Paternity probability calculation in "exclusion" cases

A. Amorim, J. Rocha

Instituto de Antropologia, Universidade do Porto, Portugal

INTRODUCTION

Recent advances of paternity investigation have shown two apparently contradictory developments: (a) the growing information power of the tests, corresponding to very high paternity probabilities among true fathers, and (b) the loss of absoluteness of the "exclusion" concept, since the number of false exclusions is expected to grow with the increasing number of genetic systems employed.

This has been widely recognized, and operative solutions have been advanced for "difficult" cases where single 2nd. rule exclusions are observed (Gürtler, 1977; Polesky and Souhrada, 1988).

The discovery of false 1st. rule exclusions at a relatively common rate (from 10^{-4} up to more than 10^{-3}) in some genetic systems (Wetterling, 1988; Zang and Blin, 1988; Werret et al., 1988) demonstrates that even this kind of exclusion, when isolated, can no longer be treated in a qualitative fashion, but, on the contrary, must be quantified along with the rest of the genetic evidence gathered in the case under dispute.

Since we have already discussed in a previous work (Amorim and Rocha, 1988) the disadvantages of a non-uniform treatment of the genetic results when an exclusion is present, we shall limit ourselves here to the derivation of an algorithm able to deal in a biostatistically uniform manner with results of 1st. order exclusions.

DEFINITIONS

We will call paternity index (L) to the ratio between the probability of occurrence of the genetic results under the hypothesis of paternity and the probability of occurrence of

Advances in Forensic Haemogenetics 3
Edited by H. F. Polesky and W. R. Mayr

the genetic results under the opposite hypothesis. For the calculation of these probabilities, estimates of population parameters are needed: gene frequencies (including silent genes) and mutation rates. Mutation will be understood here as any violation of the formal genetic model established for the system under analysis that cannot be explained by a silent gene. It includes, therefore, along with "classical" mutations, intragenic recombinations and unequal crossovers.

DERIVATION AND PROPERTIES OF THE ALGORITHM

Let us assume that we observe a trio where the child demonstrates a gene, absent from both mother and putative father (phenotypes: 1-X, 1 and 1, respectively). It is obvious that such a situation can only be reconciled with the hypothesis of paternity, assuming a mutation event during gametogenesis, either maternal or paternal. Thus, the probability of occurrence of this observation under the hypothesis of paternity is:

$$p^2.\mu.p^2(1-\mu) + p^2(1-\mu).p^2.\mu = 2\,p^4\mu(1-\mu) ,$$

where p stands for gene 1 frequency and μ for mutation rate. Correspondingly, the probability of the same observation under the hypothesis of non-paternity, is:

$$[p^2.\mu.p + p^2(1-\mu).q].p^2 = p^4[\mu p+q(1-\mu)] ,$$

where q is gene X frequency. Thus,

$$L = 2\mu(1-\mu)/[\mu p+q(1-\mu)].$$

From the expression it is easy to infer that L is minimum when μ is low and q very high; in other words, and in accordance with common sense, the reliability of the exclusion is high when mutation rate is low and the gene present only in the child is common in the population. On the other hand, the maximum value of L=1 (assuming no practical interest in a theoretical situation where μ could be higher than q) is reached when $q=\mu$ and $p=1-q$. That is to say: the reliability of the exclusion is minimal when the gene present in the child is so rare that it is equally probable to explain its occurrence by mutation or by normal gametogenesis from an unknown true father. More bluntly: if the described situation occurs in a monomorphic system and the gene is detected for the first time, the reliability of the "exclusion" is zero.

APPLICATIONS AND DISCUSSION

Reports on the occurrence of 1st. rule exclusions in situations where kinship is not doubtful have accumulated recently and data are now available in the literature for estimation of mutation rates in some genetic systems (Martin, 1981; Wetterling, 1988; Werret et al., 1988).

