Journal of Criminal Law and Criminology

Volume 84
Issue 1 *Spring*

Article 2

Spring 1993

Evaluating the Admissibility of New Genetic Identification Tests: Lessons from the DNA War

William C. Thompson

0091-4169/93/8401-0022
THE JOURNAL OF CRIMINAL LAW & CRIMINOLOGY
Vol. 84, No. 1

Printed in U.S.A.

EVALUATING THE ADMISSIBILITY OF NEW GENETIC IDENTIFICATION TESTS: LESSONS FROM THE "DNA WAR"

WILLIAM C. THOMPSON*

I. INTRODUCTION

The development of forensic tests for "typing" DNA was greeted with enormous enthusiasm. Heralded as "the greatest advance in crime fighting technology since fingerprints,"[1] the new DNA typing tests were introduced swiftly into the legal system.[2] Recently, however, a serious scientific controversy has erupted over the reliability and specificity of DNA tests. Scientific critics have charged that the new technology was rushed to court before it was ready.[3]

The scientific dispute has led to confusion and conflicting rulings in the legal system. Although DNA evidence initially faced little opposition and has already been used to obtain convictions in hun-

* Associate Professor, Department of Criminology, Law & Society, School of Social Ecology, University of California, Irvine, California; J.D., University of California, 1982; Ph.D., Stanford, 1984.

The author thanks Professors Peter D'Eustachio, Laurence Mueller, Barry Scheck and William Shields for helpful comments on the manuscript, and Professor Simon Ford for helpful discussion of numerous technical issues.

[1] People v. Wesley, 533 N.Y.S.2d 643, 644 (N.Y. Sup. Ct. 1988), *aff'd*, 589 N.Y.S.2d 197 (N.Y. App. Div. 1992).

[2] A number of journalistic accounts have heralded DNA typing as a "breakthrough" that "could revolutionize law enforcement." *See* Debra Cassens Moss, *DNA—The New Fingerprints*, A.B.A. J., May 1, 1988, at 66; Ricki Lewis, *DNA Fingerprints: Witness for the Prosecution*, DISCOVER, June 1988, at 44; Jean L. Marx, *DNA Fingerprinting Takes the Witness Stand*, 240 SCIENCE 1616 (1988).

[3] For reviews of the controversy, see William C. Thompson & Simon Ford, *DNA Testing: Debate Update*, TRIAL, Apr. 1992, at 52 [hereinafter *DNA Testing*]; Leslie Roberts, *Fight Erupts Over DNA Fingerprinting*, 254 SCIENCE 1721 (1991); Peter Neufeld & Neville Colman, *When Science Takes the Witness Stand*, 262 SCI. AM. 46 (1990); William C. Thompson & Simon Ford, *Is DNA Fingerprinting Ready for the Courts?*, 125 NEW SCIENTIST 38 (1990) [hereinafter *DNA Fingerprinting*]; Eric S. Lander, *DNA Fingerprinting On Trial*, 339 NATURE 501 (1989); Rorie Sherman, *DNA Unraveling*, NAT'L L.J., Feb. 1, 1991, at 1.

dreds of cases,[4] a number of courts have recently ruled DNA evidence inadmissible on grounds that it is not generally accepted as reliable by the scientific community.[5] Some commentators have called for reexamination of the DNA evidence in closed cases.[6] The dispute has generated such heated exchanges, both inside and outside the courtroom, that commentators[7] and participants have called it a "war."[8]

Not surprisingly, proponents of forensic DNA evidence offer very different accounts of the controversy than do critics. Proponents claim that an excellent forensic procedure has been unfairly maligned by a small group of scientists who have garnered large fees for expert testimony by making exaggerated criticisms.[9] Ac-

---

[4] For a review of the early cases, see Thomas M. Fleming, Annotation, *Admissibility of DNA Identification Evidence*, 84 A.L.R. 4TH 313 (1991). Several excellent reviews of subsequent caselaw developments (as well as scientific issues raised in court) have been published in the BNA Criminal Practice Manual. *See Judicial Tide Turns Against Admission of FBI's DNA Tests*, 7 BNA CRIM. PRAC. MAN. 16 (1993); *DNA Typing Critics Say Report Validates Their Concerns*, 6 BNA CRIM. PRAC. MAN. 195 (1992).

[5] Appellate decisions holding DNA evidence inadmissible include, State v. Schwartz, 447 N.W.2d 422 (Minn. 1989); Commonwealth v. Curnin, 565 N.E.2d 440 (Mass. 1991); Commonwealth v. Lanigan, 596 N.E.2d 311 (Mass. 1992); People v. Barney, 10 Cal. Rptr. 2d 731 (Cal. Ct. App. 1992); State v. Vandebogart, 616 A.2d 843 (N.H. 1992); State v. Anderson, 853 P.2d 135 (N.M. Ct. App. 1993); United States v. Porter, 618 A.2d 629 (D.C. 1992); People v. Atoigue, No. CR 91-95A, 1992 WL 245628 (Guam Dist. Ct. App. Div. Sept. 11, 1992); People v. Wallace, 17 Cal. Rptr. 2d 721 (Cal. Ct. App. 1993); State v. Bible, No. Cr-90-0167-AP, 1993 WL 306544 (Ariz. filed Aug. 12, 1993) (en banc). Trial court decisions holding DNA tests inadmissible include People v. Castro, 545 N.Y.S.2d 985 (N.Y. Sup. Ct. 1989); State v. Wheeler, No. C89-0901CR (Wash. Co., Or. Cir. Ct. Mar. 8, 1990); State v. Despain, No. 15589 (Yuma Co., Ariz. Super. Ct. Feb. 12, 1991); State v. Fleming, No. 90-CR-2716 (Cook Co., Ill. Cir. Ct. Mar. 12, 1991); State. v. Watson, No. 90-CR-5546 (Cook Co., Ill. Cir. Ct. Mar. 12, 1991); State v. Passino, No. 185-1-90 Fcr (Franklin Co., Vt. Dist. Ct. May 13, 1991); People v. Halik, No. VA00843 (Los Angeles Co., Cal. Super. Ct. Sept. 26, 1991); United States v. Porter, 120 DAILY WASH. L. RPTR. 477 (D.C. Super. Ct. Sept. 20, 1991); People v. Keene, 591 N.Y.S.2d 733 (N.Y. Sup. Ct. 1992).

Recently, DNA evidence has been ruled inadmissible by two trial courts in the United Kingdom as well. Regina v. Levy, No. 920188 (Central Crim. Ct. Dec. 7, 1992); Regina v. Borham, No. 920857 (Central Crim. Ct. Nov. 3, 1992). These cases are discussed in William Brown, *DNA Fingerprinting Back in the Dock*, NEW SCIENTIST, Mar. 6, 1993, at 14.

[6] Neufeld & Colman, *supra* note 3.

[7] Edward Humes, *DNA War*, L.A. TIMES MAG., Nov. 29, 1992, at 29. For other accounts of the heated nature of the dispute, see Leslie Roberts, *Science in Court: A Culture Clash*, 257 SCIENCE 732 (1992) ("Vituperative arguments over DNA fingerprinting epitomize the difficulties of settling scientific disagreements in a highly charged legal environment."); Roberts, *supra* note 3, at 1721; Christopher Anderson, *DNA Fingerprinting Discord*, 354 NATURE 500 (1991).

[8] "This is no [longer] a search for the truth, it is a war, the way people are behaving." Roberts, *supra* note 7, at 736 (quoting John Hicks, head of the FBI's Laboratory Division).

[9] Andre Moenssens, *DNA Evidence and Its Critics—How Valid Are the Challenges?*, 31

cording to proponents, the major lesson of the controversy is that the legal system offers powerful incentives for ersatz scientific dissent and that courts take such dissent too seriously when evaluating the admissibility of new scientific techniques. By this account, the hostility surrounding the dispute arises from frustration over an untenable situation in which a few dissenting scientists are able to hold a new technology hostage by voicing ill-conceived objections to it.

On the other hand, critics claim that the new DNA tests were poorly validated and may be considerably less probative than juries have been led to believe.[10] According to critics, the major lesson of the controversy is that forensic laboratories are lacking in scientific rigor and that new forensic techniques receive inadequate scientific scrutiny before they are presented in court.[11] By this account, an atmosphere of hostility surrounds the issue because promoters of forensic DNA tests have responded to legitimate criticisms by attacking the critics rather than heeding their message.

The National Research Council ("NRC") of the National Academy of Sciences has made an important effort to resolve the underlying scientific controversy, but has not been entirely successful. Responding to a "crescendo of questions concerning DNA typing," the NRC established a distinguished panel of experts in 1990 to examine the use of DNA technology in forensic science.[12] The panel's report, based on a two-year study, was released in April 1992. Although the report firmly endorsed the general principles behind forensic DNA testing and said such testing should continue, it acknowledged that there have been problems with the implementation of the new technology and declared that improvements are necessary. A number of its recommendations for improvements are at

---

JURIMETRICS 87 (1990); George Clarke, *DNA Fingerprinting Critics Have Run Afoul of Science*, L.A. DAILY J., Jan. 28, 1992, at 5 (editorial comment of a prominent California prosecutor, arguing that scientific criticism of forensic DNA testing is "counterfeit dissent"). *See generally* Roberts, *supra* note 7; Leslie Roberts, *Hired Guns or True Believers*, 257 SCIENCE 735 (1992).

10 Richard C. Lewontin & Daniel L. Hartl, *Population Genetics in Forensic DNA Typing*, 254 SCIENCE 1745 (1991); William M. Shields, *Forensic DNA Typing as Evidence in Criminal Proceedings: Some Problems and Potential Solutions*, *in* PROCEEDINGS FROM THE THIRD INTERNATIONAL SYMPOSIUM ON HUMAN IDENTIFICATION 1 (1992); William C. Thompson & Simon Ford, *The Meaning of a Match: Sources of Ambiguity in the Interpretation of DNA Prints*, *in* FORENSIC DNA TECHNOLOGY 93 (Mark Farley & James Harrington eds., 1991).

11 Neufeld & Colman, *supra* note 3; Lander, *supra* note 3; Janet Hoeffel, *The Dark Side of DNA Profiling: Unreliable Scientific Evidence Meets the Criminal Defendant*, 42 STAN. L. REV. 465 (1990).

12 COMMITTEE ON DNA TECH. IN FORENSIC SCIENCE, NAT'L. RES. COUNCIL, DNA TECHNOLOGY IN FORENSIC SCIENCE vii (1992) [hereinafter NRC REPORT].

odds with the current practices of most forensic laboratories[13] and consistent with the suggestions of critics. Although the report clarifies the issues, it has generated further controversy, including charges that the committee was biased and subject to improper influence,[14] and that some sections of the report are scientifically flawed.[15] Its recommendations have generally been hailed by critics as essential for assuring the accuracy of DNA tests, and have been dismissed by proponents of the tests as unnecessary and unworkable.

