



### B26 Random Match Probabilities and Database Search Estimates Provide Different Answers for Different Questions

Bruce Budowle, PhD\*, and F. Samuel Baechtel, PhD, FBI Laboratory, 2501 Investigation Parkway, Quantico, VA 22135

After attending this presentation, attendees will learn how forensic scientists might better articulate the bases for statistical estimates of DNA profile frequencies: what are the proper questions to be answered, and what approach best addresses a particular question.

This presentation will impact the forensic community and/or humanity by demonstrating recent misrepresentations that have led to confusion in the legal arena regarding the best approaches for estimating the rarity of a DNA profile when the suspect is identified first by a database search. This presentation will bring clarity to this issue, demonstrate that there is general acceptance of current practices, and point out that the perceived debate is nothing more than an application of wrong answers to proper questions.

Recent misrepresentations have led to confusion in the legal arena on the best approaches for estimating the rarity of a DNA profile when the suspect is identified first by a database search. This presentation will bring clarity to this issue, demonstrate that there is general acceptance of current practices, and point out that the perceived debate is nothing more than an application of wrong answers to proper questions.

When a comparison of DNA profiles derived from evidence and reference samples fails to exclude an individual as a contributor of the evidence sample, the weight of the evidence is determined using a statistical assessment. For forensic applications, it is important that the statistical conclusions be conveyed meaningfully. To derive appropriate statistical inferences the question to be answered must be properly formulated. One particularly useful question for the fact finder is how common or rare is an evidence profile (calculated by either the random match probability or by using the likelihood ratio). Because evidentiary profiles are routinely being searched for matching profiles in felon databases another question may be raised: What is the probability of finding the DNA profile in the database searched? This latter question could have investigative value and addresses a different issue than that of the rarity of the profile. The scientific bases for the estimates for each question are the same. That is the profile frequencies can be estimated by multiplying allele frequencies and correcting for sub-structure and sampling error. There is little dispute today about such fundamental practices. Recent court deliberations (e.g., U.S. v Jenkins 2005), there has been confusion regarding an answer to the question of profile rarity with the database statistical search estimate. The NRC II Report (1996) advocates using the formula  $1/(N p_x)$ , where  $N$  is the size of the database and  $p_x$  is the random match probability, for the database search estimate. Proponents of applying this calculation as the true random match probability erroneously cite the language of the NRC II Report for supporting their contention. The report written by the NRC committee must be read in its entirety to appreciate the proper application. Clearly, on page 40 (Recommendation 5.1) the report defines the proper question as: "If one wishes to describe the impact of the DNA evidence *under the hypothesis* (Italics added) that the source of the evidence sample is someone in the database, then the likelihood ratio should be divided by  $N$ ." Thus, the above formula was never meant to supersede the random match probability estimate. It should also be obvious that the different questions produce different answers and should not be construed as a conflict. In reality, there is no issue regarding general acceptance of the random match probability approach.

Another approach to contest the use of the random match probability is to focus on differences of opinion on how best to calculate the database search estimate. An alternate treatment to that of the NRC II report suggests that the evidentiary weight is underestimated. Again, this estimate does not address the question of how rare is the profile, and thus a debate on what questions to ask becomes an interesting academic exercise for some.

To appreciate better the subtleties of these various positions, examples will be provided to demonstrate the erroneous practice of proffering incorrect answers to meaningful forensic questions. These include the approaches already described by the DNA Advisory Board (Statistical and Population Genetics Issues Affecting the Evaluation of the Frequency of Occurrence of DNA Profiles Calculated From Pertinent Population Database(s), Forensic Science Communication, July 2000 Volume 2 Number 3) and other simple models.

This presentation will help forensic scientists articulate the bases for statistical estimates of DNA profile frequencies, what are the proper questions to be answered, and what approach best addresses a particular question.

**Statistics, Database Search, Random Match Probability**