

Rounding Up the Usual Suspects

David H. Kaye

AALS Evidence Section Newsletter

Spring / Summer 2008, pp. 2–4

Countries around the world have established databases consisting of the DNA profiles of suspected or convicted offenders. In the United States, state and federal databases combined in the FBI's National DNA Index System hold over five million convicted offender DNA profiles as well as those of some people who are merely arrested or detained. These identification databases have helped solve cases that have baffled investigators for decades. In one case, a federal database search linked a 58-year-old man suspected of raping at least 25 women in three states to semen on underwear from a 1973 rape.

When the DNA profile from a crime-scene stain matches one of those on file, the person identified by this “cold hit” will become a target of the investigation. A fresh sample will be taken from the suspect to verify the DNA match, and other evidence normally will reinforce the investigatory lead. In rare cases, prosecutors will even proceed with no other evidence. In one such case, a San Francisco jury convicted John Davis, already behind bars for robbery and other crimes, of the murder of his neighbor, Barbara Martz, nearly 22 years [-----Page 3-----] earlier. The database match was all the jurors had to go on. This was enough for a conviction, at least where the probability that a randomly selected, unrelated individual would match the crime-scene DNA sample — the “random-match probability” — was said to be “quadrillions-to-one.” Cases like *Davis* that emanate from cold hits have been called “trawl cases” because “the DNA match itself made the defendant a suspect, and the match was discovered only by searching through a database of previously obtained DNA samples.” Peter Donnelly & Richard D. Friedman, *DNA Database Searches and the Legal Consumption of Scientific Evidence*, 97 Mich. L. Rev. 931, 932 (1999). These database-trawl cases can be contrasted with traditional “confirmation cases” in which “other evidence has made the defendant a suspect and so warranted testing his DNA.” *Id.* In terms of this dichotomy, we must ask whether the fact that the defendant was selected for prosecution by trawling requires some adjustment to the random-match probability. Two committees of the National Academy of Sciences (NAS) thought so. In their influential reports on “DNA Forensic Science,” they reasoned that a match coming from a trawl is much less impressive than a match in a confirmation case — just as finding a tasty apple on the very first bite is more impressive than pawing through the whole barrel of apples to locate a succulent one. To account for the extra bites at the apples, they described approaches that would inflate the normal random-match probability.

The response has been disputation and litigation. Two early commentators, Bill Thompson and Simon Ford, gave “a Bayesian analysis” to suggest that “this evidence has no probative value.” William C. Thompson & Simon Ford, *DNA Typing: Acceptance and Weight of the New Genetic Identification Tests*, 75 Va. L. Rev. 45, 100 (1989). Ten years later, Donnelly and Friedman reached precisely the opposite conclusion. They applied Bayes' rule in more detail -- and correctly -- to show that the trawl actually increases the probative value of the match.

Recently, three appellate opinions on the issue have emerged -- *United States v. Jenkins*, 887 A.2d 1013 (D.C. 2005), *People v. Johnson*, 43 Cal.Rptr.3d 587 (Ct. App. 2006), and *People v. Nelson*, 48 Cal.Rptr.3d 399 (Ct. App. 3 Dist. 2006), rev. granted, 147 P.3d 1011 (Cal. 2006).

In these cases, defendants argued that until the scientific community can agree on a single statistic to characterize the import of a database trawl, even the fact of a match should not be admitted.

Even though the dispute in the scientific community is limited to the question of whether there is any reason to bother with the NAS adjustments to the probability figure, the trial court in *Jenkins* felt compelled to exclude the DNA evidence in its entirety.

The appellate courts all rejected the defense challenges, but their opinions fail to address the dispute in the scientific and legal literature. The avoidance mechanisms they employ are singularly unimpressive. In *Nelson*, for example, the court of appeal claimed that California's general acceptance standard for scientific evidence does not apply because after the database trawl identifies the suspect, a fresh sample from the suspect is typed. If the fresh sample matches, only this match is introduced at trial. In the court's view, it is as if the database trawl never took place.

To a statistician, this is a jaw-dropping claim. The challenge is not to the use of a convicted-offender DNA database as an investigatory tool. The objection [-----Page 4-----] is to the use of the random-match probability at trial to gauge the power of the later match when the defendant has not been selected for DNA testing “at random” — that is to say, on the basis of factors that are uncorrelated with his DNA profile. When the defendant is selected for a later test precisely because of his known DNA profile, the replication adds no new information about the hypothesis that the defendant is unrelated to the actual perpetrator and just happens to have the matching DNA profile. It adds no information because the datum — a matching profile in the new sample — is just as probable when this hypothesis is true as when it is false. Replication helps eliminate the risk of a laboratory error in determining or reporting the DNA profile, but it has no further value in probing the possibility of a coincidental match.

Because the rationales presented in the three cases to date are unconvincing, the emerging case law needs to be reoriented to confront directly the competing statistical arguments about the meaning of a database match. Recent statistical literature seems to favor the view that no adjustment to the random-match probability is necessary, but this may just reflect the fact that most statisticians writing about forensic science are Bayesians rather than frequentists. Although Donnelly and Friedman have presented the Bayesian perspective forcefully and simply, it appears that it will take more to convince the courts that they need to think more deeply -- and more clearly -- about the subject.

© 2008 D.H. Kaye. These comments are adapted from a forthcoming book, *The Double Helix and the Law of Evidence* (Harvard Univ. Press).