  
Marshfield Medical Research Foundation  
Marshfield, WI

### References

- Boehnke M, Cox NJ (1997) Accurate inference of relationships in sib-pair linkage studies. *Am J Hum Genet* 61:423–429
- Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am J Hum Genet* 63:861–869
- Yuan B, Vaske D, Weber JL, Beck J, Sheffield VC (1997) Improved set of short-tandem-repeat polymorphisms for screening the human genome. *Am J Hum Genet* 60:459–460

Address for correspondence and reprints: Karl W. Broman, Center for Medical Genetics, Marshfield Medical Research Foundation, 1000 North Oak Avenue, Marshfield, WI 54449. E-mail: BromanK@cmg.mfldclin.edu

© 1998 by The American Society of Human Genetics. All rights reserved.  
0002-9297/98/6305-0041\$02.00