

Likelihood Calculations in Paternity Testing on the Basis of DNA-Fingerprints

R. Fimmers^{*}, J.T. Epplen^{**}, P.M. Schneider^{***}, M.P. Baur^{*}

^{*} Institute for Medical Statistics, University of Bonn, Bonn, FRG

^{**} Max-Planck-Institute for Psychiatry, Martinsried, FRG

^{***} Institute of Legal Medicine, University of Mainz, Mainz, FRG

INTRODUCTION

DNA-fingerprints seem to be a powerful tool in paternity testing. But the quantification of the results in terms of likelihood and likelihood ratios is a problem, because in most cases the correct genetic model and its parameters are not known. Two approaches have been suggested to circumvent these problems. The use of band sharing rates to distinguish between pairs of relatives and pairs of unrelated individuals, and the calculation of likelihood ratios on the basis of simplifying assumptions. The first approach reduces the available genetic evidence to "phenotypic" similarities. The second one makes unjustified simplifying assumptions. These two decision strategies have to be examined with respect to their correctness. We applied both strategies to a set of real data and to a set of simulated data. The bandsharing rates and likelihood ratios for paternity were calculated for mother/child/father and for mother/child/non-father combinations, and the distributions were compared.

DATASETS

The real dataset consists of 27 families which were typed for a multilocus probe (Zischler, 1989). The probe was $(CAC)_5 / (GTG)_5$, the restriction enzyme was Hinf I. From these 27 families with 107 children we obtained 107 mother/child/father trios. By substituting one of the parents by an unrelated individual typed on the same blot, we got 87 mother/child/non-father trios out of 25 families.

Additionally we simulated two (father / non-father) sets of 5000 trios each. We allowed 200 band positions for 40 independent loci, which produced visible bands (bands within these 200 positions) with a probability of about 50%. The allele frequencies were chosen to be unequal.

METHODS

As stated above, we tested two different strategies. The first one was based on bandsharing. The bandsharing for a pair of individuals (here alleged father and child) was calculated as the average of the relative number of mutually shared bands for both individuals. The bandsharing was calculated for all possible combinations father/child and non-father/child in the four datasets.

The likelihood calculation was done according to a modified version of the method of Evett (Evett et al., 1989). This method calculates a likelihood ratio (paternity / non-paternity) independently for each possible band position. The total likelihood is the

product of all possible band positions. It is based on the level of band frequencies and the probability to inherit a specific band. This requires several assumption for the genetic model. Independent treatment of the different band positions requires linkage equilibrium, the absence of linkage and no allelism. To derive the transmission probabilities for the bands from the bandfrequencies, the assumption of Hardy-Weinberg equilibrium is necessary. Additionally we assumed, that the band frequencies are equal for all band positions. Eight different combinations of having a band or not are possible for mother, child and alleged father. In their paper Evett et al. did not consider the combination: child has a band, which mother and father do not have. If the alleged father is the true father, this has to be interpreted as a mutation. Because in our real data this kind of "mutation" occurred quite often (6% of the children), we modified Evett's procedure allowing for this band combination and taking it into account with an estimated probability for mutation. Another critical point of the method as proposed by Evett is the use of the band combinations with no band either for mother, alleged father or child, as an suggestion for paternity. This means, by increasing the number of "possible" band positions the likelihood ratio changes towards paternity. Because the number of possible band positions usually is unknown we decided not to use the information from such band positions. This results probably in an underestimation of the "true" likelihoodratio. To get a reliable estimate of the band frequencies we subdivided the range of occurring kb-value in 100 equal sized intervals for the real data as well as for the simulated data.

RESULTS

The distributions of bandsharing rates for the real data (paternity, non-paternity) were only scarcely overlapping. Only two bandsharing values under non-paternity (44.5% and 45.7%) were larger than the minimal bandsharing under paternity (44.2%). An optimal cutpoint for this data may be found 45%.

Table 1. Distribution of bandsharing (real data)

paternity	n	mean	minimum	maximum	> 45%	< 45%
yes	107	65.4%	44.2%	80.0%	106	1
no	87	18.6%	0.0%	45.7%	1	86

For the simulated dataset the result is different. The smallest bandsharing value for true fathers was 25.1%. 1330 trios (26.6%) with non-fathers were higher than this value. The highest value under non paternity was 45.3%. 262 (5.2%) of the bandsharing values under paternity were smaller than this value. An optimal cutpoint may be found with 39%.

Table 2. Distribution of bandsharing (simulated data)

paternity	n	mean	minimum	maximum	> 39%	< 39%
yes	5000	58.0%	25.1%	81.6%	4944	56
no	5000	20.9%	0.0%	45.3%	11	4989

For the real data the discrimination power of the likelihood ratio approach was not as good as that of the bandsharing method. The smallest ratio calculated under paternity was 9.55, but 3 (3.4%) of the likelihood ratios with non-fathers were higher. The highest ratio under non-paternity was 97.1 and 9 (8.4%) of the ratios under paternity were smaller.

Table 3. Distribution of likelihood ratios (real data)

paternity	n	mean	minimum	maximum	> 1	< 1
yes	107	1.1×10^4	9.6×10^0	4.2×10^8	107	0
no	87	4.2×10^{-2}	1.1×10^{-5}	9.7×10^1	19	68

For the simulated data our modified likelihood approach seems to be better than the bandsharing. The smallest value under paternity was 6.31×10^{-3} and 46 (0.9%) likelihood ratios for non-fathers were higher. The maximum ratio for non-father trios was 2.9×10^2 and 1272 (25.4%) of the ratios under paternity were smaller.

Table 4. Distribution of likelihood ratios (simulated data)

paternity	n	mean	minimum	maximum	> 1	< 1
yes	5000	5.0×10^3	6.3×10^{-3}	6.4×10^{10}	4919	81
no	5000	6.9×10^{-8}	6.8×10^{-16}	2.9×10^2	2	4998

DISCUSSION

The discrimination power of both methods seems to be very good. The better results for the real data in comparison to the simulated data may be due to the construction of trios out of the same families, which means dependence of bandsharing rates and likelihood ratios. The results for the two methods are similar. The cutpoint of likelihood ratio 1 is not an optimal cutpoint as chosen for the bandsharing. For both methods one should use a region of uncertainty, where no decision is made. These results do not confirm the calculated likelihood ratios. They only show, that the qualitative decision is correct in most cases.

REFERENCES

Evett IW, Werrett DJ, Buckleton JS (1989) Paternity calculations from DNA multilocus profiles. *J Forens Sci Soc* 29: 249-254

Zischler H, Nanda I, Schäfer R, Schmidt M, Epplen JT (1989) Digoxigenated oligonucleotide probes specific for simple repeats in DNA fingerprinting and hybridization in situ. *Hum Genet* 82: 227-233

Supported by DFG Ba 660/5-3 and the VW-Stiftung.