Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics Supplement Series 1 (2008) 635–637
www.elsevier.com/locate/FSIGSS

Research article

Towards a systematic probabilistic evaluation of parentage casework in forensic genetics: A modest attempt to define a general standardized approach to simple and complex cases

D. Abrantes [a,*], M.L. Pontes [a], M.F. Pinheiro [a,b,c], M. Andrade [d], M.A.M. Ferreira [d]

[a] *Delegação do Norte do Instituto Nacional de Medicina Legal, Jardim Carrilho Videira, 4050-167 Porto, Portugal*
[b] *Faculdade de Ciências da Saúde – Universidade Fernando Pessoa, Portugal*
[c] *Instituto de Ciências Biomédicas de "Abel Salazar" – Universidade do Porto, Portugal*
[d] *ISCTE Business School, Departamento de Métodos Quantitativos, Portugal*

Received 17 August 2007; accepted 9 October 2007

Abstract

In view of the increasing demand of ever-complex parentage casework, our lab saw itself in the position to define a strategy as standardized as possible to deal with the bulk of those cases. We searched for a coherent and exhaustive way to find answers. The major aim of this communication is to present the heuristical, mathematical and other aspects of a probabilistic evaluation procedure that has been developed since 2003. We also present a theoretical casework to illustrate one of the many problems that have to be faced: the choice of suitable statistical hypotheses.

Keywords: Probabilistic evaluation; Parentage casework; Complex cases

1. Introduction

In the last decade, the Forensic Genetics Lab (FGL) of D.N.I.N.M.L. (affiliation a, above) has been experiencing an ever-increasing demand, in quantity and complexity, of parentage testing casework. This was already mentioned before [1].

For that reason, the FGL sought a probabilistic evaluation and interpretation procedure as coherent and exhaustive as possible. Such procedure would accommodate in the same conceptual framework the global treatment of ''simple'' and ''complex'' casework, without disrespecting the specificities of each case. This framework, developed since 2003, and continually adjusted to every new situation that meanwhile emerged, does not intend to be an absolutely exhaustive evidence evaluation procedure, and by no means a prescriptive one. Its main characteristics are briefly outlined below.

2. Methods

The FGL's probabilistic evaluation and interpretation methodology can be arbitrarily divided in five stages: (1) scientific hypotheses definition; (2) hypotheses' prior probabilities assessments; (3) Likelihood ratio calculations; (4) hypotheses' posterior probabilities assessments and (5) interpretation according to Hummel's chart.

FGL's forensic experts and ISCTE statisticians are aware that all of these stages except stage (3) are not really the province of the expert, they are the province of the trier-of-fact. Nevertheless, in Portuguese courts of law there is some tradition (hard to change) in presenting the evaluation as posterior probability, interpreted under Hummel's verbal predicates.

Stage (1): This stage offers the greatest problems. The choice of suitable statistical hypotheses is not necessarily automatic [2]. For that purpose, the following set of heuristic guidelines was developed:

– In a case where a set of individuals is to be tested, ''fixed'' genealogical relationships (i.e., assumed as biologically true by the court) among them are recorded by the experts.
– Hypotheses' enumeration is made according to the combinatorial possibilities of genealogical relations for individuals

* Corresponding author. Tel.: +351 222 073 850; fax: +351 223 325 931.
E-mail address: davidabrantes@gmail.com (D. Abrantes).

1875-1768/$ – see front matter
doi:10.1016/j.fsigss.2007.10.132

who do not have fixed relations between them. Each particular combination constitutes a pedigree. Each hypothesis may include one or more of such pedigrees. Pedigrees gathered under a particular hypothesis represent the same global genealogical reality. They differ in combinatorial details (see Fig. 1) that will not yield differences in likelihood (see below) among them. Hypotheses' listing is subjected to constraints (e.g., information provided by the court, age/generational hierarchy of individuals, etc.).

– To assure a hypotheses' set as exhaustive as possible sometimes the presence of extra individuals, non-phenotyped and hence with no known genetic information, has to be considered (see individuals PX, PY and PZ, Fig. 1.). Thereby, a minimum number of phenotyped and non-phenotyped individuals are considered. That number does not change among hypotheses. What changes is the non-fixed relations among those individuals.

Stage (2): A discrete, uniform prior distribution is adopted. If the hypotheses' set contains m pedigrees and hypothesis H_i ($i = \{1, 2, \ldots, i, \ldots, n\}$; $n \leq m$) includes k pedigrees, then the prior probability of hypothesis H_i is $P(H_i) = k/m$.

