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 technical aspects, but certain arguments advanced in the section concerning "DNA typing in criminal need to be clarified.

This part of the article is, in our opinion, fundamental, because the expert witness in forensic medicine 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 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’ theorem is illustrated by the following formula:

\[ P(F|E) = \frac{P(E|F) \cdot O(F)}{P(E|F) \cdot O(F) + P(E|\neg F) \cdot O(\neg F)} \]

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

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

\[ P(F|E) = \frac{P(E|F) \cdot O(F)}{P(E|F) \cdot O(F) + P(E|\neg F) \cdot O(\neg F)} \]

LR

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

\[ P(F|E) = \frac{P(E|F) \cdot O(F)}{P(E|F) \cdot O(F) + P(E|\neg F) \cdot O(\neg F)} \]

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

\[ P(E|F) \]

The probability of concordant evidence, given that the suspect is involved.

\[ P(E|\neg F) \]

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 justices deal with the odds on F.1

The statement (A) by Decorte and Cassiman leads us to believe that the genotype frequencies involved are the same as those represented in the presentation of the DNA evidence P(F|E) is equal to the genotype frequency (F). Consider an example in which a juror hears that the probability that an innocent person would possess the traits that characterise the offender is 1 in 1000 (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 is essentially an elementary but common error called "prosecutor fallacy" or "inversion fallacy," many recent examples of which are presented in the article by Kaye.2 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:

\[ P(F|E) = 1 - f \] and P(F|E) = 1 - P(F) = f.

The correct equation is:

\[ P(F|E) = 1 - P(E|F) \]

The fallacy consists then in replacing P(F|E) by 1-P(F) in the probability (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 (E) about the case before the DNA evidence.

The second statement (B) is completely at odds even though 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. If we speak of chance co-incidence 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).3 This population is determined solely by the facts found on the crime scene and by no means by the suspect. This is a key point in forensic science.

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.4 A trace is recovered at the scene under such conditions that it could only come from one person. A suspect is then taken into account and his blood type corresponds to the crime stain (type 0\(\alpha\), 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 \(\alpha\) being \(f\). The likelihood ratio is, in this particular case, simply \(1/f\) as described in a paper under review, but the first hypothesis (the trace did come from one perpetrator) is not always possible in this DNA analysis (e.g. in blood transfusions), and therefore the LR is not always easy to calculate.5

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 the perpetrator has four times the chance of being a person X, while the suspect X has been apprehended and an examination of his clothing shows a blood stain on his jacket which is of type 0\(\alpha\). The likelihood ratio is that of another blood type (\(\beta\)). The interpretation of the facts becomes more complex here, but that is not necessarily relevant to the occurrence. The expert must consider the possibility that the transfer trace came about under many unusual circumstances. In these circumstances, LR is not simply equal to 1 and the denominator is not needed, either.

Some of the questions that the expert must answer, to appreciate the force of the ratio of DNA evidence in any case of trace blood stain of this observed shape and size, if the suspect did in fact stab the victim? What is the probability that a blood stain from someone else on clothing of a person taken at random would be type \(\alpha\)?

Information about other technical and investigative observations, and particularly about multiple equate surveys, can help the expert in the LR evaluation. We can easily see that in this type of transfer, the frequency of type \(\alpha\) 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 LR allows us to choose the relevant questions and consider the physical evidence from two opposite perspectives.6 These different arguments are, in our opinion, of the utmost importance. The expert must be aware that the statistical evaluation of DNA evidence is not 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.

P. TARONI
C CHAMPIONS
Institut de Police Scientifique et de Criminologie, Université Lausanne, Place du Château 3, CH-1005 Lausanne, Switzerland