The scientific and legal issues involved in the DNA war warrant close examination because they can tell us much about the evaluation of new techniques in forensic science. The current DNA tests are the first spin-offs of the molecular revolution in genetics to reach the courtroom, but they will not be the last.[16] The lessons taken from the current DNA war will shape the response of the legal system to battles over the DNA technology of the future. These lessons may also influence the legal system's response to disputes about the reliability of other scientific techniques.[17]

This article provides a detailed account of the controversy over forensic DNA testing and suggests some lessons that might properly

---

13 *See* State v. Alt, No. K4-90-1437 (Olmsted Co., Minn. Dist. Ct. May 29, 1992) ("The FBI Laboratory procedures are not totally defective, but neither do they meet the criteria of The National Research Council and others."). *See also* Richard Lewontin, *The Dream of the Human Genome*, N.Y. REV., May 28, 1992, at 17. Lewontin notes that the NRC report contains "numerous recommendations which, taken seriously, will lead any moderately businesslike defense attorney to file an immediate appeal of any case lost on DNA evidence." *Id.* at 39. The NRC Report cannot properly be read as an endorsement of current laboratory practices, Lewontin argues, "[s]ince no laboratory currently meets [the NRC's] requirements" for reliable DNA testing. *Id.*

14 Christopher Anderson, *Conflict Concerns Disrupt Panels, Cloud Testimony*, 355 NATURE 753 (1992); Celia Hooper, *Rancor Precedes National Academy of Science's DNA Fingerprinting Report*, 4 J. NIH RES. 76 (1992); Leslie Roberts, *DNA Fingerprinting: Academy Reports*, 256 SCIENCE 300 (1992).

15 For an account of scientific attacks on the NRC Report, see Peter Aldhous, *Geneticists Attack NRC Report as Scientifically Flawed*, 259 SCIENCE 755 (1993). Scientific articles criticizing the NRC's position include, Bernard Devlin et al., *Statistical Evaluation of DNA Fingerprinting: A Critique of the NRC's Report*, 259 SCIENCE 748 (1993); Bruce S. Weir, *Population Genetics in the Forensic DNA Debate*, 89 PROC. NAT'L. ACAD. SCI. 11654 (1993).

16 A number of new technologies are being developed. *See infra* notes 36-38 and accompanying text.

17 *See* Michael J. Saks & Jonathan J. Koehler, *What DNA "Fingerprinting" Can Teach the Law About the Rest of Forensic Science*, 13 CARDOZO L. REV. 361 (1991):

> The most important legacy of DNA "fingerprinting" and the debate surrounding it will likely be a spillover of standards of empirical testing and statistical rigor to many other forensic sciences which, somehow, have exempted themselves from the conventional standards of scientific rigor. In short, the debate over DNA fingerprinting may compel the rest of forensic science to become more recognizably scientific.

be taken from it. It discusses the nature of the disagreements among experts and analyzes how courts should respond to such disagreements.

A. OVERVIEW OF DNA TECHNOLOGY

The DNA tests discussed here are the product of basic advances in molecular biology over the past fifteen years that have revealed a number of potential methods for detecting genetic differences among individuals.[18]

Two of these methods have been developed and used for forensic testing so far, and additional methods are currently being developed. Continuing advances in molecular biology and genetics will undoubtedly prove a wellspring for more technology, sending a continuing stream of new genetic tests toward the courtroom.

The most widely used method at present, and the method that is the primary focus of this article, is commonly called DNA fingerprinting or DNA profiling. It employs a technique known as restriction fragment length polymorphism (RFLP) analysis in which long strands of DNA are extracted from biological material and broken into fragments, and then measurements are made of the length of certain fragments that tend to vary in length from person to person.[19] If two samples have fragments of different lengths, they could not have a common source; if they have fragments of the same

---

[18] Unlike traditional serology tests, which detect genetic differences by examining proteins and enzymes in bodily fluids, the new tests examine the DNA itself. DNA is a long, double-stranded molecule found in the chromosomes carried in all cell nuclei. The structure of DNA is similar to a long, twisted ladder in which the "rungs" consist of pairs of molecules called base-pairs. There are four different bases, labeled *A*, *T*, *G* and *C* which always form into the base-pairs *A-T* and *C-G*; the sequence of these base-pairs on the DNA strand constitutes the genetic code. Although most sections of the DNA molecule vary little from individual to individual within a species, some sections are polymorphic, which means that they have different forms in different individuals. The different forms are called alleles. *See generally* NRC REPORT, *supra* note 12, at 78; William C. Thompson & Simon Ford, *DNA Typing: Acceptance and Weight of the New Genetic Identification Tests*, 75 U. VA. L. REV. 45 (1989).

[19] In this process, DNA is broken into fragments through the use of restriction enzymes, and the resulting fragments are called restriction fragments. Hence, the polymorphic areas are known as restriction fragment length polymorphisms ("RFLPs") and examination of these areas is called RFLP analysis. For a detailed discussion of these techniques, see *infra* notes 59-75 and accompanying text.

About three thousand RFLPs have been identified. Since the early 1980s, geneticists have relied heavily on RFLPs as genetic markers for studying heredity. Most of the major advances in detecting disease genes that have occurred during the last decade have resulted from this application of RFLP analysis. For a general background on the development of RFLP analysis, see Jan Witkowski, *Milestones in the Development of DNA Technology*, *in* FORENSIC DNA TECHNOLOGY 1 (Mark Farley & James Harrington eds., 1991) and NRC REPORT, *supra* note 12, at 78.

lengths, they might have a common source (although they might also match by chance). To reduce the likelihood of a chance match, laboratories attempt to measure six or more distinct fragments.[20] Although RFLP analysis has been used in scientific research since the early 1980s, it became useful for identification of individuals only with the discovery of certain areas of DNA, known as variable number tandem repeats ("VNTRs"), where there is a great deal of variability in the length of DNA fragments, and thus a low likelihood that any two people will match with regard to the length of six or more fragments. By some accounts, RFLP analysis of VNTRs produces DNA profiles that are virtually unique to each individual, and thus are akin to fingerprints.[21]

DNA profiling was swiftly introduced in the legal system. Although the term "DNA fingerprint" did not enter the scientific lexicon until 1985,[22] two commercial laboratories in the United States were offering DNA profiling services by 1987;[23] and by 1988 the FBI laboratory began offering such services to law enforcement agencies nationwide.

There was almost no opposition to the tests at first, but critics gradually began to emerge. No one has questioned the theory underlying DNA testing nor the fundamental validity of RFLP analysis. Critics' concerns focus solely on forensic applications of RFLP analysis which, they argue, are more demanding than other applications of the technique.[24] The three major issues that have consistently

---

20 In some cases, fewer fragments are measured due to limitations imposed by technical problems or the limited quality of samples.

21 When DNA profiling tests were first introduced, it was frequently reported that the likelihood two people would be found, by chance, to have matching DNA prints was around one in thirty billion. *See, e.g.*, Moss, *supra* note 2, at 66. Considering that the population of the earth is approximately five billion, this is a rather impressive claim. The true likelihood is currently the subject of serious scientific debate.

22 The term "DNA fingerprint" was coined by British scientist Alec J. Jeffreys and his colleagues in 1985. Alec J. Jeffreys et al., *Individual-Specific Fingerprints of Human DNA*, 316 NATURE 76 (1985). It has been used in the United States as a trade name for forensic and paternity tests marketed by Cellmark Diagnostics Corp., a subsidiary of ICI International, which holds a license for the use of the genetic probes developed by Jeffreys. This article will use the generic term DNA profiling to refer to forsenic techniques that employ RFLP analysis of VNTRs.

23 The two laboratories were Cellmark Diagnostics of Germantown, Md., and Lifecodes Corporation, formerly of Valhalla, N.Y., now in Stamford, Conn.

24 *See* Lander, *supra* note 3, at 501:

> [T]rial judges have raced to admit DNA fingerprinting as evidence on the grounds that the methods are "generally accepted in the scientific community," citing the application of RFLPs in DNA diagnostics and accepting claims that false positives are virtually impossible.
>
> With due respect, the courts have been too hasty. Although DNA fingerprinting clearly offers tremendous potential as a forensic tool, the rush to court has ob-

been raised are: (1) the adequacy of standards and controls for assuring the reliability of the forensic tests, (2) the adequacy of the procedures used by forensic laboratories to determine whether DNA profiles "match," and (3) the accuracy of the procedures used to determine the statistical frequency or rarity of DNA profiles. In subsequent sections, these concerns, and the reactions of courts to them, will be discussed in detail.

A second method for DNA testing, less widely used than DNA profiling, employs a technique known as polymerase chain reaction (PCR). PCR tests identify an area of DNA in which there tends to be variation from one person to another and then "amplify" that area by causing the DNA strands to replicate themselves multiple times.[25] Once it is amplified, the DNA can be "typed" through the use of genetic probes.[26] If two samples have the same type, they may have a common source; if they do not have the same type, they could not have a common source.

The first technique to be developed for "typing" amplified DNA, and the one that has been most widely used in forensics, examines the HLA DQ-alpha gene.[27] DQ-alpha typing is not as discriminating as DNA profiling: there are twenty-one different DQ alpha types[28] which range in frequency from about one to fifteen

---

scured two critical points: first, DNA fingerprinting is far more technically demanding than DNA diagnostics; and second, the scientific community has not yet agreed on standards that ensure the reliability of the evidence.

25 DNA is extracted from samples and "primers" are used to locate a section of the DNA to be studied. This section of DNA is "amplified" by a process known as polymerase chain reaction ("PCR"), which mimics the natural process by which DNA copies itself. Even if the sample initially contains only a few copies of the fragment of interest, PCR will rapidly increase that number to about ten million, which is sufficient for analysis. The polymerase chain reaction was discovered by Kary B. Mullis, an employee of Cetus Corporation, a biotechnology firm, and was developed as a technology by Cetus. *See generally* NRC REPORT, *supra* note 12, at 63-70; Thompson & Ford, *supra* note 18, at 76-81.

26 The probes are engineered to detect specific forms (alleles) of a given gene, and are therefore called "allele specific oligonucleotide probes." *See* Russell Higuchi et al., *DNA Typing from Single Hairs*, 332 NATURE 543, 544 (1988); Thompson & Ford, *supra* note 18, at 76-81.