Stage (3): After having the phenotyping data, the likelihood of each hypothesis given the genetical data (D) – $L(H_i|D)$ – is obtained through the calculation of probability of genetical data assuming that the relations described by the hypothesis' typical pedigree (one chosen among the k possible, denoted

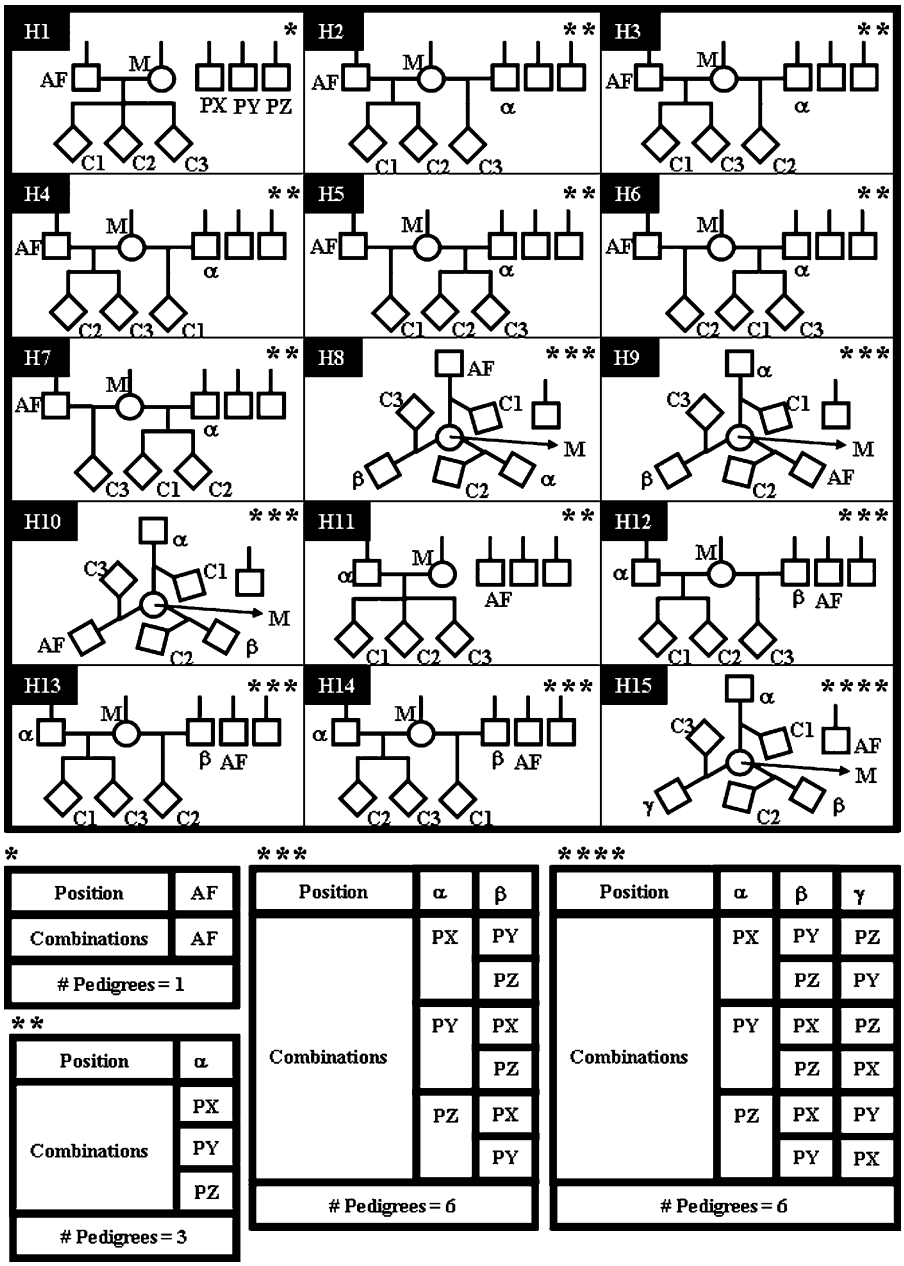

Fig. 1. Hypotheses and pedigrees of the theoretical casework example. Each hypothesis H_i ($i = \{1, \ldots, 15\}$) consists of one or more pedigrees which represent the same global genealogical reality, differing in combinatorial details represented in the tables underneath. The asterisk(s) in the superior right corner of each hypothesis indicate(s) the respective bottom table where such details are displayed. "# Pedigrees" stands for the number of pedigrees.

