NATIONAL COMMISSION ON FORENSIC SCIENCE

National Institute of Standards and Technology
U.S. Department of Commerce

Views of the Commission
Statistical Statements in Forensic Testimony

Subcommittee
Reporting and Testimony
Status
Initial Draft

Date of Current Version	13/11/16
Approved by Subcommittee	16/11/16
Approved by Commission	[*dd/mm/yy*]

Note: This document reflects the views of the National Commission on Forensic Science and does not necessarily represent the views of the Department of Justice or the National Institute of Standards and Technology. The portion of the document directly labeled "Views of The Commission" represents the formal Views of the Commission. Information beyond that section is provided for context. Views documents do not request specific action by the Attorney General, and thus do not require further action by the Department of Justice upon their approval by the Commission. The National Commission on Forensic Science is a Federal Advisory Committee established by the Department of Justice. For more information, please visit: https://www.justice.gov/ncfs.

Overview

This Views document presents background information and views on the following question: When experts present the results of forensic science and medicine examinations, tests, or measurements in reports or testimony, what types of quantitative or qualitative statements should they provide to indicate the accuracy of measurements or observations and the significance of these findings? This document refers to such statements as "statistical statements." These statistical statements may describe measurement accuracy (or conversely, measurement uncertainty), weight of evidence (the extent to which measurements or observations support particular conclusions), or the probability or certainty of the conclusions themselves. Such statements occur with many types of forensic evidence. Five examples follow.

1. *Pattern and impression evidence*. A shoe print found in the dirt next to the deceased is compared with a print made from the suspect's sneaker. What degree of similarity exists between the two impressions, and how strongly does it support a claim that the suspect's sneaker is, or is not, the source of the print in the dirt?

2. *Trace evidence*. A burglar smashed a pane of window glass to open a window. Various physical and chemical properties of a glass fragment collected from the suspect's clothing and a fragment known to be from the pane are measured. How accurate are the measurements, and how strongly does the similarity between the two sets of measurements support or refute the claim that the broken pane is the source of the fragment on the suspect as opposed to a claim that the fragments are from a different source?

3. *Qualitative analysis*. An oily substance is analyzed with thin layer chromatography to ascertain whether it contains tetrahydracannibinol (THC) and hence is liquid cannabis. How sensitive and specific is the procedure for detecting THC?

4. *Quantitative analysis with extrapolation*. A blood sample is collected from a driver 2 hours after an accident. A chemical test indicates a blood alcohol concentration (BAC) of 0.04%, which is below the 0.08% legal limit. What is the standard error of measurement? How likely is it that the true BAC at the time of the accident was above 0.08%, given that the measurement was 0.04% 2 hours later? What is the probability of observing 0.04% at the time of the accident if the true BAC at that time was below the legal limit?

5. *Cause, manner, and time of death*. An autopsy reveals injuries believed to be indicators of child abuse. Can we infer that child abuse had occurred and caused the injuries and death?[1]

These statistical statements and those that appear in connection with many other forms of evidence should be based on: (1) the existence of a relevant database describing characteristics, images, observed data, or experimental results; (2) a statistical model that accurately assesses the strength of the inference in question or describes the process that gives rise to the data linked to the question at hand; (3) information on variability and errors in measurements or in statistics or inferences derived from measurements; and (4) a statistical statement regarding the probative value of any comparisons done or calculations performed (e.g., how rare is an observed positive association when two items arise from the same source and when they arise from different sources?).

For many types of evidence, forensic science experts may not currently be making statistical assessments explicitly, but they may nevertheless be presenting their findings in a manner that connotes a statistical assessment. For example, the statement that "the latent print comes from the defendant's thumb" or "it is unlikely that the print came from anyone else" suggests a high probability that the print came from the defendant as determined by an understanding of the frequencies of similar features in fingerprints from the same individual and in prints from different individuals. As forensic science moves forward, the Commission anticipates efforts to make the presentation of analyses more overtly statistical and quantitative, and it is toward this end that the Commission expresses its views. The Commission advances a preference for an explicit statistical foundation for statements because a mathematical analysis tends to provide a ready means for assessing and expressing uncertainty.