27 To be useful for criminal identification, the section of DNA chosen for analysis must be polymorphic and must have alleles (forms) that can reliably be identified. The HLA DQ-alpha gene was the first to be used in forensic PCR tests because it is the first polymorphic gene that could be identified with primers for which reliable allele-specific probes were developed. *See* Higuchi et al., *supra* note 26, at 544.

28 The DQ-alpha gene has six alleles. People have two copies of this gene in their DNA, one inherited from the mother and one from the father. Hence, there are 21 different "genotypes" in the HLA DQ-alpha system, representing all combinations of the six alleles. Each person has 1 of the 21 genotypes. Catherine T. Comey & Bruce Budowle, *Validation Studies on the Analysis of the HLA DQ-alpha Locus Using the Polymerase Chain Reaction*, 36 J. FORENSIC SCI. 1633 (1991).

percent,[29] making the likelihood of a coincidental match between different samples far higher than with DNA profiling tests. But the PCR DQ-alpha test is far more sensitive than DNA profiling, allowing it to "type" samples that are much smaller and older.[30]

The first forensic analysis employing the PCR DQ-alpha system was performed in 1986 by Dr. Edward Blake of Forensic Science Associates ("FSA"), a private forensic laboratory.[31] FSA was the first laboratory, and for several years was the only laboratory, to offer forensic PCR DQ-alpha tests. However, the technology has recently been licensed to a number of other forensic laboratories. Hence, evidence from PCR DQ-alpha tests is likely to be offered in court with increasing frequency.

The reliability of the PCR DQ-alpha test has been subject to a great deal of debate.[32] Although no one questions the theory underlying the procedure, concerns have arisen about its susceptibility to errors caused by inadvertant contamination of samples or poor "fidelity" of DNA amplification.[33] Concerns have also been raised about some of the procedures used by FSA for interpreting results[34] and for estimating the frequency of DQ-alpha types.[35]

---

29 The frequency varies across the different types, and the frequency of each type varies somewhat across different reference populations. *Id.* at 1643.

30 *See* Higuchi et al., *supra* note 26, at 545.

31 *See* Edward Blake et al., *Polymerase Chain Reaction (PCR) Amplification and Human Leukocyte Antigen (HLA) DQ-alpha Oligonucleotide Typing on Biological Evidence Samples: Casework Experience*, 37 J. FORENSIC SCI. 700 (1992); George Sensabaugh & Cecilia von Beroldingen, *The Polymerase Chain Reaction: Application to the Analysis of Biological Evidence*, *in* FORENSIC DNA TECHNOLOGY 63 (Mark Farley & James Harrington eds., 1991).

32 Like DNA profiling tests, these tests are subject to artifacts that can complicate the interpretation of results in forensic cases. As with DNA profiling, there is considerable debate about the susceptibility of PCR procedures to artifacts, the conditions under which artifacts might occur, and the reliability of the procedures used by forensic analysts to detect artifacts and avoid being misled by them. *See* NRC REPORT, *supra* note 12, at 63-70; Comey & Budowle, *supra* note 28, *passim*.

33 "In some cases, PCR can be qualitatively faithful but quantitatively unfaithful, because some alleles amplify more efficiently than others. A sample might contain a 50:50 mixture of two alleles and yield an amplified product with a 90:10 ratio." NRC REPORT, *supra* note 12, at 64 (citing Catherine Comey et al., *Use of Formamide to Improve Amplification of HLA DQ Alpha Sequences*, 10 BIOTECHNIQUES 60-61 (1991)).

34 An important point of controversy is whether PCR DQ-alpha tests provide a reliable basis for determining which alleles belong to which donor in a mixed sample. Analysts frequently assign multiple alleles in a mixed stain to "primary" and "secondary" donors based on the intensity of the "dots" indicating the presence of the alleles. The reliability of these assignments is of crucial importance in many cases. However, the NRC Report expresses concern about the reliability of such determinations. NRC REPORT, *supra* note 12, at 65. Another important set of issues involves the adequacy of controls used to detect contamination and the manner in which the controls are interpreted.

35 FSA uses a statistical estimation procedure that is based on the assumption that

A number of new techniques, currently at the experimental stage, will soon reach the courtroom. Several new techniques employ PCR but use methods for "typing" amplified DNA that afford greater specificity than the DQ-alpha system.[36] Another new method allows analysis of mitochondrial DNA.[37] Still another allows analysis of another type of variation in DNA called a minisatellite variant repeat ("MVR").[38] The latter two techniques represent fundamentally different approaches to typing DNA than the current RFLP and PCR-based methods. All of these new techniques will require independent validation and review. Hence, the challenge posed to the legal system by the need to evaluate new genetic technology is unlikely to diminish, and may well increase, over the foreseeable future.

B. LEGAL STANDARDS FOR ADMISSIBILITY OF SCIENTIFIC EVIDENCE

The results of new DNA tests are reviewed by courts for admissibility in pretrial evidentiary hearings. These hearings have been the primary arena in which the DNA war has been fought. In the majority of jurisdictions, courts have applied the rule established in *Frye v. United States*,[39] under which evidence derived from a novel scientific technique may be presented to a jury only if the court first

---

the frequencies of the various DQ-alpha alleles are statistically independent. The assumption of independence is problematic, however, because there is evidence of population structure—that is, evidence that some combinations of alleles (genotypes) are far more common in some groups than others. *See, e.g.*, Rhea Helmuth et al., *HLA-DQ alpha Allele and Genotype Frequencies in Various Human Populations, Determined by Using Enzyme Amplification and Oligonucleotide Probes*, 47 AM. J. HUM. GENETICS 515 (1990); Comey & Budowle, *supra* note 28, at 1642-44. For a general discussion of the problem posed by population structure for statistical estimation, see notes 171-87 and accompanying text; NRC REPORT, *supra* note 12, at 74-95.

36 Recently, primers and probes have been developed for other polymorphic genes. Methods for measuring the length of polymorphic fragments of amplified DNA known as amplified fragment length polymorphisms ("AMP-FLPs") and short tandem repeats ("STRs") are also under development and promise to increase the specificity of PCR-based tests. Chantal J. Fregeau-Aubin et al., An Evaluation of STRs and MVRs for Forensic DNA Typing, Abstract of Paper presented at the meeting of the American Academy of Forensic Science (Feb. 1993) (on file with William C. Thompson). However, the use of these newer methods in forensic identification appears to be still at the experimental stage. Most forensic PCR tests conducted to date have examined only the HLA DQ-alpha gene.

37 *See* Mark D. Stoneking et al., *Population Variation of Human Mitochondrial DNA Control Region Sequence Detected by Enzymatic Amplification and Sequence-specific Oligonucleotide Probes*, 48 AM. J. HUM. GENETICS 370 (1991). This novel technique was recently used for the first time in a case in San Diego, California.

38 Alec J. Jeffreys et al., *Minisatellite Repeat Coding as a Digital Approach to DNA Typing*, 354 NATURE 204 (1991).

39 293 F. 1013 (D.C. Cir. 1923).

determines that the technique has "gained general acceptance in the particular field in which it belongs."[40] In other jurisdictions, courts have followed the "relevancy" approach, under which novel scientific evidence is admitted if the judge determines that it will be more helpful than misleading to the jury.[41]

These admissibility rules primarily serve to shield juries from misleading or prejudicial scientific testimony. A common theme in appellate opinions and commentary is that, because lay jurors are overawed by science and lack the capacity to evaluate it critically, unreliable scientific testimony is likely to be given more weight than it deserves. There is also concern that efforts to challenge unreliable scientific evidence through cross-examination and expert testimony will be inordinately time-consuming, expensive and confusing, and, particularly in criminal cases where the prosecution presents novel evidence, that the defense may be handicapped by an inability to find experts who are knowledgeable about the new technique.

The *Frye* rule is generally viewed as a more restrictive standard for admissibility than the relevancy standard,[42] and has proved to be so with respect to DNA evidence. It is in *Frye* jurisdictions that DNA evidence has received the most thorough scrutiny during pretrial admissibility hearings and has most often been ruled inadmissible.

The *Frye* standard will be used less commonly in the future, however, due to the U.S. Supreme Court's decision in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,[43] which abrogated the *Frye* standard in federal trials. *Daubert* held that *Frye* had been superseded by the Federal Rules of Evidence, which do not require that scientific evidence meet the test of general acceptance to be admissible.[44]

---

40 *Id.* at 1014. For a general discussion of the application of the *Frye* rule to forensic DNA testing, see Thompson & Ford, *supra* note 18, at 53-60.

41 For a general discussion of the evaluation of DNA evidence under relevancy rules, see Hoeffel, *supra* note 11, at 507-19.

42 In relevancy jurisdictions, there is no need to show that scientists generally regard the procedure as reliable. Hence, new procedures about which there is substantial scientific controversy may nevertheless be admitted if the court either finds scientific proponents more persuasive than critics or finds that, notwithstanding the dispute, the jury will be able to evaluate the evidence appropriately. A necessary consequence of the greater discretion granted courts in relevancy jurisdictions is that judges bear the burden of making a correct assessment of both the reliability of scientific evidence and the jury's competence to evaluate that evidence, which, in the case of forensic DNA evidence, is a formidable task indeed.

43 113 S. Ct. 2786 (1993).

44 Expert testimony is governed by Federal Rule of Evidence 702, which states:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an

Although *Daubert* is binding only on federal courts, it may have significant persuasive effect at the state level, particularly in states that have adopted evidence codes paralleling the Federal Rules.[45]

But *Daubert* will not necessarily assure the admissibility of DNA evidence. Although the Federal Rules do not require that scientific evidence be "generally accepted," the Supreme Court emphasized that the Rules *do* require that scientific evidence be valid and that it assist the trier of fact.

The validity requirement stems partly from the language of Rule 702, which states that the subject of an expert's testimony be "scientific knowledge": "[I]n order to qualify as 'scientific knowledge,' an inference or assertion must be derived by the scientific method. Proposed testimony must be supported by appropriate validation—i.e., 'good grounds,' based on what is known."[46] Rule 702 also requires that the evidence or testimony "assist the trier of fact to understand the evidence or to determine a fact in issue."[47] Whether forensic DNA tests have been appropriately validated is the heart of the dispute over DNA evidence. Parties opposing DNA evidence will find plenty of scientific support for claims that the scientific validation is inadequate for at least some aspects of the forensic tests. Judges will need to weigh this evidence against the claims of scientific supporters that the techniques are valid.

The Supreme Court mentioned a number of factors that should bear on the admissibility of scientific evidence under Rule 702, noting that no single factor is dispositive.[48] The factors include "whether [the theory or technique] can be (and has been) tested,"[49] "whether the theory or technique has been subjected to peer review and publication,"[50] "the known or potential rate of error,"[51] "the existence and maintenance of standards controlling the technique's operation,"[52] and, finally, whether the theory or technique is gener-

expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.

