

LETTERS TO THE EDITOR

Forensic medicine, PCR, and Bayesian approach

The recent article by Decorte and Cassiman, entitled "Forensic medicine and the polymerase chain reaction technique" (*J Med Genet* 1993;30:625-33) is strong in technological aspects, but certain arguments advanced in the section concerning "DNA typing in court" need to be clarified.

This part of the article is, in our opinion, fundamental, because the expert witness in forensic science has to provide to the courtroom not only a technology but also an evaluation of the strength of the link(s) established by the analysis of physical evidence.

In this context, the purpose of this letter is to develop three arguments cited in the paper of Decorte and Cassiman (p 631).

(A) "In case of a match, the court wants to know the significance of this event, or what the probability is that somebody else has committed the crime. In answering this question, the forensic expert calculates the population frequencies of the genotypes for the different loci."

(B) "In forensics, it is appropriate to use reference data on the population from which the suspect originated."

(C) "An alternative approach would be to use a likelihood ratio (LR) which is simply the inverse of the genotype probability."

To support our remarks, we shall briefly recapitulate the essentials of the Bayesian approach in forensic science. This probabilistic model allows one to evolve the given "prior" odds on the existence of a fact (F) in the light of new information (E), to obtain "posterior" odds on the existence of the fact, through simple multiplication by a ratio called the likelihood ratio (LR). In forensic science, F can be seen as the event in which "the suspect is involved in the trace transfer" and E as the event in which "physical traces link a suspect with a crime, for example, a blood stain".

Bayes's theorem is illustrated by the following formula:

$$O(F|I,E) = \frac{P(E|F)}{P(E|\bar{F})} O(F|I)$$

LR

where I is the background information necessary to evaluate "prior" odds.

In what follows, I is omitted for the sake of brevity.

$$O(F|E) = \frac{P(E|F)}{P(E|\bar{F})} O(F)$$

LR

O(F) The odds on the suspect's involvement before scientific analysis, the "prior" odds on F which is equal to P(F)/P(\bar{F}).

O(F|E) The odds that the suspect is involved given the evidence, the "posterior" odds on F which is equal to P(F|E)/P(\bar{F} |E).

P(E|F) The probability of concordant evidence, given that the suspect is involved.

P(E| \bar{F}) The probability of the concordant evidence, given that the suspect is not involved.

This formulation clarifies the position of the scientist as well as that of the jury. The

scientist is concerned solely with the LR, whereas juries deal with the odds on F.¹

The statement (A) by Decorte and Cassiman leads us to believe that the genotype frequency (f) will directly determine the probability that someone other than the defendant is involved. The "reasoning" is that the probability that the suspect is not involved after the presentation of the DNA evidence P(\bar{F} |E) is equal to the genotype frequency (f).

Consider an example² in which a juror hears that the probability that an innocent person would possess the traits that characterise the offender is 1 in a 1000 (f=1/1000). He or she may misconstrue this figure to mean that 99.9% (1-1/1000) must represent the probability of guilt.

This argument constitutes an elementary but common error called "prosecutor fallacy" or "inversion fallacy",³ many recent examples of which are presented in the article by Kaye.⁴ This "inversion fallacy" tempts the jury to interpret the probability without knowledge of the size of the reference population (factor determining the "prior" odds).

The fallacious argument is posed in the following equation:

$$P(F|E) = 1 - f \text{ then } P(\bar{F}|E) = 1 - P(F|E) = f.$$

The correct equation is:

$$P(F|E) = 1 - P(\bar{F}|E).$$

The fallacy consists then in replacing P(\bar{F} |E) by f, whereas f is in reality P(E| \bar{F}).

The Bayesian approach allows us to avoid this classic error, since f must influence only the LR and not O(F). O(F) depends on information (I) about the case before the DNA evidence.

The second statement (B) is completely in error even though unfortunately it has been adopted in several different criminal trials. The solution to the question of the reference population is found in the question: "Why are scientists interested in chance coincidence probability?" This probability implies the acceptance of the hypothesis that the discovered trace at the crime scene came from some other person than the suspect. In the light of this statement, the ethnic group of the suspect has absolutely no bearing on this probability.⁵ When we speak of chance coincidence probability, it is implicitly admitted that the hypothesis of the suspect's non-implementation is taken into consideration.