[1] Inferring causation is usually a statistical activity. Scientists usually infer from a treatment or cause to the outcome or effect, and one of the strongest forms of evidence to support such inference comes from randomized experiment. In the present context, we are in effect reversing that process and attempting to infer the cause from the effect, as more often is the case in the context of the law.

The National Research Council Committee on Identifying the Needs of the Forensic Sciences Community emphasized the importance of describing uncertainties in measurements and inferences.[2] Statistics is concerned with the study of variability, uncertainty, and decision-making in the face of uncertainty. It supplies a set of principles, based on probability, for drawing conclusions from data and for expressing the risks of certain types of errors in measurements and conclusions. This framework applies throughout forensic science and medicine, but the discussion that follows is of special relevance to pattern, impression, and trace evidence.

Trace, impression, or pattern evidence examiners should follow a valid and reliable process to determine whether there is a positive or negative association between the item in question (often called a "questioned" sample or specimen) and a sample whose source is known (such as a reference sample from the defendant). Reliability and external validity should be established via scientific studies that have been the subject of independent scientific scrutiny.[3] Only when the reliability and validity of the process have been studied quantitatively can a probabilistic or statistical model for indicating the uncertainty in measurements and inferences be credible.

Such models are most convincing when a scientific understanding of the physical process that generates the features exists. Sufficient knowledge of the process permits a mathematical model to be developed. This approach has been successful for determining the probability that associations in particular DNA features will exist among different individuals. For other types of trace and pattern evidence, however, no widely accepted probabilistic models of the phenomena that give rise to the features are available. Consequently, most efforts to provide probabilistic statements about features and their degree of association often rest on the personal impressions of examiners, supported by their subjective judgment developed through individual training and experience, or by reference to empirical studies of the reliability of the judgments of examiners. Training and experience are important in applying valid techniques, but they are not a sufficient basis for establishing the uncertainty in measurements or inferences.

When forensic examiners do provide a statistical statement—with or without a numerical articulation of probability, odds, or likelihoods—such a statement must be supported by an empirical assessment of the underlying statistical model. Statistical calculations used in judicial proceedings should be replicable, given the data and statistical model; however, when observations are largely subjective or when different statistical models are in use, the quantitative summary of the significance of the findings may vary from examiner to examiner and from laboratory to laboratory. Consequently, an essential element of an examiner's report is a statement of the measurements and the models to assist other experts in replicating the statistical quantities reported.

At the core of all of such statistical calculations, there must be data from a *relevant population or sampling*. Impressions of the soles of shoes gathered in Israel may not be relevant to a case

[2] Comm. on Identifying the Needs of the Forensic Sci. Cmty., Nat'l Research Council, Strengthening Forensic Science in the United States: A Path Forward 184 (2009).

[3] Nat'l Comm'n on Forensic Science, Views Document on Technical Merit Evaluation of Forensic Science Methods and Practices, June 21, 2016, https://www.justice.gov/ncfs/file/881796/download.

involving a shoe print in Alaska. If a clear population of relevance can be defined, then one needs to consider the extent to which the actual data in the database represent the population. To be applicable to casework, empirical studies of the reliability and accuracy of examiners' judgments must involve materials and comparisons that are representative of actual cases and rely on data from a relevant population or sample base.

In comparing forensic evidence recovered from crime scenes or on victims or suspects with known samples, the forensic examiner primarily focuses on ascertaining corresponding features and, traditionally, in deciding whether there is a positive association (often referred to as a "match," an "inclusion," or "consistent" or "indistinguishable" features) or a negative one (an "exclusion" or inconsistency) to the known sample. But a "positive association" is not probative unless it is more probable when the items have a common source than when they originate from different sources. Indicating the statistical weight of the positive association therefore requires a statement of how common or rare the association is, based on a database linked to the case at hand. For example, a positive association for the presence or absence of pigment in a hair cuticle is some evidence that the hairs have a common origin, but the significance of this association is unknown without data from relevant populations.