45 For example, the New Mexico Supreme Court, following the example of *Daubert*, recently abandoned the *Frye* standard. The Court noted that the state's rules regarding scientific evidence are identical to the Federal Rules of Evidence. State v. Alberico, No. 20,282, 1993 WL 387950, at *1 (N.M. Aug. 30, 1993) ("Today we abandon the *Frye* test as a predicate for the admissibility of scientific evidence . . . relying instead on our Rules of Evidence.").

46 *Daubert*, 113 S. Ct. at 2795.

47 FED. R. EVID. 702; *Daubert*, 113 S. Ct. at 2795.

48 *Daubert*, 113 S. Ct. at 2796.

49 *Id.*

50 *Id.* at 2797.

51 *Id.*

52 *Id.*

ally accepted in the relevant scientific community.[53] Parties supporting and opposing the admissibility of DNA evidence have disagreed with respect to each of these factors.

Rule 702 also requires that scientific evidence or testimony "assist the trier of fact."[54] According to *Daubert*, this helpfulness standard "requires a valid scientific connection to the pertinent inquiry as a precondition to admissibility."[55] Some scientific testimony about forensic DNA tests may well be challenged on grounds that it lacks a valid scientific connection to facts properly at issue.[56]

Finally, the *Daubert* court emphasized the authority of the judge to exclude scientific evidence under Rule 403, which permits the exclusion of relevant evidence "if its probative value is substantially outweighed by the danger of unfair prejudice, confusions of the issues, or misleading the jury. . . ."[57] The Court noted the special dangers of prejudice surrounding scientific testimony, which "can be both powerful and quite misleading because of the difficulty in evaluating it."[58] Following *Daubert*, opponents of DNA evidence are likely to place special emphasis on its potential for prejudice.

II. MATCHING OF DNA PROFILES: SCIENTIFIC AND LEGAL ISSUES

DNA profiling tests have three distinct steps: creating DNA prints, determining whether DNA prints match, and (if a match is declared) estimating the frequency of such a match in a reference population. This section discusses the scientific and legal issues surrounding the first two steps. The third step (frequency estimation) is discussed below in Section III.

A. CREATING A DNA "PRINT" OR "PROFILE"

A DNA print (or profile) is a pattern of dark bands on an X-ray plate, known as an autorad (short for autoradiogram). The positions of the bands on the autorad indicate the lengths of the DNA fragments being compared. The techniques for creating DNA prints are drawn from the field of molecular biology and biochemis-

---

53 *Id.* at 2798.

54 FED. R. EVID. 702.

55 *Daubert*, 113 S. Ct. at 2795.

56 For example, testimony about the frequency of matching DNA types in a reference population, even if valid, may be unhelpful if the reference population is inappropriate or if other factors suggest a tenuous connection between the reported frequency and the probability of a false match in the case at hand. *See infra* part IV.

57 FED. R. EVID. 403.

58 *Daubert*, 113 S. Ct. at 2798 (citing Jack Weinstein, *Rule 702 of the Federal Rules of Evidence is Sound, it Should not be Amended*, 138 F.R.D. 631, 632 (1991)).

try and are known, collectively, as restriction fragment length polymorphism ("RFLP") analysis.

The procedures involved in forensic RFLP analysis are illustrated in Figure 1.[59] DNA is first extracted from biological samples and then exposed to a restriction enzyme, which cuts the long chain-like DNA molecules into restriction fragments. The restriction fragments are separated and sorted by length using a process known as electrophoresis[60] and then bound to a nylon membrane using a process known as Southern Transfer. At this point, DNA fragments are arrayed across the membrane according to their length, with the larger fragments at the top and the smaller fragments at the bottom. To determine the length of the fragments containing a particular VNTR, one need only locate their position in the array. This is done through the use of radioactive probes which selectively bind to fragments of interest. When X-ray film is placed on the membrane, the radioactive probes expose the film (known as an autorad), producing dark bands. The position of the bands on the autorad corresponds to the position of the VNTR fragments in the array, and thus reveals their length.

Because humans have two copies of most chromosomes, one from the mother and the other from the father, each DNA sample exposed to a single probe will ordinarily show two bands.[61] In Figure 1, Suspect 2 and the biological stain have bands in different locations, which indicates that the samples have VNTRs of different lengths and therefore must have come from different individuals. Suspect 1 and the stain have bands in the same positions, indicating that the two samples could have come from the same individual.

To reduce the chances of a coincidental match, forensic laboratories typically use three or four different probes, each of which detects a different VNTR.[62] Because VNTRs are highly variable in

59 A good general discussion of test procedures may be found in NRC REPORT, *supra* note 12, at 27-49; *see also* Thompson & Ford, *supra* note 18.

60 In electrophoresis, the DNA samples are placed at one end of an agarose gel, an electric current is applied, and the DNA, which carries a negative charge, is drawn through the gel. The shorter restriction fragments move more quickly than the longer fragments so that, after a time, the fragments are arrayed across the gel according to their length. Typically a number of DNA samples are run together in separate lanes on the same gel.

61 A person who inherits an identical fragment of DNA from both parents, however, will show only one band. Individuals who have two bands are called heterozygotes; those with one band are called homozygotes.

62 Sometimes the probes are mixed together in a cocktail, which produces DNA prints of six to eight bands on a single autorad. Figure 2 is an example of a DNA print produced by a cocktail of probes. Other times, the probes are run separately, producing two bands per sample on each of three or four autorads. In some cases, fewer fragments

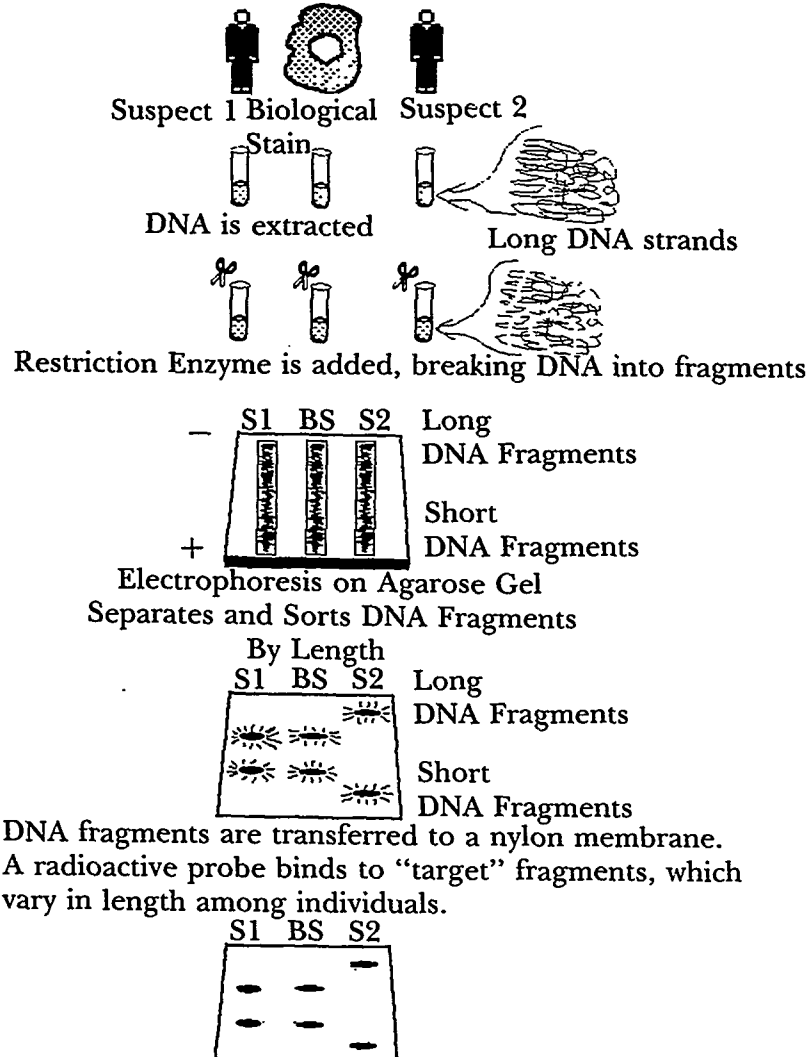

FIGURE 1. Schematic representation of the creation of DNA prints for RFLP analysis. In the example shown here, DNA of two suspects is compared to DNA in a biological stain using a single probe. DNA is extracted from specimens, broken into fragments with restriction enzymes and sorted by length using agarose gel electrophoresis. The DNA fragments in the agarose gel are then imprinted on a nylon membrane and exposed to a radioactive probe. The probe binds to fragments containing VNTRs, which tend to vary in length among individuals. When an X-ray film is exposed to the membrane, the radioactivity creates dark bands, indicating the position (and thus the length) of the variable fragments. In the example, Suspect 2 has fragments that differ in length from those of the biological stain, and he is therefore excluded as a possible contributor. Suspect 1 has bands of similar lengths, indicating he might have been the contributor. Exposing the membrane to additional probes can reduce the likelihood of a coincidental match, thereby increasing confidence that a suspect with a matching pattern of bands is the contributor of the stain.

length, there presumably is a low likelihood that any two people will match with regard to the length of six or more of them.

In order to adapt RFLP analysis for use in criminal identification, forensic scientists had to modify existing procedures and introduce several novel elements. Techniques had to be adapted to the special requirements of crime scene samples, which sometimes produce ambiguous results because they are old, contaminated, or composed of the DNA of more than one individual.[63] Quality control procedures needed to be tightened due to the need, in forensic DNA testing, to make decisions of critical importance based on a single result.[64] New procedures had to be developed for measuring and matching bands because forensic tests require greater precision than other applications of RFLP analysis.[65] Also, methods had to be developed for estimating the rarity of DNA prints,[66] a task not required for research or genetic diagnosis.

The adaptation occurred quickly (too quickly, some critics suggest) and RFLP-based forensic tests were rapidly introduced in the legal system. Some observers have suggested that Lifecodes and Cellmark, the two commercial laboratories that introduced forensic DNA profiling in 1987,[67] engaged in a "race to the courthouse," each hoping to gain a competitive advantage by more quickly establishing the admissibility of its DNA test.[68] There is some evidence that the FBI, which began accepting casework in 1988, introduced its test precipitously as well.[69]

---

are measured due to limitations imposed by technical problems or the limited quality of samples.

63 *See* Thompson & Ford, *supra* note 10, at 94; NRC REPORT, *supra* note 12, at 51-73.

64 "In research and medical diagnostics, scientists have ample samples and can repeat experiments that produce equivocal results. In many instances, conclusions are based on multiple samples; an error in one observation is of little consequence because it is averaged out. In forensic DNA testing, however, limited samples often make it impossible to run a test more than once." Thompson & Ford, *supra* note 10, at 95.

