The Honest Scientist's Guide to DNA Evidence

Richard O. Lempert
University of Michigan Law School, rlempert@umich.edu

Genetica **96:** 119–124, 1995.

The honest scientist's guide to DNA evidence

Received 18 May 1994 Accepted 22 July 1994

Dear Bruce,

Thank you for your invitation to participate in the DNA symposium. As you know DNA has never been a prime research focus of mine, and I have been so preoccupied with my own work on ITPT (intertemporal personal transportation) that I thought I must decline. Happily, however, the two projects came together, for I recently had an amazing breakthrough during which by coincidence I stumbled across a book entitled *A Century of DNA Testing* and holocopied (a fancy form of Xeroxing) the following few pages for you.

Even by 1996, before the second NRC Report, which fostered the switch to Bayesian methods, appeared and several years before the general abandonment of RFLP procedures, the statistical issues that had so roiled the early days of DNA testing had largely been resolved. This occurred not because the statistical and genetic disputes we describe above had all been resolved, but rather because the need to confirm identifications from burgeoning data bases stimulated several different innovations that increased the informativeness of the RFLP approach to DNA testing, thus reducing the chances of coincidental matches to minuscule proportions even when the suspect population included relatives or had the same narrow ethnic heritage. In one Minnesota case, for example, the state crime lab simply tested alleles at nine loci. But Minnesota at the time was laboring under a short-sighted (and soon overturned) court ruling which precluded attaching statistical probabilities to DNA matches, and thus, without resorting to statistics, the lab's scientists had to be able to testify that an identification was virtually certain.

Nevertheless, problems stemming from what Professor Junipurr, some fifty years ago, in his study of forensic laboratories called the 'infallibility complex' continued to appear. Some of the most important of these problems and ways that were suggested at the time for meeting them are captured in the following extracts from the original edition (Jack Point, ed., 1996) of the still essential handbook, *The Honest Scientist's Guide to DNA Evidence*. We can in this history do no better than reproduce the relevant portions, leaving the original spelling and grammar, including archaic gendered pronouns, untouched. The somewhat preachy nature of the original handbook was, we believe, influenced by the religious fundamentalism resurgent at that period. Also, though it may not be evident to readers today, many of the Handbook's suggestions were regarded at the time as quite Utopian.

Section 7.1: (Subjectivity)

The honest scientist recognizes that she herself is a test instrument, and a fallible one at that. Subjectivity inescapably enters into any human endeavor, and should not be denied. DNA testing is rife with subjective elements, no place more so than at the crucial stage of deciding whether a match exists (1). On the one hand, non-matching extraneous bands may sometimes be properly disregarded and patterns that do not quite meet objective matching criteria may be appropriately regarded as incriminatory matches. On the other hand, band patterns that do meet objective matching criteria may be treated as exonerative depending on how they deviate from perfect matches. The DNA expert should not hide behind the cloak of science to deny the role of human judgement. White coats should not be worn into the courtroom either literally or figuratively.

At the same time, the honest scientist tries to be as objective as possible in her judgements. She realizes that this is inconsistent with a strong *a priori* belief that the donor of a suspect sample is guilty. Thus she avoids any information suggesting the involvement or uninvolvement of the accused until after she has prepared her report and, in the ideal case, until after she has testified. Laboratories should cooperate to make this easy. Crime-related information should be stripped from all information sent to the analyst unless it is essential to the test and its interpretation (e.g. information that several people are suspected of participating in a rape or that the suspected rapist is the victim's brother).

Scrupulous laboratories should also realize that being continually identified with a side may also bias judgements. Thus, they should seek out business from defense attorneys who seek to exonerate clients and encourage their scientists to take the rare opportunities that come along to testify on behalf of defendants.

An additional and highly recommended way of coping with subjectivity is to transform DNA tests from true-false into multiple-choice tests by routinely providing scientists with three test samples. One or two of these, varying on some random schedule, should be the suspect's and the other(s) should come from one or two non-suspects. This marginally increases the cost of DNA tests, but we are dealing with decisions that can take human lives. Moreover, multiple testing carries the added benefit of providing laboratory-specific proficiency information.

