Charles Kingston,[1] D. Crim.

A Perspective on Probability and Physical Evidence

REFERENCE: Kingston, C., "**A Perspective on Probability and Physical Evidence,**" *Journal of Forensic Sciences*, JFSCA, Vol. 34, No. 6, Nov. 1989, pp. 1336–1342.

ABSTRACT: Evaluating the significance of physical evidence often requires a consideration of the probability of chance occurrence of the evidential properties and an appropriate target population. These are used to calculate the probability of duplication of the properties within the target population, which in turn is used to evaluate the significance of the evidence. The target population chosen for consideration by the expert witness generally cannot be the same as that used by investigators or a jury. This paper discusses the choice of an appropriate target population by the expert and offers a suggestion for aiding the jury in evaluating the expert's opinion with respect to other evidence in a case.

KEYWORDS: forensic science, jurisprudence, physical evidence, witnesses, probability, jury, testimony

Recent discussions in the literature and among colleagues have suggested some issues in the use of probability in forensic science that call for further consideration. This paper examines some aspects of interpreting probability calculations in forensic science examinations and the presentation of conclusions by an expert witness that are based partly or totally upon such considerations.

The discussion here is based upon those properties of evidence that occur according to some distribution D within some population P. A specific instance occurs with probability p within the population P. Some examples of physical evidence are (1) blood groups within a sampled population, (2) head hair characteristics within a sampled population, and (3) glass refractive index values within a sampled population. The properties represented in the examples have respectively discrete, discrete with fuzzy boundaries, and continuous distributions.

The significance of an observed property (or set of properties) of an item of evidence depends upon the value of p for that property (or set of properties), and the size (N) of the population (P) under consideration. These can be combined to provide the crucial probability of duplication of the evidence within the population P. See Refs *1* and *2* for a detailed discussion of this with respect to physical evidence.

If we are dealing with a situation in which we are trying to establish that a suspect is the source of the physical evidence, a very small probability of duplication within P would be supportive of a connection between the suspect and the evidence. As p gets larger than $1/N$, the probability of duplication approaches unity, and the weight of the evidence may best be stated in terms of the expected number of people in P who could leave

Received for publication 30 Sept. 1988; revised manuscript received 16 Nov. 1988; accepted for publication 21 Dec. 1988.

[1]Professor of criminalistics, John Jay College of Criminal Justice, New York, NY.

physical evidence with the same characteristics. This is roughly equivalent to estimating the expected number of duplications within P.

Unfortunately, it is not always clear how the population under consideration should be defined. For individuals, the size of P may range from a few suspects in an aircraft in flight to the population of the world, depending upon the circumstances.

With objects, the situation is somewhat problematic. For instance, if a small piece of glass occurs as evidence and we want to know the significance of its properties matching those of glass from a suspect source, it can be a problem to specify the size of the population of glass objects to be considered. This can often be circumvented by considering the relation of the suspect to the evidence (for example, determining the probability of an individual having a piece of glass on his or her clothing with the particular properties). For this discussion, only populations of individuals will be considered.

We can evaluate evidence from the point of view of estimating the probability of occurrence or from the point of view of reducing the size of the (suspect) population to be considered. It seems more natural to estimate probabilities for physical evidence and to reduce the population size for other types of evidential considerations, such as motive and opportunity. Investigators will initially use the available evidence to define a reduced population within which they will look for the person or persons responsible for a crime or other incident.

Frequency of Occurrence Analysis

Because the role of forensic scientists (excluding psychiatry and allied fields) is to provide expert opinion about physical evidence, they will generally deal with the (objective or subjective) probability of such evidence. An opinion about the specific significance of the examined physical evidence in a case requires either an estimation of the probability of duplication of the evidence, or perhaps an estimation of the expected number of duplications (when the expected value is more than 1), within some target population. Different experts who examine the same evidence may be in general agreement about the probability of occurrence, but differ in their opinions about the significance of the evidence because they have chosen to use different target populations.

It may be useful before continuing to examine a question (posed by a reviewer) that might occur to the reader:

Question—Why not simply use the frequency of occurrence and just report this to the jury (and thus avoid worrying about a target population)?