65 *See infra* Part II.B. for a discussion of the matching proceduress used in DNA profiling.

66 *See infra* Part III. for a discussion of these procedures.

67 *See supra* note 23.

68 Neufeld & Colman, *supra* note 3, at 46-53. By this account, each laboratory hoped to licence its testing procedure to as many crime laboratories as possible in order to gain a lucrative market for its proprietary reagents and testing materials. Although both DNA tests employed RFLP analysis, they used different reagents and materials and produced incompatible results—that is, the DNA print of the Lifecodes test looks different from and cannot be compared to the DNA print of the Cellmark test.

69 In January 1988, the FBI published a "validation protocol," developed at the FBI's Forensic Science Research Unit, that listed a number of steps "that should be ascended at the research level before DNA typing techniques will be certified for use on case evidence . . . ." Bruce Budowle et al., *An Introduction to the Methods of DNA Analysis Under*

No one questions that RFLP analysis is reliable and generally accepted within the scientific community for *some* purposes. Scientific criticism has focused solely on forensic applications of RFLP analysis, which are more demanding than other applications of the technique.[70] The criticism primarily concerns the steps of the forensic tests that are unique to forensic applications—that is, the procedures for matching and statistical estimation. Although the way forensic laboratories go about producing DNA prints (the first step) has been questioned as well, these questions largely stem from concerns that unreliability in the first step will complicate the process of matching and statistical estimation.[71] Critics also charge that inadequate safeguards in the procedures for producing DNA prints create a potential for false positives—that is, matches between samples with different DNA profiles—and that such events should be taken into account (but are not) when estimating the statistical likelihood of a match between different people.[72]

The NRC Report provided a measure of support for these criticisms. Although the report concluded that "[t]he current laboratory procedure for detecting DNA variation . . . is fundamentally sound,"[73] it declared a need for standardization of laboratory procedures, proficiency testing and accreditation of laboratories in order to assure the quality of forensic analyses.[74] It also recommended that laboratory error rates as determined by proficiency testing should be disclosed to juries.[75]

Investigation in the FBI Laboratory, CRIME LABORATORY DIG., Jan. 1988, at 19. However, the FBI began accepting casework later that year before completing those steps.

There has been speculation that the FBI rushed its test into service in order to promote a national standard for DNA testing that would allow development of a national index of DNA prints. Such an index would have been impossible if crime laboratories had adopted a hodgepodge of incompatible DNA testing procedures. By offering free training and other assistance to state and local crime laboratories, the FBI has persuaded them to use its test, thereby creating a *de facto* national standard that has enabled the creation of a national DNA index, which will be administered by the FBI.

70 For good reviews of the positions taken by defense experts in early DNA litigation, see People v. Axell, 1 Cal. Rptr. 2d 411, 419 (Cal. Ct. App. 1991), and People v. Castro, 545 N.Y.S.2d 985 (N.Y. Sup. Ct. 1989). *See also* Lander, *supra* note 3.

71 Some critics charge that the procedures used to produce forensic DNA prints introduce unnecessary and unpredictable variability in the position and number of bands, making the process of matching more complicated and less reliable. One issue, for example, concerns forensic laboratories' use of the chemical ethidium bromide during the electrophoresis process. Some critics have suggested that this chemical causes unpredictable shifts in the position of bands in DNA prints. *See* NRC REPORT, *supra* note 12, at 57, 58, 63; *see also* discussion *infra* note 122.

72 *See infra* part III.B. for further discussion of this issue.

73 NRC REPORT, *supra* note 12, at 149.

74 *Id.* at 16, 98, 108-09.

75 *Id.* at 88-89.

B. DETERMINING WHETHER DNA PRINTS MATCH

In the second step of DNA profiling, an analyst determines which samples should be declared to match. Under the procedures currently followed in all forensic laboratories, pairs of samples are typically classified as either a match or non-match (exclusion), although some laboratories allow a third category, inconclusive, for comparisons that do not meet the formal criteria for a match, but come close.[76]

Matching is difficult in some cases due to uncertainty about the number and position of bands. Bands may partially be obscured due to technical problems; spurious dark spots, easily mistaken for bands, may appear on the autorads; and the prints themselves may be faint or blurry.[77] Forensic laboratories have no objective standards for "scoring" bands; this task is left to the judgment of laboratory analysts.[78] An example of the ambiguities that can arise is shown in Figure 2.

Most forensic laboratories use computer-assisted imaging devices to help score bands. These devices scan the autorads and determine the number and position of bands based on fixed criteria.[79]

---

[76] Alternative approaches are possible and have been proposed. Some experts consider the match/no match dichotomy arbitrary and favor, instead, a so-called likelihood ratio or Bayesian approach in which the weight given to a particular result would vary according to how close the correspondence is between the DNA patterns being compared—there would be no bright-line cut-off between categories. I.W. Evett, *Evaluation of DNA Profiles: Sense and Nonsense*, 31 J. FORENSIC SCI. SOC'Y 205 (1991) ("a rigid cut off point where the evidence switches from good positive to exclusionary negative is quite illogical"); I.W. Evett et al., *An Efficient Statistical Procedure for Interpreting DNA Single Locus Profiling Data in Crime Cases*, 32 J. FORENSIC SCI. SOC'Y 307 (1992).

[77] Thompson & Ford, *supra* note 10, at 93-95.

[78] Critics have faulted the laboratories for leaving the scoring of bands to subjective judgment and for allowing analysts to exercise that judgment without being blinded to the identity of the samples or the expected outcome of the comparison. Whether a suspect is incriminated or exculpated by the test sometimes depends on the scoring of a few ambiguous bands. It is inappropriate, say the critics, that the forensic analysts who perform this task not only know the identity of the samples (i.e., which sample is the suspect's) but also often know details of the underlying case that could color their view of the likelihood the suspect is guilty. For a review of this issue, see Thompson & Ford, *supra* note 10.

Forensic analysts respond that the scoring of bands is clear and unambiguous in most cases and that in every case the autorads are available for examination by other experts should anyone wish to question their judgment. Mark D. Stolorow & George W. Clarke, *Forensic DNA Testing: A New Dimension in Criminal Evidence Gains Broad Acceptance*, 25 PROSECUTOR 13, 23 (1992).

[79] By comparing the position of a band in a DNA print with the positions of "marker" bands produced by DNA fragments of known length, it is possible to estimate the length of the polymorphic fragments represented by the band. This process is known as sizing the bands; the estimated length of the DNA fragment represented by a band is often called the band size and is expressed as a number of base pairs. A DNA

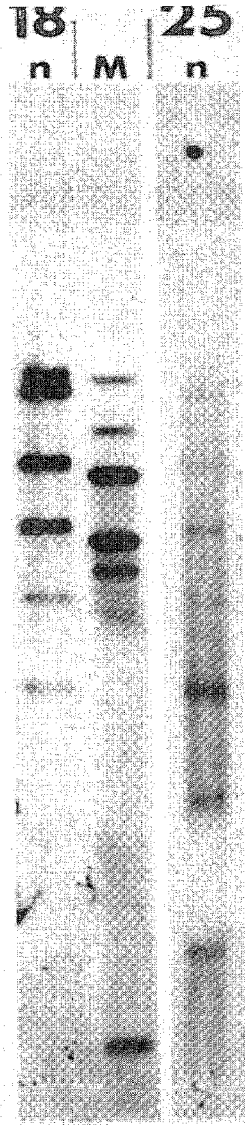

FIGURE 2. Ambiguous DNA Print Comparison. This figure shows two DNA prints of the same individual processed by Cellmark Diagnostics as part of a proficiency test. Lane 18n is a relatively fresh blood stain; lane 25n is a five year-old blood stain from the same individual. Lane M is DNA from a different individual and is being used as a control. Cellmark employed a cocktail of four probes, producing patterns with multiple bands. In an initial report, Cellmark reported that samples 18 and 25 were suitable for comparison but did not match. A subsequent report corrected the error. The comparison of 18 and 25 is complicated by degradation of the DNA in 25, which reduced the intensity of the upper bands, and by artifactual dark spots in the lower half of the print, which may have been mistaken for bands.

However, all forensic laboratories allow the analyst to override the computer's determinations manually and to cause the computer to reassign bands to whatever locations the analyst chooses.[80] Reassignment of bands through manual override can convert exculpatory results into incriminating ones, and vice versa.[81]

Matching is also made difficult by the potential for variability in DNA prints. The number and position of bands in someone's DNA print may change a bit depending on the quality of the biological samples and the testing conditions.[82] Hence, DNA prints of the same person are not always identical.

Forensic matching procedures are designed to take the variability of DNA prints into account. Two DNA prints may differ somewhat but nevertheless be declared to match if the discrepancies are thought to be within the normal range of variation for prints of the same individual. DNA prints are declared *not* to match only if the observed discrepancies are thought to exceed the normal range of variation in light of the testing conditions and quality of the samples. Hence, the match/no match determination depends on knowledge (or assumptions) about the range of variation possible in the results of the particular test.[83]

To deal with discrepancies in the *position* of bands, the forensic laboratories have all developed "quantitative matching rules" that specify how closely the bands in two DNA prints must align in order to be called a match.[84] If the bands in a pair of samples appear to align, the bands are "sized"—that is, an estimate is made of the length (in number of base-pairs) of the DNA fragment represented

---

profile is the set of band sizes that characterize a given DNA sample. Band sizing can be done by ruler measurement and hand calculation, but is laborious. The major advantage of the computer-assisted imaging devices is that they save time.

80 Records of these manual overrides are sketchy or non-existent; hence it is impossible to tell from laboratory reports how much role the analyst's judgment played in scoring bands in a particular case.

81 An interesting example was uncovered in State v. Despain, No. 15589 (Yuma Co., Ariz. Super. Ct. Feb. 12, 1991). After an expert retained by the defendant expressed concern about the accuracy of the FBI's band sizings, the court, upon a motion by the defendant, ordered the FBI to rescore the autorads using its computer imaging device but without any manual override. The computer's sizing of the autorads differed in a number of respects from the results the FBI had reported (which had been based on manual override of a previous computer sizing). Several bands that the FBI relied upon to establish a match between two key samples were placed in a different position by the computer than by the analyst. Had the computer's sizings been accepted, the FBI would have reported an exclusion rather than a match. Record at 1-98, *Despain* (No. 15589) (testimony of Dr. Simon Ford on Oct. 10, 1991).

82 *See generally* Thompson & Ford, *supra* note 10; Shields, *supra* note 10.

83 *See generally* NRC REPORT, *supra* note 12, at 51-73.

