It is straightforward to implement these formulas in a computer program that makes it possible to calculate the distribution of $Z_n$ for a broader range of values for the number $a$ of alleles. For instance, it took less than half a second on an IBM RS/6000 to compute the first missing column in table 3 of Ott (1992), resulting in $P(Z_{100}=5) + P(Z_{100}=4) = .9831$, $P(Z_{160}=5) + P(Z_{160}=4) = .9985$, and $P(Z_{200}=5) + P(Z_{200}=4) = .9997$, for $a = 5$ and $g_k = k^3/\sum_{i=1}^{5} i^3$.

MICHAEL KNAPP
*Institut für Medizinische Statistik*
*Dokumentation und Datenverarbeitung*
*Universität Bonn*
*Bonn*

## Reference

Ott J (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. Am J Hum Genet 51:283–290


0002-9297/93/5201-0030$02.00

---

*Am. J. Hum. Genet. 52:213–214, 1993*

## Robustness of the Maximum-Likelihood (LOD) Method for Detecting Linkage

*To the Editor:*

Ott (1992) recently described a pathological situation in which results from a linkage study can be misinterpreted to yield systematically biased estimates of the recombination fraction ($\theta$) between a marker and a trait and, moreover, in which falsely positive evidence for linkage can be generated. The conditions Ott used which yielded false-positive evidence for linkage consist of both a misspecification of parameters describing population frequencies of alleles at a marker locus and selection through trait phenotypes. Similarly, Green (1990) provided an example of asymptotic bias in a linkage study in which both the frequencies of the marker alleles and the segregation parameters for the trait locus are misspecified. In a previous study (Williamson and Amos 1990), we showed that, subject to some mild regularity conditions and to conditions that are *usually* met in a linkage study, maximum likelihood (ML) methods for detecting linkage are robust to misspecification of the trait-related parameters, i.e., these methods do not falsely detect evidence for linkage more often than is specified by the nominal significance level. The purpose of the present study is to clarify the conditions which led Ott (1992) and Green (1990) to obtain false-positive evidence for linkage and to contrast those conditions with the conditions under which our result holds.

In our previous analysis of this problem, we tacitly assumed either that all the marker genotypes in a pedigree were known or that the parameters describing the segregation of the marker alleles in the population studied were known and correctly specified in the analysis. When sampling of pedigrees is random, or when pedigrees are ascertained with respect to the trait phenotypes, we showed that, in the absence of linkage, the estimator of $\theta$ is asymptotically unbiased, regardless of whether the trait-related parameters are correctly specified. We also showed that, in the absence of linkage, the distribution of the likelihood-ratio test for linkage is not affected by misspecification of the trait-related parameters. Adapting our previous approach to the case in which the parameters describing the marker are misspecified allows us to make several statements regarding the robustness of ML linkage in this situation. Throughout the rest of this discussion we assume that the true value of $\theta$ is ½ and that neither the trait- nor the marker-related parameters are being estimated during the linkage analysis but that these have previously been specified.

1. By symmetry arguments, with respect to the marker and the trait loci, provided that either the marker- or the trait-related parameters are correctly specified and that pedigrees are randomly selected, the ML test is robust.
2. If both the trait and the marker parameters are misspecified, then an asymptotic bias may exist in the estimator of $\theta$, and the ML approach may yield false-positive results in excess of what would be expected under the level of significance being employed. Our previous results do not assess the effects of misspecifying both sets of parameters.
3. If sampling of pedigrees is not random with respect to the trait or the marker locus, then the estimator of $\theta$ might be asymptotically biased. If the trait-related parameters are incorrectly specified, then asymptotic bias may accrue if the ascertainment event affects selection of the marker phenotypes. Similarly, if the marker-related parameters are misspecified, then asymptotic bias may accrue if the

ascertainment event is through the trait phenotypes. Asymptotically, $\theta$ will not be biased if the marker-related parameters are misspecified and selection is through the marker phenotypes or if the trait-related parameters are misspecified and selection is through the trait phenotypes, provided that, in each of these situations, the other set of parameters is correctly specified. Selection through both the marker and the trait is not covered by our previous work.

4. Another tacit assumption made in our work was that there were not two different configurations of trait, marker, and linkage parameters that both lead to the assignment of the same probabilities to all possible outcomes. If this condition were violated, it would be impossible, of course, to distinguish between the two parameter configurations by using only the trait and the marker data.

The situation which Ott (1992) describes falls under category 3, and one therefore may encounter asymptotic bias if the marker-related parameters are incorrectly specified in analysis and if ascertainment is through the trait, even when the trait parameters are correctly specified. In the particular situation of affected sibships typed at a locus with unknown parental marker genotypes and misspecified marker-allele frequencies, the bias can be expressed as a function of the true and assumed marker-gene frequencies. For other family structures, such as the one described by Ott, it may be necessary to perform simulation studies to identify the possible asymptotic bias conferred by errors in specification of the marker-allele frequencies. In the case that marker genotypes are known for all of the pedigree members, the population frequencies of the marker alleles are irrelevant to the linkage analysis.

CHRISTOPHER I. AMOS* AND JOHN A. WILLIAMSON†
*Genetic Studies Section, National Institute of Arthritis, Musculoskeletal and Skin Diseases, Bethesda; and †Program in Applied Mathematics, University of Colorado, Boulder*

## References

Green P (1990) Comment. Genet Epidemiol 7:25–28

Ott J (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. Am J Hum Genet 51:283–290

Williamson JA, Amos CI (1990) On the asymptotic behavior of the estimate of the recombination fraction under the null hypothesis of no linkage when the model is misspecified. Genet Epidemiol 7:309–318



---

*Am. J. Hum. Genet. 52:214, 1993*

## Defining the Location of the Huntington Disease Gene

*To the Editor:*

As one not involved in research on Huntington disease (HD) but involved in offering presymptomatic and prenatal testing to families, I have been forced to follow the restless wanderings of the HD locus along distal 4p, at least as claimed by various research groups. The recent report in the *Journal,* by Snell et al. (1992), is one of the most recent of the many articles on this subject. Most of these articles purport to discuss the location of the HD locus when they are really reporting results regarding the location(s) of the mutation(s) in the HD gene in a particular family(ies). Linkage analysis does not measure distance between genes or DNA probes. It measures the distance between the unique region (or base) in the gene or anonymous DNA sequence that constitutes the mutation or polymorphism. I recognize that some believe that there have been very few founding HD mutations, but this is not a firmly established fact and cannot achieve this status until the gene is cloned and all mutations are identified. Perhaps some of the marker loci used in linkage studies lie within the HD gene and have quite different linkage relationships to different causative mutations. The two peaks of linkage disequilibrium 1 Mb apart (MacDonald et al. 1989) might be explained by two frequent sites of mutation this far apart in a very large gene.

DAVID M. DANKS
*The Murdoch Institute*
*Melbourne*

## References

MacDonald ME, Cheng SV, Zimmer M, Hains JL, Poustka AM, Allitto BA, Smith B, et al (1989) Clustering of multiallele DNA markers near the Huntington's disease gene. J Clin Invest 84:1013–1016

Snell RG, Thompson LM, Tagle DA, Holloway TL, Barnes G, Harley HG, Sandkuijl LA, et al (1992) A recombination event that redefines the Huntington disease region. Am J Hum Genet 51:357–362


0002-9297/93/5201-0032$02.00