E34   Cold Hit Statistics and Database Access

Bicka Barlow, JD*, Office of the Public Defender, 555 7th Street, San Francisco, CA 94103

After attending this presentation, attendees will learn about why cold hit DNA cases require a different statistical analysis. Additionally, disclosure of CODIS database information will be discussed and attendees will learn more about privacy issues and the statistical analysis that can be done on the data.

This presentation will impact the forensic science community by challenging some long held beliefs regarding the use of statistics in cold hit DNA cases.

Even though both reports from the National Research Council on DNA evidence (known as NRC 1 and NRC 2) propose two statistical methods that should be applied to matches found through a database search, the forensic community has ignored these recommendations. Forensic labs in the United States have continued to use the Random Match Probability (RMP) statistic in cold hit cases. A review of academic journals shows that there is a split in the statistical community regarding what is the appropriate statistic method. However, all sides agree that the RMP is not the correct statistic because it does not take into account the manner in which the defendant was identified.

The purpose of applying a statistical analysis to matches is to provide a jury with a way to give weight to the match. The RMP provides a measure of the probability of a coincidental match, meaning that the defendant did not leave the biological material but matches purely by coincidence. Many people - including DNA analysts - misunderstand this number to mean the probability that anyone besides the defendant left the DNA. This is called the prosecutor's fallacy and a case addressing this issue has recently been taken up by the U.S. Supreme Court.

The discussion around the appropriate statistic to be used in cold cases centers on how one measures the probability of a coincidental match given the manner in which the match is made. Unlike in "probable cause" cases where the defendant is already a suspect, in a cold hit case the defendant only becomes a suspect based on a match. There are three groups in this debate, only two of which will be described in detail. All alter the RMP statistic to account for ascertainment bias.

One camp adopts the NRC 1 approach. This method uses the RMP but only on loci that are not used to identify the defendant in the first place. For instance, the database search would be done with nine loci and once a possible suspect is identified additional testing would be done and the RMP applied only to those additional loci. Although there are 13 core loci that are routinely used by labs in the CODIS system, additional loci have been identified and are in use that allow testing up to 15 loci.

Another camp adopts the NRC 2 approach. This method also uses the RMP but then takes into account the number of people who were searched in the database. So the RMP for the evidence profile is 1 in 1 million and 300,000 profiles are searched to find one match, the NRC 2 number would be one in three. In other words, the probability of choosing a person who matches the crime scene evidence from a database of innocent people, is one in three.

The third approach is Bayesian. The statistic applied is approximately the same as the RMP.

Many statisticians believe that one way to determine the appropriate measure would be to study the CODIS databases. Due to the huge size of these databases, the statistical value of the research would be much higher than any previous study. Questions relating to the heterogeneous nature of the database which include mixed racial groups and relatives can be addressed through statistical modeling.

Privacy issues also are extremely low. No individual information is necessary and the data could be treated in the same manner as all human data used for research is handled.

Cold Hit, Statistics, CODIS Database