84 *Id.*

by each band.[85] The sizes of all corresponding bands must fall within a specified tolerance for the prints to be declared a match. Each laboratory has a different tolerance based, supposedly, on the possible range of variation of its results. For example, the FBI declares two DNA profiles to match only if the sizes of all corresponding bands are within approximately five percent of each other,[86] and Lifecodes requires that the band sizes differ by no more than 1.8%, although exceptions are sometimes made.[87]

To deal with discrepancies in the *number* of bands, laboratory analysts rely on the interpretation of indicators and controls designed to detect problems that could cause the disappearance of

---

85 To say that the "size" of a band is 1000 base-pairs, for example, means that the band is thought to reflect a DNA fragment of that length. Bands are "sized" by comparing their position on the autorad with the positions of "marker" bands resulting from fragments of a variety of known sizes. Based on the position of the markers, the laboratory computes a mathematical function relating band position and size that is used to translate the former to the latter.

86 The FBI has described its match criterion as a "±2.5% window." Bruce Budowle et al., *Fixed Bin Analysis for Statistical Evaluation of Continuous Distributions of Allelic Data from VNTR Loci, for Use in Forensic Comparisons*, 48 AM. J. HUM. GENETICS 841, 844-45 (1991). This rather ambiguous characterization has left some commentators with the erroneous impression that the FBI requires the measured sizes of two bands to be within 2.5% of each other to be declared a match. In fact, the FBI protocol specifies that the ±2.5% window be drawn around each band and that a match may be called if these windows overlap. Federal Bureau of Investigation, Procedures for the Detection of Restriction Fragment Length Polymorphisms in Human DNA at 21 (Dec. 4, 1989) (unpublished laboratory protocol); Shields, *supra* note 10, at 11. Hence, bands that differ in measured size by up to 5% may be declared to match. Moreover, one defense expert has charged that there are "fudge factors" in the FBI's matching procedure that expand the range of potential matches even beyond 5%. Record at 1-98, State v. Despain, No. 15589 (Yuma Co., Ariz. Super. Ct. Feb. 12, 1991) (testimony of Dr. Simon Ford). Among the "fudge factors" is the ability of the analyst to override the computer scoring system if it showed band comparisons to be slightly outside the FBI's match "window" and to manually reassign the position of bands to make them match, an event that apparently occurred in that case. *See supra* note 81; *see also* State v. Bruno, Nos. CR 89-7443, 7444, & 7445 (New Hanover Co., N.C., Super. Ct., 1990), *aff'd*, 424 S.E.2d 440 (N.C. 1993) (band measurements declared to match by FBI).

87 *See* Lander, *supra* note 3, at 502. Critics have questioned whether there is an adequate scientific foundation for the quantitative matching rules—that is, whether the forensic laboratories know enough about the variability of their own procedures to devise appropriate quantitative criteria for a match. Eric Lander, *Invited Editorial: Research on DNA Typing Catching Up with Courtroom Application*, 48 AM. J. HUM. GENETICS 819 (1991); Shields, *supra* note 10, at 8-17. Critics have also charged that some laboratories use quantitative matching rules that are too broad, such that prints so discrepant that they are unlikely to be from the same individual may nevertheless be declared to match. *Id.*; *see also infra* note 121 and accompanying text. Proponents generally respond that concerns about matching are of marginal importance. The precise size of the matching standard matters little, they say, because the tests are so discriminating that DNA prints of different individuals are unlikely even to come close to matching under the matching standards currently in use. *See infra* notes 127-28 and accompanying text.

true bands or the appearance of spurious extra bands.[88] The analyst may need to review the entire process of creating the DNA print, or conduct additional experiments, to determine whether the observed discrepancies were caused by technical problems rather than true genetic differences between the samples. To make such determinations accurately, analysts must have a sophisticated knowledge of both the procedures used to create the prints and factors that could cause variability in the results.[89]

A third factor complicating the determination of matches is the potential, in forensic work, for samples to contain the DNA of more than one individual. In rape cases, for example, vaginal swabs may contain DNA of both the rapist and the victim, as well as the DNA of other sexual partners of the victim. Blood stains at crime scenes occasionally are mixtures, containing more than one person's DNA. Samples can also become mixed if mishandling by police or laboratory technicians causes cross-contamination. Interpretation of forensic DNA tests often requires analysts to distinguish the extra bands or alleles that indicate a mixed sample from extra bands and alleles that result from laboratory artifacts or accidental cross-contamination. The adequacy and reliability of the forensic procedures used to make such distinctions has also been questioned.[90]

C. THE DEBATE OVER MATCHING IN THE COURTS

The difficulty of distinguishing matching and non-matching DNA prints first came to public attention in the widely publicized case of *People v. Castro*.[91] Lifecodes Corporation had declared a match between blood samples from a homicide victim and a blood stain on the defendant's watch. This evidence was offered to prove the defendant's identity as the killer. In a pretrial hearing on the admissibility of the DNA evidence, several prominent scientists ini-

88 *See* Thompson & Ford, *supra* note 10, at 110.

89 NRC REPORT, *supra* note 12, at 52-63. Critics have charged that forensic analysts sometimes make these determinations in a cavalier manner, without adequate scientific foundation. *See, e.g.*, Lander, *supra* note 3, *passim*. The failure of forensic laboratories to "blind" analysts to the identity of the samples and the expected outcome of the analysis is also an issue for these judgments. *See* Thompson & Ford, *supra* note 10, at 110-14. Proponents respond that misinterpretations have been rare and represent, at worst, a misuse of DNA technology rather than a problem with the technology per se. Moenssens, *supra* note 9, at 107; Stolorow & Clarke, *supra* note 78, at 24-25. These proponents assert that misinterpretation of DNA tests, if and when it occurs, can be dealt with appropriately in the adversarial process by allowing criminal defendants to hire their own experts to give a second opinion.

90 Thompson & Ford, *supra* note 10, at 116-20; NRC REPORT, *supra* note 12, at 55, 58-59.

91 545 N.Y.S.2d 985 (N.Y. Sup. Ct. 1989).

tially vouched for the reliability of Lifecodes' findings. But experts retained by the defendant criticized Lifecodes' procedure on a number of grounds.[92] First, there were disagreements about the scoring of bands. According to the defense experts, Lifecodes drew conclusions from ambiguous data, and did so in a seemingly biased manner—scoring faint bands that helped incriminate the defendant while ignoring bands that would have exculpated the defendant. Lifecodes' rationale for ignoring potentially exculpatory data was *ad hoc* and lacked scientific foundation. Lifecodes had not used adequate controls to verify that its interpretations were correct and had ignored the failure of some of the controls it did use. Finally, Lifecodes had disregarded discrepancies between the bands it had scored that exceeded its quantitative match criteria. Consequently, the defense argued that there was inadequate scientific basis for Lifecodes' conclusion that the victim's blood and the blood on the watch had a matching genetic characteristics.[93]

These criticisms proved persuasive even to some of the prosecution's expert witnesses. In an unusual move, two scientists who had testified for the prosecution met privately with two experts for the defense. The group issued a statement saying that the results in the case "are not scientifically reliable enough" to support the conclusion that the two key samples matched.[94] Thereafter, the two prosecution experts retook the stand and recanted their previous testimony in support of the test.[95]

The Court ruled the evidence of a DNA match inadmissible

92 Excellent summaries of the scientific issues in *Castro* may be found in Lander, *supra* note 3, and Hoeffel, *supra* note 11.

93 There was also extensive criticism of Lifecodes' procedures for estimating the frequency of the matching DNA prints, which will be discussed *infra* at part III.

94 Lander, *supra* note 3, at 504.

95 *Id.* After the DNA evidence was ruled inadmissible, defendant Castro pleaded guilty. Moenssens, *supra* note 9, at 98-99. In discussions of the case, some commentators have cited the guilty plea as evidence that criticisms of Lifecodes' techniques were overblown and incorrect. *Id.*; Stolorow & Clarke, *supra* note 78, at 18-19. This claim is fatuous. The defense experts did not argue that Lifecodes' results were exculpatory; they argued that Lifecodes' conclusions were inadequately grounded in science. That Lifecodes' conclusions happened to be correct (if that is indeed what Castro's negotiated plea means) is not a response to these criticisms. A conclusion reached in a slapdash, haphazard manner does not suddenly become intellectually rigorous if it turns out to be correct.

A related claim, frequently made, is that the problems that created ambiguity in the *Castro* case were unique and have not occurred in other cases. Moenssens, *supra* note 9, at 107. This claim is demonstrably wrong. *See* Thompson & Ford, *supra* note 10, at 128-30 (documenting numerous examples from forensic casework of problems similar to those in *Castro*); Shields, *supra* note 10, at 14-16 (citing examples from casework); Thompson & Ford, *Is the Probative Value of Forensic DNA Evidence Undermined by Subjectivity in Determination of Matches*, *in* PROCEEDINGS OF THE SECOND INTERNATIONAL CONFERENCE

under an innovative legal theory.[96] Despite the spectacular clash of experts that had occurred before him, the judge found that the prosecution had met its burden under *Frye* of proving that "DNA forensic techniques and experiments are generally accepted in the scientific community and can produce reliable results."[97] The judge construed the scientific clash as a dispute over how properly to apply a technique that "can produce reliable results" rather than a dispute about whether the technique is inherently unreliable. The DNA evidence of a match was ruled inadmissible because "[t]he testing laboratory failed in several major respects to use the generally accepted scientific techniques and experiments. . . ."[98] In sum, *Castro* held that DNA profiling meets the *Frye* standard, but imposed an additional foundational requirement on the proponent of the technology of demonstrating that the correct techniques had been used, and held that the prosecution had failed to meet that foundational requirement.[99]

Following *Castro*, disputes over Lifecodes' matching procedures have continued to surface in litigation.[100] Scientific critics have questioned the scientific basis for Lifecodes' quantitative matching

---

ON FORENSIC STATISTICS (forthcoming) (documenting examples from recent casework). *See also* cases discussed *infra* part II.C.

96 Lifecodes' conclusion that the blood on the watch did not match Castro was not challenged by the defense; the challenge focused solely on the conclusion that the blood matched the victim in the case.