Section 9.3: (Proficiency Testing—continued)

The honest scientist works for the Virtuous laboratory. The Virtuous laboratory maintains a rigorous system of proficiency testing, both to identify and correct sources of error and as an incentive for continual high quality staff work. Four types of samples may be analyzed as part of a proficiency testing program: known samples, unknown samples of pure quality, unknown samples of case quality, and apparent case samples. While these different types of samples all have a role to play in measuring and maintaining laboratory proficiency, only the last, test samples regarded by both the laboratory and analyst as true casework samples, provides a true picture of the quality of laboratory procedures, for only apparent case samples are certain to be treated by laboratories as if they were casework samples. Conversely, if a laboratory knows that any apparent case sample it analyzes may be a test, the incentive is to treat each sample submitted for analysis as if it were a test sample, and the quality of all the laboratory's work should increase.

It is, even today, difficult to generate convincing casework test samples, for such samples must appear to have come from specific police departments or prosecutors' offices, and any communications with the police or prosecutor's office during the course of analysis must be consistent with the otherwise apparent genuineness of the sample. For this reason, laboratories, police and prosecutors should agree to join the proposed APTLAB (Accredited Proficiency Tested Laboratory Analysis Bureau) organization. As proposed APTLAB would serve as an intermediary between parties seeking DNA tests and laboratories that do them. APTLAB would receive all DNA test samples from police departments or other agencies and after removing unnecessary case-related information that might accompany samples (see Section 7.1) would forward the material to the laboratory that the agency chose to conduct the test. Approximately one out of every eight samples forwarded to a laboratory, on some random schedule, would be a casework test sample. Although the problem of false positive matches has been the target of most concern, APTLAB, according to the proposal, would forward as many matching test samples as non-matching ones. Missing matches also reveals weaknesses or biases in laboratory procedures, and the error of mistakenly freeing an actual rapist or murderer may be as socially harmful as mistakenly convicting an innocent person. Only after test results were communicated to APTLAB would the testing laboratory receive the name of the submitting agency. Thereafter all communications would be between the laboratory and the agency. Statistics on error rates, aggregated across laboratories using similar procedures and also by laboratories, would be made freely available.

At the moment it does not appear that APTLAB will get off the ground, for the suggestion has met with fierce resistance from forensic scientists and laboratory directors. Three arguments are made against it. Least is made of the first argument, but we suspect it is the strongest of the stated motivations behind the heated objections. This is that APTLAB will interfere with the relations that many laboratories and forensic scientists have with the agencies and individuals that submit samples to them. While this is true, the honest scientist should regard this as a virtue of the APTLAB proposal. Relations between laboratories, scientists and police agencies today may include conversation about the heinousness of the crime, which can cause a scientist to overly-identify with the police function, and other evidence that incriminates the accused. Ordinarily this information is at best irrelevant to the scientific validity of the test, and it can bias judgements and so detract from it.

A second objection one often hears is that interposing an agency between the law enforcement body and the testing laboratory simply adds a source of error. This objection cannot be dismissed out of hand, but since all test samples will be sent by courier under seal to APTLAB and the seal will not be broken in the transmission process, the possibility of introducing additional error will be minimal.

The third objection concerns costs. This objection is substantial, for running a new agency, even a genuine non-profit agency costs money as does the proposed $12\frac{1}{2}\%$ increase in the number of DNA tests conducted. These costs will, of course, be reflected in the price of DNA tests and may decrease somewhat the propensity of policy agencies to order DNA tests. However, the cases where testing will be forgone will be most likely those in which evidence of guilt apart from DNA is overwhelming and DNA evidence is window dressing not needed for a conviction. The cost issue also appears different when one thinks in terms of total social costs. It can cost between $20,000 and $50,000 to keep a rapist in prison for a year and many times that to keep a person on death row and eventually execute him. This is totally apart from the costs to the wrongfully convicted man. Conversely, mistaken acquittals can lead to the rape or murder of future victims. Thus, if the APTLAB procedures reduce errors as expected, they will be cost justified. If after some experience with APTLAB, its procedures do not appear cost justified, the consortium can be terminated, and we can have some confidence that there is no problem here.