If we look upon PGM1, for instance, 4 exclusions by the first rule out of 12,682 mother-child pairs were found (Wetterling, 1988). The same author concludes that "at the present, it is preferable not to base any paternity exclusion on the PGM1 system alone".

We agree in the sense that no exclusion should be made in a qualitative way, specially if based on an isolated incompatibility, disregarding the rest of the genetic evidence. However, we think it is very difficult to sustain before experts and laymen that the same system can be reliably used when "including" but it is suspicious when "excluding".

In fact, using the algorithm above, a unified biostatistical approach to such cases is possible, allowing a quantified opinion on paternity using all the information available.

Indeed, the available data lead to a conservative estimate of $\mu = 4.4 \times 10^{-4}$ (see Appendix). Using this estimate, gene frequencies from the same source and the above algorithm, we get L values ranging from 10^{-2} to 10^{-3}.

Since the test battery used now in many laboratories is sufficiently powerful to produce an average L among true fathers exceeding 3×10^{2}, we believe that no longer a single 1st. rule exclusion in PGM1 is a major difficulty for paternity expertises.

Another illustration of the use of the algorithm comes out from a paternity case we analysed where, after the routine battery, the child evidenced a previously undescribed gene product, absent from both mother and accused man in ESD system (phenotypes: 1, 1 and 1-V, respectively). However, using the other systems, paternity index was 9.4×10^{3}. Extended investigations, including HLA, were performed and L raised to 2.1×10^{7}, no other exclusion being found.

In order to join all the genetic evidence in a single figure we would need estimates for both mutation rate and gene frequency, both non-existing in the present case. Anyway, even assuming a very low mutation rate (10^{-6}) and an absurdly high gene frequency (10^{-2}), the global paternity index would be still 41.

CONCLUSIONS

Paternity cases where a single exclusion is detected can now be dealt with using a uniform biostatistical approach weighting all the evidence available, no matter what class of exclusion is observed. The difficulties arising from the expression, in the same case report, of quantitative and qualitative expert's opinions are avoided, allowing a better understanding of the facts at stake by a layman.

REFERENCES

Amorim A, Rocha J (1988) The bias of the dichotomy exclusion / non-exclusion and the evidential value of L (or W). In: Mayr WR (ed) Advances in Forensic Haemogenetics 2. Springer Verlag, Berlin, p 603-606

Gürtler, (1977) Evaluation of the reliability of exclusions of paternity in Denmark. In: Proc. 7th International Congress International Society for Forensic Haemogenetics, Hamburg, p 577-581

Martin W (1981) Red cell enzyme groups in paternity testing. In: Proc. 9th. International Congress International Society for Forensic Haemogenetics, Bern, p 221-239

Polesky HF, Souhrada JM (1988) A protocol for reporting single indirect exclusions. In: Mayr WR (ed) Advances in Forensic Haemogenetics 2. Springer Verlag, Berlin, p 600-602

Werret DJ, Gill PD, Lygo JE, Fowler SJ (1988) DNA polymorphisms - practical use. In: Mayr WR (ed) Advances in Forensic Haemogenetics 2. Springer Verlag, Berlin, p 320-338

Wetterling G (1988) Discrepancy between gene and protein products within the PGM1 system shown by improved resolution on immobiline gels. In: Mayr WR (ed) Advances in Forensic Haemogenetics 2. Springer Verlag, Berlin, p 97-101

Zang KD, Blin N (1988) Recombinant DNA technology and human DNA-polymorphism. In: Mayr WR (ed) Advances in Forensic Haemogenetics 2. Springer Verlag, Berlin, p 311-319

APPENDIX

For a n (>2) allele codominant genetic system, μ can be estimated from the frequency of mother-child 1st. order exclusions (f):

$$f = (n-2)\ \mu\ [\sum_{i=1}^{n} p_i^2(1-p_i) + \sum_{i>j}\sum 2p_ip_j(1-p_i-p_j)]$$

4 Varia