97 People v. Castro, 545 N.Y.S.2d 985, 999 (N.Y. Sup. Ct. 1989).

98 *Id.*

99 A number of courts have followed *Castro* in requiring proponents of DNA evidence to make a foundational showing that "correct procedures" were used in the particular case. United States v. Two Bulls, 918 F.2d 56 (8th Cir. 1990); Smith v. Deppish, 807 P.2d 144 (Kan. 1991); People v. Fishback, 829 P.2d 489 (Colo. Ct. App. 1992); Perry v. State, 586 So. 2d 242 (Ala. 1991). *See generally* Edward J. Imwinkelried, *The Debate in the DNA Cases Over the Foundation for the Admission of Scientific Evidence: The Importance of Human Error as a Cause of Forensic Misanalysis*, 69 WASH. U. L.Q. 19 (Cal. Ct. App. 1991). In some jurisdictions, proof that "correct procedures" were used is considered part of the required showing under *Frye*. *See* People v. Axell, 1 Cal. Rptr. 2d 411 (Cal. Ct. App. 1991); People v. Barney, 10 Cal. Rptr. 2d 731 (1992). For arguments in favor of the *Castro* approach, see Hoeffel, *supra* note 11, at 503-07.

However, other courts have taken a more limited view of the requirements for admissibility under *Frye*. *See* People v. Wesley, 589 N.Y.S.2d 197, 199 (N.Y. App. Div. 1992) ("[A]ncillary issues regarding integrity of the particular forensic sample from which the DNA fingerprint was obtained and whether the laboratory followed the accepted procedures in carrying out the tests on the particular sample at issue speak to the weight the evidence is accorded and thus are not relevant to the initial determination of admissibility. . . ."); Hopkins v. State, 579 N.E.2d 1297 (Ind. 1991).

100 *See* Thompson & Ford, *supra* note 10, at 98-110 (discussing technical details of several cases); Alun Anderson, *DNA Fingerprinting on Trial*, 342 NATURE 844 (1989); Colin Norman, *Maine Case Deals Blow to DNA Fingerprinting*, 246 SCIENCE 1556 (1989).

rule[101] and have criticized Lifecodes for relying on ambiguous data and for calling matches between bands that differ in size more than allowed by Lifecodes' quantitative match criteria.[102] Disputes have arisen in some cases over Lifecodes' use of "monomorphic probes" (which produce a band of the same size in every sample) to adjust for spurious misalignment of bands caused by technical problems.[103] In the majority of cases, courts have held that disputes over the matching procedure raise issues going to the weight, rather than the admissibility, of the DNA evidence.

At least one court, however, has held that DNA evidence is inadmissible where the "match" did not meet Lifecodes' quantitative matching rule. In *People v. Keene*,[104] experts for the prosecution "testified that there was a 'satisfying impression of a match' here. . . ."[105] But two experts retained by the defendant were neither satisfied with the match, nor impressed by Lifecodes' use of monomorphic probes to "adjust" for the misalignment of bands. They testified that the use of monomorphic probes for this purpose is not generally accepted in the scientific community, and that Lifecodes had performed the adjustment improperly in any case. Acknowledging the complexity of the issue, the court declared "it would be judicial foolhardiness to submit the issue . . . to the jury to determine the weight of such evidence."[106] A key factor in the court's decision was the NRC Report, published shortly after the hearing in the case, which advised against the use of monomorphic probes to "adjust" for misalignment of bands, pending further

---

101 The reproducibility studies on which Lifecodes based its match criteria were performed on fresh, clean blood samples. *See* Michael Baird et al., *Allele Frequency Distribution of Two Highly Polymorphic DNA Sequences in Three Ethnic Groups and Its Application to the Determination of Paternity*, 39 AM. J. HUM. GENETICS 489, 494 (1986). Some experts find such studies inadequate because they neglect "the artifacts that can arise with degraded and contaminated evidence samples" and may therefore give an inaccurate picture of the actual variability of the DNA typing system. Lander, *supra* note 87, at 820. *See also* NRC REPORT, *supra* note 12, at 61-62 ("The match criterion must be based on reproducibility studies that show the actual degree of variability observed when multiple samples from the same person are separately prepared and analyzed *under typical forensic conditions*.") (emphasis added).

102 *See, e.g.*, Caldwell v. State, 393 S.E.2d 436 (Ga. 1990). *See also* Anderson, *supra* note 100, at 844 (discussing such a case); Norman, *supra* note 100, at 1556 (discussing such a case).

103 For a discussion of the problem of spurious band misalignment (sometimes called "band shift") and the use of monomorphic probes to correct for it, see NRC REPORT, *supra* note 12, at 59-61; Thompson & Ford, *supra* note 10, at 98-110; Lander, *supra* note 87, at 821 (characterizing use of monomorphic probes to correct for band shift as "venturesome").

104 591 N.Y.S.2d 733 (N.Y. Sup. Ct. 1992).

105 *Id.* at 737.

106 *Id.* at 740.

study,[107] and thus strongly supported the defense claim that this procedure is not generally accepted in the scientific community.

Disputes have also arisen over the admissibility of Cellmark's test.[108] Cellmark has been criticised for failing to publish research on the reproducibility of measurements in its DNA typing system,[109] for failing to take the variability of measurements into account when it established its quantitative match criterion,[110] and for calling matches in cases where the discrepancies between bands exceeded its quantitative match criterion.[111] Cellmark has also been accused of drawing conclusions in some cases in a cavalier manner, without adequate scientific foundation.[112] So far, all challenges on these grounds have been held to go to the weight rather than the admissibility of Cellmark's DNA test.[113]

The FBI's matching procedure has been similarly disputed.[114] The most extensive and important clash occurred in *United States v. Yee*.[115] A central issue in *Yee* was the adequacy of the FBI's research

---

107 NRC REPORT, *supra* note 12, at 60-61.

108 People v. Axell, 1 Cal. Rptr. 2d 411 (Cal. Ct. App. 1991); People v. Barney, 10 Cal. Rptr. 2d 731 (Cal. Ct. App. 1992).

109 *Axell*, 1 Cal. Rptr. 2d at 426; *Barney*, 10 Cal. Rptr. 2d at 739.

110 Cellmark did not perform reproducibility studies before establishing its match criterion. It initially based its match criterion on the amount of variability in band measurements that its system could reliably detect, rather than the amount of variability that actually occurred. Lander has declared this to be an invalid basis for a match criterion because it "ignor[es] all sources of experimental variability other than the final step of [scoring the band]." Lander, *supra* note 87, at 820. In response, Cellmark's scientific staff noted that they have since performed a reproducibility study. Robin W. Cotton et al., Response Letter, *Research on DNA Typing Validated in the Literature*, 49 AM. J. HUM. GENETICS 898 (1991). Lander replied that Cellmark's reproducibility study "remains unpublished to my knowldege and thus is not yet in the scientific literature." Eric Lander, *Lander Reply*, 49 AM. J. HUM. GENETICS 899 (1991) [hereinafter *Lander Reply*]. *See also* Shields, *supra* note 10, at 11 ("It is at least possible that [Cellmark's match criteria] can be validated statistically . . . but until they publish their studies this remains an open question."). Cellmark's Laboratory Director refused a request by the author of this article for a copy of its unpublished reproducibility data. Interview with Dr. Daniel Garner, Cellmark Laboratory Director, in Anaheim, Cal. (Feb. 21, 1991). These data remain unpublished and are apparently available only by court order.

111 *See, e.g.*, State v. Schwartz, 447 N.W.2d 422, 426 (Minn. 1989) ("[T]he Cellmark report opined that the DNA samples from the stained blue jeans and from [the victim's] blood 'are from the same individual,' even though the banding patterns did not fit their match criteria.").

112 *See* Thompson & Ford, *supra* note 10, at 116-120 (discussing technical details of several cases); Thompson & Ford, *supra* note 95, at 11-15 (manuscript)(discussing a case).

113 *See, e.g.*, *Axell*, 1 Cal. Rptr. 2d at 411. *But see Barney*, 10 Cal. Rptr. 2d at 731, and discussion thereof, *infra* notes 156-59 and accompanying text.

114 For a particularly cogent critique of the FBI's matching standards, see Shields, *supra* note 10.

115 134 F.R.D. 161 (N.D. Ohio 1990).

to validate its matching procedure—research that is described in several published articles.[116] Concerned about ambiguities and inconsistencies in these articles,[117] the defense in *Yee* sought access to the underlying data. After protracted litigation, access was granted.[118]

The defense in *Yee* made a multi-pronged attack on the FBI's matching standards based on evaluations of these data by experts they had retained. First, they argued that the FBI's validation research was inadequate and, indeed, that many of the conclusions in the FBI's published articles were either groundless or wrong.[119] Second, they argued that key findings reported in the FBI's validation research could not be replicated.[120] Third, they argued that the FBI's system for DNA profiling is subject to unpredictable variabil-

116 The key publications are Bruce Budowle et al., *Validation with Regard to Environmental Insults of the RFLP Procedure for Forensic Purposes, in* FORENSIC DNA TECHNOLOGY (Mark Farley & James Harrington eds., 1991); Budowle et al., *supra* note 86, at 841.

117 Prepublication copies of these articles were available in manuscript form during the *Yee* hearing in 1990.

118 In United States v. Yee, 129 F.R.D. 629 (N.D. Ohio 1990), the court granted defendants' motion for discovery.

119 For example, Professor Peter D'Eustachio, a molecular biologist at New York University, in an Expert's Report placed in evidence in *Yee*, stated:

> I have reviewed the two FBI manuscripts [Environmental Insult Article and Fixed Bin Article] as well as the lab notebooks and autoradiograms from which the results reported in these manuscripts were taken. Based on this review, I find no scientific basis for their conclusions. Many of the key experiments were badly designed and poorly executed, and many interpretations of data appear to be arbitrary, often because key controls failed or were omitted entirely . . . .
>
> Thus, instead of concluding that the FBI has developed and validated a reliable, sensitive procedure for identification of forensic DNA specimens, I conclude that they have not, and that the validation procedures themselves are badly flawed.

Report of Professor Peter D'Eustachio, *An Evaluation of the FBI's Environmental Insult Validation Study and the FBI's Quantitative Matching Criteria* at 5-7, *Yee*, 129 F.R.D. 629 [hereinafter D'Eustachio Report].

120 The data that the FBI's scientific staff characterize as "most important for determination of a tolerance window for matching criteria for forensic analysis" are contained in Table 3 of Budowle et al., *supra* note 86, at 842. This table compares the measurement of DNA fragments in different samples known to have come from the same person (the victim) in actual forensic cases. When the defense in *Yee* sought to discover the data underlying this table, the FBI could not produce it. The agency reported that it had not kept adequate scientific records. Instead, the FBI produced more recent data of the same type. However, these new data did not replicate the findings of Table 3. The data in Table 3 indicate that the measured size of bands in evidentiary samples (vaginal swabs) tends to be slightly larger than the measured size of the same bands in blood samples. The new data also showed such a measurement bias, but in the opposite direction. Professor D'Eustachio's evaluation of this finding was scathing:

> The data from which the FBI derived their plus or minus 2.5% matching window reveal a significant bias in DNA band size measurements that furthermore has

ity[121] that can affect the sizing of bands and greatly complicate the task of distinguishing matching and non-matching samples.[122] The

changed over time for reasons that are not apparent. These data simply do not provide a reliable and reproducible basis for making decisions as to window sizes.