There are also objections to APTLAB that have not figured in the debate because no one dares make them publicly. The first is the fear which some scientists privately voice, that uncomfortably high laboratory error rates will be revealed, and the second is that APTLAB will displace several other organizations that in a less rigorous but nonetheless serious fashion have begun to develop proficiency testing programs. Perhaps for these reasons more than the three that have been articulated, objections to the APTLAB proposal have been so fierce in the forensic science community that absent legal mandate the best that can be hoped for in the way of proficiency testing may be procedures, like those that have been implemented in a number of conscientious laboratories, consistent with the recommendations for proficiency testing in the first NRC report. If so, it may be that the honest scientist will find that there is no place that quite meets her standards.

Section 13.2: (Statistics and Error—continued)

The honest scientist recognizes that no matter what the odds that a random man's DNA would match evidence DNA, the probative value of a reported match between the defendant's DNA and evidence DNA is always affected by the likelihood that a match will be mistakenly reported (2). This probability is greater than the likelihood that a test error might somehow mistakenly yield a match, for it includes the chance that either inadvertently or intentionally a second sample of evidence DNA has been substituted for or contaminated the DNA of the defendant in a way that will not be obvious in the testing (3). In testifying, the honest scientist should never attach more probative weight to the evidence that this probability, which we will call the false positive rate, reflects. In particular, the extremely low probabilities often associated with random matches are likely to mislead the trier of fact and should not be mentioned or, if mentioned, should be placed very carefully in context so that they are clearly subordinate to the false positive rate.

Such careful, limited testimony poses, however, a difficulty. Unscrupulous defense attorneys may seek to interpret false positive statistics so as to induce the jury to make what Thompson and Schumann (4) have called the 'defense attorney's fallacy.' The defense attorney's fallacy falsely presupposes that there is no reason to think the defendant is guilty apart from the identification evidence. Without stating this presupposition, defense attorneys argue that statistical identification evidence simply locates the defendant as one member of a large group who might have committed the crime, and so does little to finger the suspect.

In the example that follows, drawn from the transcript of the case called *State v. Evett*, the DNA expert is careful not to overweight the probability of a reported DNA match. The cross-examiner then tries to take advantage of this honest scientist by attempting to induce the defense attorney's fallacy. On redirect examination, this effort to mislead the jury is successfully countered. In a portion of the transcript we omit, the scientist testified that a false positive rate of one in a hundred, which she later uses, is a conservative estimate. She also explains that while ordinarily the police agencies she works with divide evidence DNA and send it to several laboratories for independent testing (thus justifying a far lower false positive estimate), division was impossible in this case because of the limited amount of available evidence DNA. To conserve space we have also eliminated other aspects of the direct, cross and redirect examinations that do not bear on the issue discussed in this section.

Direct examination

Prosecutor [P]:
When you conducted the analysis you described, what exactly were your findings?

Scientist [S]:
I found that, within the limits I have already mentioned, each of the seven alleles identified in the DNA extracted from the defendant's blood matched an allele found in the DNA extracted from the semen found in the victim's vagina.

P: Did you find any alleles that did not match?

S: No, every allele we identified in the DNA extracted from the defendant's blood corresponded in length to an allele extracted from the semen.

* * *

P: Can you give the jury some idea of how unusual it is to find the seven allele match you have reported?