More generally, when dealing with features, such as the refractive index of glass or the heights of peaks in an electropherogram of DNA fragments, that have more values than "absent" and "present," the classification of "matching" and "not matching" omits statistical information related to the degree of similarity. The weight that should be given to any degree of association depends on (1) the probability of the degree of correspondence in the features, given that the samples came from the same source, and (2) the probability for the same measurement, given that the samples came from different sources. When the former probability is larger than the latter—when the observed degree of similarity occurs much more often for same-source samples than for different-source samples—the evidence supports the conclusion of a common source.

Any recommendation on presenting explicit probabilities, however derived for specific forensic evidence, might distinguish between probabilities based on some statistical model and ones that characterize the examiner's subjective sense of how probable the evidence is under alternative hypotheses. The latter are difficult to validate, but it also must be understood that statistical models are approximations, and, inevitably, there is some uncertainty in the selection of a model. Furthermore, the statistical model and method used to analyze the evidence do not always admit naturally to the simple form of the likelihood ratio favored in the forensic science literature. Some statistical problems, especially those focused on issues of causation, may not involve source comparisons leading to likelihood ratios. In light of the limitations on both statistical modeling and more intuitive judgments of the significance of similarities, we offer the following views on the presentation of forensic science findings:

Views of the Commission

It is the view of the Commission that:

1. Forensic experts, both in their reports and in testimony, should present and describe the features of the questioned and known samples (the data), and similarities and differences in those features as well as the process used to arrive at determining them. The presentation should include statements of the limitations and uncertainties in the measurements or observations.

2. No one form of statistical calculation or statement is most appropriate to all forensic evidence comparisons or other inference tasks. Thus, the expert needs to be able to support, as part of a report and in testimony, the choice used in the specific analysis carried out and the assumptions on which it was based. When the statistical calculation relies on a specific database, the report should make clear which one and its relevance for the case at hand.

3. The expert should report the limitations and uncertainty associated with measurements and the inferences that could be drawn from them. This report might take the form of an interval for an estimated value, or of separate statements regarding errors and uncertainties associated with the analysis of the evidence. If the expert has no information on sources of error in measurements and inferences, the expert must state this fact.

4. Forensic science experts should not state that a specific individual or object is the source of the forensic science evidence and should make it clear that, even in circumstances involving extremely strong statistical evidence, it is possible that other individuals or objects could possess or have left a similar set of observed features.[4] Forensic science experts should confine their evaluative statements to the support that the findings provide for the claim linked to the forensic evidence.

5. To explain the value of the data in addressing claims as to the source of a questioned sample, forensic examiners may:

   A. Refer to relative frequencies of individual features in a sample of individuals or objects in a relevant population (as sampled and then represented in a reference database). The examiner should note the uncertainties in these frequencies as estimates of the frequencies of particular features in the population.

   B. Present estimates of the relative frequency of an observed combination of features in a relevant population based on a probabilistic model that is well grounded in theory and data. The model may relate the probability of the combination to the probabilities of individual features.

   C. Present probabilities (or ratios of probabilities) of the observed features under different claims as to the origin of the questioned sample. The examiner should note the uncertainties in any such values.

---

[4] Similarly, to avoid implying that a statistical foundation exists when there is no statistical model or method and related database to properly characterize the evidence, forensic experts should not use phrases such as "to a reasonable degree of scientific certainty." Nat'l Comm'n on Forensic Science, Views Document on Use of the Term "Reasonable Scientific Certainty," Mar. 22, 2016, https://www.justice.gov/ncfs/file/839731/download.

D. When the statistical statement is derived from an automated computer-based system for making classifications, present not only the classification but also the operating characteristics of the system (the sensitivity and specificity of the system as established in relevant experiments using data from a relevant population). If the expert has no information or limited information about such operating characteristics, the expert must state this fact.

6. Not all forensic subdisciplines currently can support a probabilistic or statistical statement. There may still be value to the factfinder in learning whatever comparisons the expert in those subdisciplines has carried out. But the absence of models and empirical evidence needs to be expressed both in testimony and written reports.