Answer—The frequency of occurrence is too easily misunderstood and misinterpreted by those without adequate statistical training. A frequency, such as $1/N$, is easily interpreted as a probability of duplication (which it is not), or as implying that the characteristic will not be duplicated in a population of size N (which it does not).[2]

When neutron activation analysis (NAA) was first being presented as a wonder tool for tracing hair and other possible evidential materials back to their source, the claims were justified by statements of frequencies about finding any particular combination of elemental quantities. These impressively low frequencies, which were often misinterpreted, served to generate expectations that NAA would indeed provide positive linkage of materials such as hair to their source. Consider the following [*3*]:

[2]If some event, say E, has a probability of chance occurrence (frequency) of p within some population of size N, then the most likely single outcome is that E occurs pN times. Thus, if $p = 1/N$, the most likely *single* event is that E occurs once. However, one of the alternative occurrences of E (1, 2, 3, . . . times) is more likely. If we know that E has occurred once, the probability that it has actually occurred more than once is about 0.4.

> Jervis and Perkons computed the odds for the duplication of one and the same element pattern in both quantity and quality. The results were as follows:
> If only 3 elements were measured, the probability of duplication came to 1:4800. With 5 elements measured, it was 1:126,000. With 7 elements, it rose to 1:4,250,000; and for 11 elements, 1:1,140,000,000. From these figures it followed that if in comparison of two samples of hair 11 elements of the same kind and quantity were found, identity was as good as established.

It is clear from reports on the original work that the figures are actually calculated frequencies of occurrence, and not probabilities of duplication (although essentially interpreted as that by the authors). See Ref *4* for an example of this. The evaluation of NAA results for hair with respect to forensic examinations was eventually put into its proper perspective (for example, see Ref *5*) by combining some well-designed experimental work with a reasonable statistical model.

A similar pattern seems to be emerging with respect to a more recent wonder tool (or set of tools) used for deoxyribonucleic acid (DNA) typing that is or are receiving current publicity. The following appeared in an article discussing a recent trial in which "DNA prints" were used [*6*]:

> The frequency of Andrews's pattern was one in 10 billion. "In a world population of just over five billion, he's the only guy who could have left his semen there," says Baird.

Considering only the DNA pattern, the conclusion does not logically follow from the frequency given. The chance of the pattern being duplicated in the world population (provided that the frequency figure is correct) is about 0.23 (23%).

When an expert offers an opinion that the evidence was definitely left by a specific individual, that expert should essentially be saying that it is his or her opinion that *nobody else* out there (however "out there" is defined) could have left it. This cannot be inferred from the figures in the above examples. If scientists can confuse frequencies of occurrence with probabilities of duplication, how can juries be expected to make sense of them?

In 1968, the California Supreme Court reversed a judgment (of guilty) in People vs Collins [*7*]. A main element of error cited was the presentation by the prosecution of a frequency of occurrence (1 in 12 million) as essentially a probability of duplication. This was considered as being prejudicial to the accused. The decision pointed out that the actual probability of duplication was at least 0.4 (based on the population of the Los Angeles area at the time of the case). Although the figures in the case were not presented with respect to a scientific examination of physical evidence, the principles are the same.

The above considerations strongly suggest that it would not be appropriate to simply use a population frequency and present it to a jury without further (and valid) interpretation. It would be unreasonable to expect a jury to be able to make a logical evaluation of the frequency without further guidance. Neither can an expert make a logical evaluation of the significance of the evidence without considering the probability of duplication or some related value, In order to do this, it is necessary to consider some target population.

Example Application

Let us continue by looking at a simplified example of a situation and the effect of a forensic expert's choice of target population. Assume a trial where there are two types of evidence supporting the prosecution. First, an eyewitness claims to have seen the suspect in the location of a crime at or about the time that the crime was committed. Since the area is remotely located, there would only be a few people there at any time (roughly 10 to 20). The suspect claims to have been out of town at the time, but can produce no witnesses to support that assertion. Second, some blood was found at the scene that matches the suspect's blood with respect to certain genetic markers.