S: If DNA tests were perfect, the chance that one would find DNA matching that in the evidence sample by testing one Caucasian man drawn randomly from the country's population is about one in fifty million. However, as with all human endeavors DNA tests are not perfect. To the best we can estimate, in about one test of every hundred we run a DNA match is erroneously reported due to some mix up or mistake in the field or in the laboratory. So the best way of thinking about this evidence is that there is about one chance in 100 that somewhere a mistake was made, and the defendant's DNA did not match the DNA extracted from the semen. But if we did not make a mistake, then there is only about one chance in fifty million that some random person, which is to say some white person not related to the defendant, could have left the evidence DNA. Ultimately, however, the former figure is more important than the latter, so I will summarize my testimony by saying that the defendant's DNA appeared to match the evidence DNA but there is about one chance in 100 that this conclusion is mistaken. If the defendant's DNA did match, it is very unlikely he is the victim of coincidence.

P: Thank you, no further questions.

Cross examination

Defense Counsel [DC]:
Now Dr., you said that about one time in 100 when you test for DNA you report a match as you did in this case, but that match turns out to be false—isn't that so?

S: Well, not exactly. The one in a hundred figure is a conservatively estimated error rate based on what are known as 'double blind' proficiency tests.

DC: But one in a hundred is your best estimate of the false positive rate, isn't that so?

S: Best conservative estimate.

DC: Now there are about two million Caucasian males living within 40 miles of where the crime occurred, isn't that so?

S: If you say so; I don't live here.

DC: Well you can take my word on it. Now if any of these men had been chosen at random and tested there is about a one in a hundred chance that you would be here telling me there was a match, isn't that so?

S: Yes, assuming they had been suspected as the defendant was.

DC: Dr. please just answer 'yes' or 'no.' Now it follows doesn't it that if all two million Caucasian men had been tested, there would be about 20,000 positive tests reported?

S: Well, it's more complicated that that.

DC: Dr. 1/100 times two million is 20,000 is it not?

S: Yes.

DC: So basically what your test establishes is that there is about one chance in 20,000 that my client is the culprit.

S: No, that's . . .

DC: Dr. I didn't make any mathematical mistakes, did I?
S: No.
DC: Thank you, no further questions.

Redirect examination

P: Dr. do you feel that the defense counsel was accurately stating the import of the match you reported, when she suggested that it meant that there was one chance in 20,000 that the defendant was the culprit?
S: No, I don't.
P: Can you explain why?
S: Yes, the defense counsel was trying to induce the jury to buy into what is known in the literature as the defense attorney's fallacy. It misleads because it assumes that there is no other evidence in the case.
P: Can you say more?
S: Yes, the defense counsel's calculations would not be misleading had the defendant's name simply been drawn from the phone book and had he then been required to undergo a DNA test. If the police drew 100 such names from the phone book and processed each individual as they would if he were the prime suspect, handling his DNA and the evidence DNA as they would that of a suspect, proficiency test results suggest that even if all the people tested were innocent, there is a good chance that we would mistakenly conclude in the case of one person that there was a DNA match. If we tested more people at random the chance of finding additional false positive results would increase, with the expectation being that one such result would occur in every hundred people tested, provided the DNA and tests were handled exactly as they are when actual suspects are tested.

But this defendant's name was not drawn at random from a phone book. Defense counsel's fundamental error is to ignore this fact; the defendant was not selected for the DNA test at random. Instead there was a reason the defendant's DNA was tested. As I understand the state's evidence in this case . . .

DC: Objection.Your Honor, the witness's understanding of the evidence is irrelevant and interferes with the province of the jury.

Judge [J]

Were you going to interpret the evidence in this case Dr. or just list what has been introduced?

S: I was just going to list it, Your Honor.

J: In that case you may proceed. Please do not give your opinion of the evidence or mention any evidence the state has not introduced, for if you do I shall have to sustain defense counsel's objection.

S: As I was saying the defendant was not just a random person selected for testing. In this case the defendant was picked up because the last three digits on his license plate matched the information provided by a woman who witnessed the abduction, the victim testified that though the defendant had worn a ski mask, she recognized his eyes and his voice, when he was arrested the defendant had scratches on his face, which he refused to explain to the police, and a hair was found in the defendant's car consistent with hair that might have come from the victim. In these circumstances, even with a false positive rate of one in one hundred, the evidence of a DNA match I presented is powerful confirmatory evidence. Had the defendant been innocent despite the other evidence against him 99 times out of 100 the DNA test would have demonstrated his innocence. If his alibi were true, the DNA test would fail to confirm it only one time in a hundred.

