European Journal of Human Genetics (2003) 11, 637

www.nature.com/ejhg

LETTER

Comment on ''The Impact of genotyping error on haplotype reconstruction and frequency estimation''

European Journal of Human Genetics (2003) **11,** 637. doi:10.1038/sj.ejhg.5201030

In a recent article in the European Journal of Human Genetics, Kirk and Cardon[1] examined the relative accuracy of haplotype estimation using families *vs* unrelateds in the presence of genotyping error. The authors come to the conclusion that parent–child trios 'offer little or no gain over unrelated samples' for the purpose of haplotype frequency estimation and that the 'accuracy of haplotype inference in trios is generally closer to that in unrelateds than to larger families'. These statements, however, stand in sharp contrast to recently published work,[2,3] which found that frequency estimates gained from trios are in general very close to estimates gained from directly observed haplotypes. Moreover, when the haplotype diversity is not too low trios are more efficient than unrelated individuals even on a per genotype base, that is, *N* genotyped unrelated individuals are less efficient than *N*/3 genotyped parent–child trios. In addition, Schaid[3] investigated nuclear families with more than one child and found that further children do not yield much improvement in the precision of haplotype frequency estimates.

We believe that the discrepancy between Schaid's and our results on the one side and the results of Kirk and Cardon on the other side cannot be explained by the fact that Kirk and Cardon employed a different measure to assess the accuracy of haplotype estimates or that we did not include the possibility of typing errors. In fact, Kirk and Cardon miss to detect the benefit of child information also under simulations with an error rate of zero. We rather think that the discrepancy is due to their usage of an inappropriate methodology. While they use the EM-algorithm to estimate haplotype frequencies from individuals, for no apparent reason they do not use the EM-algorithm for haplotype frequency estimation for nuclear families, as described for instance by Rohde and Fuerst.[4] Instead they use Lander-Green (which is implemented in programs such as GENEHUNTER and MERLIN) to reconstruct phase in their nuclear families. However, GENEHUNTER and MERLIN perform well for that purpose only when there are few haplotype ambiguities, for instance in the case of multiallelic markers and large pedigrees. Since Lander-Green ignores the linkage disequilibrium between the markers, all possible haplotype explanations are equally likely in case of ambiguity. In particular, for tightly linked SNPs and small families, the inferred haplotypes by GENEHUNTER even depend on the order of alleles in the input file.[5]

We agree with Kirk and Cardon that it would be important to incorporate the impact of genotyping errors into Schaid's and our considerations. Intuitively, one would expect that the difference in efficiency between individuals and trios would increase in the presence of erroneous genotypes. On the other side, it is known that on average only 30% of all genotyping errors can be detected in trios.[6] So it may be possible that the effect of error detection on the accuracy of haplotype frequency estimates in trios is negligible or may even be negative. The occurrence of the latter effect will certainly depend on the strategy false genotypes are dealt with. We fear that due to the inappropriate method used to estimate haplotype frequencies from pedigrees, the work by Kirk and Cardon does not provide a reliable answer to these questions.

Tim Becker and Michael Knapp
Institute for Medical Biometry, Informatics and Epidemiology
University of Bonn, Bonn, Germany

Correspondence to: Tim Becker, Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn
Sigmund-Freud-Str. 25, D-53105 Bonn, Germany
Tel: +49 228 287 5564; Fax: +49 228 287 5032;
E-mail: becker@imsdd.meb.uni-bonn.de

References

1 Kirk KM, Cardon LR: The impact of genotyping error on haplotype reconstruction and frequency estimation. *Eur J Hum Genet* 2002; **10**: 616–622.
2 Becker T, Knapp M: Efficiency of haplotype frequency estimation when nuclear family information is included. *Hum Hered* 2002; **54**: 45–53.
3 Schaid DJ: Relative efficiency of ambiguous *vs* directly measured haplotype frequencies. *Genet Epidemiol* 2002; **23**: 426–443.
4 Rohde K, Fuerst R: Haplotyping and estimation of haplotype frequencies for closely linked biallelic multilocus genetic phenotypes including nuclear family information. *Hum Mut* 2001; **17**: 289–295.
5 Schaid DJ, McDonnell SK, Wang L, Cunningham JM, Thibodeau SN: Caution on pedigree haplotype inference with software that assumes linkage equilibrium. *Am J Hum Genet* 2002; **71**: 992–995.
6 Gordon D, Heath SC, Ott J: True pedigree errors more frequent than apparent errors for single nucleotide polymorphisms. *Hum Hered* 1999; **49**: 65–70.