The chance coincidence probability must be based on the data of a reference population which is the sum of all the persons who could be at the origin of this physical trace (a blood stain, for example).⁶ This population is determined solely by the trace found on the crime scene and by no means by the suspect. This is a key point in forensic science.⁷

The statement (C) is far too restrictive: the LR is not the inverse of the genotype probability alone, as illustrated in the following examples describing the evaluation of the LR.

In a simple burglary case,⁸ a trace is recovered at the scene under such conditions that it could only come from one person. A suspect is then apprehended and his blood type corresponds to the crime stain (type α , with a frequency f in the general population). If the suspect is really the perpetrator of the crime, then the blood type must correspond, and the numerator of the LR will be 1. Under the opposite hypothesis, the suspect did not leave the stain, the probability of a random correspondence of blood type α being f. The likelihood ratio is, in this particular case, simply 1/f as described in the paper under review, but the first hypothesis (the trace did come from one perpetrator) is not always possible in this type of offender(s) to scene transfers, and therefore the LR is not always easy to calculate.^{9,10}

Also, the expert frequently deals with traces which come from the crime scene or the

victim, for example, blood stains on a suspect's piece of clothing. Let us consider an example of this type of scene to offender(s) transfer. Suppose that a person has been found, stabbed to death. A suspect has been apprehended and an examination of his clothing shows a blood stain on his jacket which is of the same blood type as that of the victim (α). The suspect is of another blood type (β). The interpretation of the facts becomes more complex here, because the trace is not necessarily relevant to the offence. The expert must consider the possibility that the transfer trace came about under purely innocent circumstances under both hypotheses F and \bar{F} , therefore the numerator of the LR is not simply equal to 1 and the denominator is not simply equal to f.

Some of the questions that the expert must answer, to appreciate the force of the relationship, are: What is the probability of a blood stain of this observed shape and size, if the suspect did in fact stab the victim? What is the probability of finding a blood stain from someone else of the observed shape and size on the clothing of a person taken at random? What is the probability that a blood stain from someone else on clothing of a person taken at random would be type α ?

Information about other technical and investigative observations, and the use of adequate surveys, can help the expert in the LR evaluation. We can easily see that in this type of transfer, the frequency f of the blood type α is only one of a series of factors influencing the LR value.

In the absence of all relevant surveys, it is impossible to calculate the LR, but the Bayesian theory constitutes, for the expert, an interesting subject for reflection on scientific proof, because the expert must search and choose the relevant questions and consider the physical evidence from two opposite points of view.¹¹

These different arguments are, in our opinion, of the utmost importance. The expert must be aware that the statistical evaluation of DNA evidence requires more than the determination of a genotype frequency in a given population. Probabilistic arguments can lead to dangerous and fallacious arguments and the force of a relationship depends also upon the circumstances of transfer.

F TARONI

C CHAMPOD

Institut de Police Scientifique et de Criminologie,
University of Lausanne,
Place du Château 3,
CH-1005 Lausanne,
Switzerland

1 Evett IW. Interpretation: a personal odyssey. In: Aitken CGG, Stoney DA, eds. *The use of statistics in forensic science*. New York: Ellis Horwood 1991:9-26.

2 Kaye DH. The admissibility of "probability evidence" in criminal trials. Part II. *Jurimetrics J* 1987;27:160-72.

3 Thompson WC, Schumann EL. Interpretation of statistical evidence in criminal trials: the prosecutors' fallacy and the defense attorney's fallacy. *Law and Human Behavior* 1987;11:167-87.

4 Kaye DH. DNA evidence: probability, population genetics, and the courts. *Harvard Journal of Law and Technology* 1993;7:101-72.

5 Evett IW, Weir BS. Flawed reasoning in court. *Chance* 1991;4:19-21.

6 Wooley JR. A response to Lander: the courtroom perspective. *Am J Hum Genet* 1991;49:892-3.

7 Stoney DA. Evaluation of associative evidence: choosing the relevant question. *J Forensic Sci Soc* 1984;24:1-23.

8 Evett IW. What is the probability that this blood came from that person? A meaningful question? *J Forensic Sci Soc* 1983;24:1-23.

9 Evett IW. On meaningful questions: a two-trace problem. *J Forensic Sci Soc* 1987;27:375-81.

10 Stoney DA. Relaxation of the assumption of relevance and an application to one-trace and two-trace problems. *J Forensic Sci Soc* 1994;34:17-21.

11 Champod C, Taroni F. Les préjugés de l'accusation ou de la défense dans l'évaluation de la preuve technique. *Revue Pénale Suisse* 1993;111:223-35.