Another way to think about this is to think about the number of innocent Caucasian males, among the two million who live in the area, who might by coincidence be associated with as much incriminatory evidence as the defendant in this case. How many others, for example, would share the same three license plate digits, have unexplained scratches, sound like the defendant and have hair that might have come from the victim in their car or have an equivalent amount of other incriminatory evidence pointing to them? What is this number? One other? Five others? Ten others? While it is for the jury to say and not I, my hunch is that the number is so small that a test with a false positive rate of one in one hundred would be very unlikely to mistakenly identify anyone.

P: Thank you, Dr., no further questions.

Considering when it was written, in some respects *The Honest Scientist's Guide to DNA Evidence* was indeed Utopian, but as we document in the pages that follow, in other respects the *Guide* sparked debates that led to needed reforms.

Well Bruce, that's it; at least that's all I could copy and bring back with me given existing mass constraints and the danger of temporal disruption. Of course, before I submit this to your volume, I must rewrite this material so it looks like these are my own ideas. And please keep what I have told you regarding the true source of my comments confidential until the equations for intertemporal transportation appear in *Nature*, and I have taken the first steps towards patenting my machine.

Sincerely,
Richard Lempert

(1) Thompson & Ford, The Meaning of a Match: Sources of Ambiguity in the Interpretation of DNA Prints. In *Forensic DNA Tech.* Ch. 7 (M. Farley & J. Harrington, eds. 1991).

(2) Lempert, Some Caveats Concerning DNA as Criminal Identification Evidence: With Thanks to the Reverend Bayes, 13 *Cardozo Law Review* 303 (1991).

(3) Establishing the likelihood of false positive errors and their implications for the probative value of DNA matches is no simple matter even putting aside the number of proficiency tests that should be conducted to achieve reliable error rate estimates. For example, the first external proficiency tests reported in the literature were conducted by the California Association of Crime Laboratory Directors which sent 50 mock forensic samples to several of the laboratories that pioneered in the forensic analysis of DNA identification evidence. In one set of tests, for example, a laboratory made a false positive error which a number of early commentators, particularly in law reviews, took as indicating that the best empirically grounded false positive error rate estimate was 2%. However, the laboratory was charged with comparing each sample with every other, which suggests that 1225 comparisons were made, leaving the impression of a far higher accuracy rate, a point also made in the literature. Neither portrait is correct. First, the laboratory in question was able to compare only 45 samples, and the presence of matching samples, including some that were matched by three other samples, further limited the number of false positives that could occur. Moreover, the likely source of the false positive was the inadvertent transfer of DNA in a sample in one test tube to an adjacent DNA containing test tube. This type of error could not affect most comparisons in the proficiency test because the test tubes were not all adjacent to each other. In actual forensic testing, however, suspect and evidence samples might routinely occupy adjacent test tubes in the test preparation process, suggesting that the 2% error rate estimate could be valid after all. However, the laboratory involved quite sensibly changed its procedures following the proficiency test results (and it claimed it had ordinarily done so in forensic work before the test) so little can be made of this figure as an estimate of the laboratory's future false positive rate. The story of this test and the error rate estimates it generated emphasizes the importance of double blind proficiency tests in which laboratories cannot distinguish proficiency test samples from ordinary forensic samples. For a discussion of further complications regarding the implications of error probabilities, see Thompson's discussion in this volume.

(4) Thompson & Schumann, Interpretation of Statistical Evidence in Criminal Trials, 11 *Law & Human Behavior* 167 (1987).

Editor's comments

The author's 'piece of whimsy' (his phrase) adds a light touch to this volume, even though it contains a serious discussion of the issue of false positives.