The probability of occurrence of the blood type can be estimated from tables. Assume that this probability is such that an expert would be willing to state that the blood at the scene came from the suspect (in the expert's opinion) if the population of people that were in the crime scene area (say *P*1) is considered as the target population, but only that the blood is consistent with having come from the suspect if the target population is that of some larger population (say *P*2, estimated to be about 2 000 000) such as that of the city, state, region or whatever.

Assume the expert decides to use *P*1 (which of course requires the belief that the suspect is a member of *P*1) and testifies that the blood came from the suspect. Upon hearing this, the jury (or other trier of fact) would probably factor this into their evaluation and conclude, that, since the blood came from the suspect, the suspect must have been at the scene. This would then give credence to the eyewitness testimony, which seems to leave little doubt as to the final conclusion.

On the other hand, suppose the expert decides to use *P*2 and testifies that the evidence is consistent with the blood having come from the suspect, but states that about 1000 persons in the target population could be expected to have blood with the same characteristics. Now the jury must carefully evaluate the significance of the eyewitness testimony and the suspect's response thereto, and the outcome is not clear.

The first decision (that is, to use *P*1 as the target population) introduces a circular reasoning process into the consideration of the evidence. From one point of view, we have "if A then B" (if the blood came from the suspect, then logically the suspect must have been at the scene), and from another point of view we have "if B then A" (if the suspect was at the scene, then the probability associated with the matching blood characteristics leads to the conclusion that the blood came from the suspect). Providing an opinion about A is the expert's function, and making a decision about B is the jury's function. The jury is acting properly to consider the expert's opinion about A in arriving at a decision about B, but the expert is not acting properly in making a decision about the significance of A in the case by considering B.

The second decision (to use *P*2) is clearly the proper one for the expert to make in this situation. The expert must assume the jury's function in order to make the first decision, which would be stepping outside of the proper role for an expert witness. It does seem, however, that the second decision leaves a somewhat incomplete picture of the significance of the blood evidence. The jury could be made aware of the firmer interpretation of the blood match if they otherwise conclude that the suspect was actually at the scene at the time of the crime. This could be done by means of a hypothetical question posed to the expert by the attorney presenting the case. Whether this is appropriate depends upon the particular case, but is worth considering in order to convey a more complete picture of the significance of the involved physical evidence to a jury or other trier of fact.

This author was involved in a case somewhat similar to the above example. Some physical evidence was present that (in this author's opinion) could not be conclusively related to the suspect, although it was certainly fairly good evidence. After discussion of the case with the expert who was examining the evidence for the prosecution, it seemed that there was an agreement on the evaluation of the evidence, and therefore presumably on roughly equivalent target populations. There seemed to be no reason for two experts to appear in court to say essentially the same thing. However, the prosecution expert testified that the evidence was definitely linked to the suspect. The reason for this apparent change of opinion was never clear; it may have been based on other evidence that was made known to the expert subsequent to the discussion. Thus this writer was called in to challenge this, and it was not difficult to explain to the jury why the evidence could not on its own definitely point to the suspect. The prosecutor made no attempt to explore the issue further on cross-examination. The suspect was found not guilty.

This case left the feeling that neither testimony presented a proper viewpoint of the value of the physical evidence to the jury. If there were no other evidence (nonphysical), then the result was probably appropriate. However, there must have been some evidence that led the police to the suspect. If so, the value of the physical evidence may have been made clearer had the expert for the prosecution testified as to the limited value of the evidence on its own, and then responded to a hypothetical question that assumed certain facts that might logically be concluded from the other evidence.

Deciding what the appropriate target population (that is, *P2*) should be is not easy. The population of the world is always proper, but not necessarily always appropriate. If a crime occurs on a flying aircraft, a remote island, a ship at sea, or some similar situation in which we have an isolated and well-defined population, then that population would generally be the appropriate one to use. (It should be noted that if we are dealing with a small, well-defined target population, and if every member of that population can be examined for matching physical evidence, then a probabilistic analysis as discussed in this paper may not be necessary.) If the evidence is a trace fingerprint, then the world population is appropriate, considering the nature of and historical use of fingerprints as evidence. But when physical evidence such as hair or blood is considered, using the world population would appear to generally lead to a gross underevaluation of the significance of such evidence.