as $\mathrm{Ped}_r(\mathrm{H}_i)$, $r = \{1, \ldots, k\}$) are true—$P(D|\mathrm{Ped}_r(\mathrm{H}_i))$. That results from the fact that $L(\mathrm{H}_i|D) \propto P(D|\mathrm{Ped}_r(\mathrm{H}_i))$. The general algorithm stems from the chain rule or multiplication rule of probability calculus, named by some the "Third Law of Probability" [3]:

$$P(D|\mathrm{Ped}_r(\mathrm{H}_i)) = P(G_1)\times P(G_2|G_1) \times \ldots \times P(G_s|G_1 \cap G_2 \cap \ldots \cap G_{s-1}) \quad (1)$$

G_j stands for the phenotype of the jth individual, $j = \{1, 2, \ldots, s\}$. The chain has chronological ordering: jth individual can only belong to the same or a later generation than $(j-1)$th person. Next, Likelihood ratios of the sort $\mathrm{LR}_i = P(D|\mathrm{Ped}_r(\mathrm{H}_i))/P(D|\mathrm{Ped}_r(\mathrm{H}_n))$, using H_n as common denominator, are calculated.

Stage (4) Hypotheses' posterior probabilities are obtained via Bayes' theorem, according to the expression:

$$P(\mathrm{H}_i|D) = \frac{(P(\mathrm{H}_i) \times \mathrm{LR}_i)}{\sum_j (P(\mathrm{H}_j) \times \mathrm{LR}_j)}, \quad j = \{1, 2, \ldots, i, \ldots, n\}$$

Stage (5): Posterior probability values are compared to Hummel's chart [4]. Hypothesis with posterior probability equal or greater than 0.9973 is accepted by the court as "practically proven"—if no objections are put to evidence's evaluation conditions. The remaining ones get, subsequently, "practically excluded".

3. Results and discussion

3.1. Theoretical casework example

In this section it is mainly intended to exemplify the set of hypotheses that is put forward, as a result of the guidelines referred earlier, in the following general case: the court wants to know about the paternity status of individual AF relative to three of his putative children (C1, C2 and C3), all of these assumed biologically mothered by M. The only fixed relations are those between M and each one of her assumed children. To have the most exhaustive set of hypotheses one has to go from the situation in which AF is most related to the children, fathering all three of them, to the one in which not only does he not father any of them but also they all have different fathers. This last situation forces one to consider the existence of at least three different individuals PX, PY and PZ, besides AF, from a population of interest, which generally will not have known genetic information. All this is depicted in Fig. 1.

4. Conclusion

It is a firm belief of the forensic experts and statisticians involved in this work, that the conceptual framework outlined earlier represents a systematic, yet flexible, way for evaluation of evidence in parentage investigation settings. Features such as: (1) multiple hypotheses testing, as needed, under a single testing universe; (2) consideration of the same number of individuals involved whichever the hypothesis considered, the difference being that between hypotheses some of the inter-individual genealogical relations will vary and (3) propagation of probability using Eq. (1), following the basic laws of probability, and employment of a general Bayesian structure, concede breadth and logical coherence to the framework.

Conflict of interest

None.

References

[1] D. Abrantes, L. Pontes, G. Lima, L. Cainé, M.J. Pereira, P Matos, M.F. Pinheiro, Complex paternity investigations: the need for more genetic information, in: A. Amorim, F. Corte-Real, N. Morling (Eds.), Progress in Forensic Genetics 11—Proceedings of the 21st International ISFG Congress, Elsevier B.V, Amsterdam, 2006, pp. 465–467.

[2] A. Stockmarr, The Choice of hypotheses in the evaluation of DNA profile evidence, in: J.L. Gastwirth (Ed.), Statistics for Social Science and Public Policy—Statistical Science in the Courtroom, Springer-Verlag, New York, Inc., New York, 2000, pp. 143–159.

[3] I.W. Evett, B.S. Weir, Interpreting DNA Evidence—Statistical Genetics for Forensic Scientists Sinauer Associates, Inc., Sunderland, U.S.A., 1998.

[4] http://dna-view.com/hummel.htm.