Target Population Selection Guidelines

The selection of an appropriate target population by a forensic science expert for evaluating evidence such as hair or blood should be done with certain guidelines in mind:

1. It should be known that the suspect is actually a member of the target population and that some member or members of the target population actually committed the crime, or at least the probability thereof should be *extremely* high.
2. The knowledge or probability in (1) should be arrived at by means other than evidence relating to the actual commission of the crime which is to be evaluated by the jury (or other experts). (Note: Evidence relating to the actual commission of a crime which is considered factual [that is, will be presented to the jury as fact rather than as evidence to be evaluated as to whether or not it is indeed fact] would seem to be acceptable as part of the criteria for the target population.)
3. The target population must be one for which the probability of chance occurrence (and thus the probability of duplication) of the evidence being evaluated by the expert can be estimated.

Let us consider two lines of reasoning that implicitly define a target population and analyze them with respect to the above guidelines. They will probably seem out of line to most readers, but have been seriously expressed by practicing forensic scientists (although not necessarily in court or relating to a case that they were working on).

> "The physical evidence is consistent with the suspect having committed the crime. I know the guy is guilty. The blood at the scene matches the suspect's. In my opinion, the blood at the scene came from the suspect."

The implicit target population is the suspect. Clearly a violation of Guideline 2. Note the implicit backward reasoning: if he is guilty, and since the blood matches, then he must have left the blood. This example shows a clear and direct influence of the backward reasoning. The same reasoning can be significant in guiding an opinion in far more subtle ways.

> "There were reasons the police arrested the suspect; they didn't just pick him at random. (Note: this is not a situation where we are dealing with a small, isolated group.) Considering that, and the probabilities involved with the physical evidence, it is my opinion that the physical evidence is definitely associated with the suspect."

The implicit target population is either the group of people arrested by the police for the crime (presumably only the suspect), or the group of people that were or would be specifically investigated as possible suspects with respect to the crime. If the police used evidence that is not sufficient for them to present in court to generate a suspect group, then this is a violation of Guideline 1. If the police used evidence that will be presented in court for evaluation by the jury, it is a violation of Guideline 2. If, however, we are dealing with a small, isolated population (for example, a crime on an aircraft in flight), then it is possible for the persons on the aircraft to be both a suspect group and a target population without violating the guidelines.

Selecting an appropriate target population for an evaluation of the significance of physical evidence by the forensic science expert depends upon the nature of the crime or incident. For many, if not most, crimes it is clear that the person or persons responsible had to be at a certain place within a certain period of time. We can therefore exclude anyone who could not have been in the area within that time period. Most persons in other countries or in distant cities or states are generally easily excluded. As the distance to the scene decreases, exclusion becomes more difficult for an increasing number of individuals. Once a boundary (not necessarily a geographical one) is selected, those persons not excluded become the target population.

For many crimes it may be appropriate to select the boundary to exclude all but the population of the area where the crime was committed (such as a city or county), plus the size of the population that routinely (or nonroutinely) came into the area (such as for work or recreation) during the time period in question. This could result in a target population size of several million for a large city (such as New York City) or only a few hundred for a small rural area. Shrinking the boundary further, such as to exclude all persons not in a specific section of a city, may be appropriate depending upon the nature of the crime and the area. However the boundary is set, care must be taken to insure that a reasonable estimate (that is, within an order of magnitude) of the number of persons who cannot be readily excluded (that is, the target population) can be made.

Once a target population (of size N) has been defined and the probability of chance occurrence of the physical evidence within that population has been determined, the forensic science expert can formulate an opinion on the significance of the evidence with respect to the target population. To do this, it is necessary to calculate the probability of duplication of the properties within the population. If p is the probability of chance occurrence (frequency), then the probability of duplication conditional on the known occurrence of the evidence can be calculated by the following formula:

$$Pr(Dup) = 1 - [(N * p * (1 - p)^{N-1}) / (1 - (1 - p)^{N})]$$

where * signifies multiplication, and / signifies division. The derivation of this formula is shown in Ref. *7*.

Pr(*Dup*) can be interpreted as essentially the probability of being wrong when stating an opinion that, based upon the properties examined, no other source (person) could have left the physical evidence under consideration. For instance, assume that *Pr*(*Dup*) is 0.25. This means that the particular evidence could be expected to have more than one source in 25% of the cases in which the evidence has the same frequency of occurrence. Therefore, if an expert consistently formed an opinion that there could not be any other source of the evidence in such cases, that expert would be expected to be wrong 25% of the time.

The following can be used as a rough guide to evaluating the significance of a frequency in a particular case (based upon Formula 9 in Ref *2*). If the expert is willing to accept an error rate of 1 error out of 10^d decisions (($d = 1, 2, 3, \ldots$), then the probability (p) of chance occurrence of the physical evidence should be about $1/(N \times 10^d)$ or less. Thus if the expert is willing to accept a risk of being wrong on the average once every 1000 cases, and the target population is 10 000, then the probability of chance occurrence should be less than or equal to 1/10 000 000 for an opinion that the evidence is definitely associated with a suspect to be formed.

Except for evidence having complex physical patterns (such as fingerprints or possibly patterns reflecting DNA polymorphism), or for very small target populations, physical evidence (such as blood or hair) is unlikely to have a probability of chance occurrence of the properties examined that is small enough for positive opinions of origin or relationship to an individual to be expressed by an expert. When this is the case, it may be worthwhile to consider in advance of a trial (but after the initial evaluation of the evidence) what target population might be defined by other evidence in the case and to formulate an opinion based thereupon which can be expressed in response to an appropriate line of questioning by the attorney or to a hypothetical question posed after the expert's opinion about the physical evidence that he or she has examined is presented.

Conclusion

To summarize, a distinction has to be made between the evaluation of the significance of physical evidence based upon the properties of that evidence alone and its evaluation with respect to other evidence in a case or situation. The primary function of the expert witness is to present an evaluation based only upon the evidence examined. That evaluation must consider both the probability of chance occurrence of the properties examined and an appropriate target population. However, it seems reasonable for the attorney to pose questions to the expert that are designed to offer guidance to a jury on the evaluation of the physical evidence within the context of different interpretations that they might make about other evidence in the case. The trier of fact (jury or judge) is the appropriate one to consider all of the evidence in rendering an opinion.

References

[*1*] Kingston, C., "Applications of Probability Theory in Criminalistics," *Journal of the American Statistical Association*, Vol. 60, March 1965, pp. 70–80.
[*2*] Kingston, C., "Applications of Probability Theory in Criminalistics-II," *Journal of the American Statistical Association*, Vol. 60, Dec. 1965, pp. 1028–1034.
[*3*] Thorwald, J., *Crime and Science*, Harcourt, Brace & World, Inc., New York, 1966, p. 435.
[*4*] Perkons, A. K., Hummel, R. L., and Jarvis, R. E., "Computer Analysis of Complex Gamma-Ray Spectra," presented at the Second National Meeting of the Society for Applied Spectroscopy, San Diego, CA, 14–18 Oct. 1963.
[*5*] Coleman, R. F., Cripps, F. H., Stimson, A., and Scott, H. D., "The Determination of Trace Elements in Human Hair by Neutron Activation and the Application to Forensic Science," Atomic Weapons Research Establishment Report No. O-86/66, United Kingdom Atomic Energy Authority, Jan. 1967.
[*6*] Lewis, R., "DNA Fingerprints Witness for the Prosecution," *Discover*, June 1988, p. 50.
[*7*] *People v. Collins*, 66 Cal.Rptr. 497 as reported in 138 P.2d 33.

Address requests for reprints or additional information to
Charles Kingston
John Jay College of Criminal Justice
445 W. 59th St.
New York City, NY 10019