. . . .

This unrepeatable bias means that there is no objective, reproducible standard for comparison of known and questioned DNA specimens.

. . . .

In terms of good science this leaves us nowhere . . . . The validation studies for devising a quantitative matching rule should be done again, and done right this time. D'Eustachio Report *supra* note 119 at 16, 18, 26.

121 During the *Yee* hearing, much was made of an unpublished FBI study in which 225 FBI agent-trainees were each DNA typed on two occasions. When defense experts examined the data they made a startling discovery: by the FBI's own standards many of the agents did not match themselves. According to a report filed with the Court by Professor Daniel Hartl of Washington University Medical School:

My analysis of the FBI test and retest data implies that the binning procedure as carried out by the FBI makes egregious errors at a truly incredible rate. In a research laboratory such quality control would be totally unacceptable, and in a medical diagnostic laboratory it would be criminally negligent. In consideration of these data, no reasonable case can be made that the FBI procedures, as presently applied, are reliable and generally acceptable.

Report of Professor Daniel Hartl, *Expert's Report in the case of United States v. Yee*, at 2, *Yee*, 129 F.R.D. 169 [hereinafter Hartl Report].

The FBI's scientists responded that it is unfair to treat the comparison as a reproducibility study, because "the samples were not analyzed under identical conditions." Bruce Budowle & John Stafford, *Response to Hartl, Expert's Report in the Case of* United States v. Yee, et al., 18 CRIME LABORATORY DIG. 101 (1991). Critics of the FBI argue that there is also uncontrollable variability in the conditions under which actual forensic samples are analyzed.

122 Another expert for the defense in *Yee*, Dr. Paul Hagerman of the University of Colorado Medical School, placed in evidence a report that is highly critical of the FBI's methods for "isolation and quantitation" of DNA and of the FBI's use of a chemical called ethidium bromide ("*EtBr*") in its gels during electrophoresis of DNA. Report of Dr. Hagerman, *Loading Variability and the Use of Ethidium Bromide: Implications for the Reliability of the FBI's Methodology for Forensic DNA Typing* at 1, *Yee*, 129 F.R.D. 629. According to Hagerman, these problems are relevant to the FBI's matching procedure, because they interfere with the FBI's ability to measure bands accurately by creating "significant and unpredictable shifts in band position" which can "result in serious problems in the interpretation of bands." *Id.* These and other problems in the FBI's procedures "call into question [the FBI's] ability either to size an individual band accurately against known standards, or to make an accurate comparison between two bands representing DNAs from different sources." *Id.* at 12. These problems also undermine the validity of the studies the FBI performed to determine the appropriate size for its matching window. *Id.* at 13. In particular, the problems "call into question the scientific reliability of the lane-to-lane comparisons described in Table 3 of the 'Fixed Bin' paper . . . from which the FBI's quantitative matching rule was derived." *Id.* Hagerman expresses concern that "the match criterion is too broad (excessively high rate of apparent inclusions)" and notes that the FBI's match criteria (window) is much broader than that of either Lifecodes or Cellmark. *Id.* Hagerman concludes that these considerations "render highly problematic [the FBI's] match window of 5%" and "lead to the conclusion that the match window is too large, i.e., that it will result in too many apparent matches." *Id.* at 16. He also expresses concern that "differential shifts may be occurring in a manner that would not be readily apparent, thus leading to potential misidentifications." *Id.* at 16-17.

defense took the position that too little is known about the variability of the FBI's DNA typing system, or its susceptibility to artifacts, to allow adequate assessment of the FBI's matching procedure, but suggested that available evidence indicated the FBI's quantitative match standards are generally too broad.[123]

Experts for the prosecution were unable to refute these criticisms directly,[124] but managed to persuade the federal magistrate (whose conclusions were adopted by the court) that they raise issues going to the weight rather than the admissibility of DNA evidence. The court, as did the magistrate, acknowledged that there is significant disagreement about the reliability of the FBI's test: "Scientists of indisputable national and international repute and stature, aided and confronted by lawyers of unusual skill and understanding of the issues, took diametrically opposed views on the issue of general acceptability, and those views reflected the division of opinion on the merits of the underlying scientific disagreement."[125] Nevertheless, the court ruled that the FBI's DNA test met the *Frye* standard.[126]

---

123 Professor Hartl states that the FBI's window "should be scaled down by approximately one half" to be consistent with accepted scientific practice. Hartl Report, *supra* note 121, at 3. He notes that "[t]he problem with using the [FBI's] window is that too many false matches are declared" and concludes that "the matching criteria employed by the FBI would not be considered as generally accepted and reliable in the scientific community." *Id.*

Prof. D'Eustachio agreed:

> [F]actors causing bands that are in fact of the same size to migrate differently on gels must be understood, or at least well controlled and reproducible . . . [and] the window size should be chosen so that it does not exceed the level of risk for false positives (matches) that is generally accepted as reliable in the research and clinical communities for similar applications of this technology . . . . The FBI fails to meet either standard.

D'Eustachio Report, *supra* note 119, at 16.

124 The magistrate generally credited the defense expert's criticisms of the validation underlying the matching standard, concluding that there were "troublesome questions about the quality of the Bureau's work." United States v. Yee, 134 F.R.D. 161, 207 (N.D. Ohio 1991). "I do not either disregard or discount the accuracy of many of the criticisms about the remarkably poor quality of the FBI's work and infidelity to important scientific principles." *Id.* at 210. The magistrate found, for example, that "Dr. Budowle [of the FBI] did not respond persuasively to Dr. D'Eustachio's criticisms, and he refused to acknowledge the potential significance or merit of a competent scientist's critique and to consider the desirability for further experimentation and confirmation." *Id.*

125 *Id.* at 206.

126 Although *Yee* purported to apply the *Frye* standard, the case was decided under the law of the Sixth Circuit, which "has overruled *Frye sub silento*." United States v. Porter, 120 DAILY WASH. L. RPTR. 477, n.79 (Super. Ct. D.C. 1991), *vacated and remanded*, 618 A.2d 629 (D.C. 1992) (citing C. MCCORMICK, EVIDENCE § 203, at 606 (3d ed. 1984)). Under the lenient standard of admissibility applied in *Yee*, the proponent of novel evidence need only produce "a set of experts" to testify "that the procedure [is] generally accepted." *Yee*, 134 F.R.D. at 165. If the court finds them persuasive, the novel proce-

A major theme of the prosecution experts in *Yee*, which the magistrate (and court) found persuasive, was that a multiple probe match (within the FBI's standards) is unlikely to occur if two samples are from different people,[127] and therefore the test is unlikely to incriminate an innocent person falsely, notwithstanding any uncertainties about match criteria.[128]

The issues raised in *Yee* have been replayed in several subsequent hearings on the admissibility of the FBI DNA test, with mixed results. Some courts have concluded that challenges to matching "involve considerations of the reliability of particular test results and not whether the FBI's techniques are generally accepted as capable of producing reliable results."[129] Consequently, such considerations "go to either the admissibility or the weight to be given the evidence in a particular case, not admissibility under *Frye*."[130]

Only one court has held that the dispute over the FBI's matching procedure precludes the admissibility of the test under *Frye*. In

---

dure is admissible "despite the *firmly held countervailing views of the opponent's experts*." *Id.* By contrast, under orthodox *Frye* analysis, a judge who observes a significant number of opposing experts with firmly held countervailing views should conclude that the scientific technique is not generally accepted and rule it inadmissible. *See, e.g.*, People v. Shirley, 723 P.2d 1354, 1377 (Cal. 1982) ("[T]he burden is on the proponent of the new technique to show scientific concensus supporting its use; if a fair overview . . . discloses that scientists significant in number or expertise publicly oppose [the technique] as unreliable, the court may safely conclude there is no such consensus at the present time.").

127 *Yee*, 134 F.R.D. at 174-77.

128 The prosecution position, adopted by the court, was that the value of a DNA match is necessarily high because the likelihood of a coincidental match is low. The defense position was that the value of a match on the FBI's system is impossible to evaluate because insufficient data exist to assess the likelihood of any particular match if the samples are from the same source, and there is reason to think the FBI *could* call matches between samples unlikely to be from the same source. For further discussion of this point, *see infra* notes 162-66 and accompanying text.

129 State v. Vandebogart, 616 A.2d 483, 492 (N.H. 1992) (FBI's DNA test inadmissible because statistical techniques fail *Frye* standard). The defense had raised challenges regarding the size of the FBI's match window, the lack of an objective criteria for matching, and the insufficiency of the FBI's validation. *Id.*

130 *Id.* at 492-93. A similar conclusion was reached in United States v. Jakobetz, 747 F. Supp. 250 (D. Vt. 1990), *aff'd*, 955 F.2d 786 (2d Cir. 1992) (FBI DNA test held admissible under relevancy standard). In that case, the district court noted that the concern that the FBI's matching criteria may allow poor quality matches "does not, in itself, render the matching criteria unreliable as a whole but rather provides fodder for effective cross examination when that condition occurs." *Id.* at 257. The defense expert "conceded that if the autorad matches in this particular case were within plus or minus 1% of the number of base pairs, he would have more confidence in the conclusion that there was in fact a match." *Id.* The prosecution pointed out that all comparisons that incriminated defendant Jakobetz were indeed within 1%, thereby sidestepping the criticism in that case.

People v. Halik,[131] a Los Angeles trial court concluded that "matching is an integral part of the total technology which, in a forensic setting, must pass the test of *Kelly-Frye*. . . . [Consequently,] [m]atching is the keystone in the archway to admissibility."[132] To determine whether the FBI's matching procedure passed muster, the court reviewed "perhaps the most comprehensive compilation of judicial and scientific materials ever assembled for a pre-trial hearing on these issues"[133] and found that scientific opinion on the FBI's matching procedure is split, with a majority of scientists who have publicly spoken on the issue taking the position that the FBI's matching procedures are unreliable and unacceptable.[134]

> There is a profound, significant and honestly-held disagreement among these men of science as to whether the protocol employed by the FBI to declare a match of DNA fragments between a known and an unknown source has gained general acceptance in their scientific community.[135]

No other court has followed *Halik*; however, no other defense counsel have raised challenges to the admissibility of the FBI's matching procedures based on such an extensive record. Thus, the debate over appropriate matching procedures is probably not resolved. Indeed, the NRC Report, with its emphasis on the importance of empirical validation of matching standards, is likely to add fuel to the debate about current procedures. Moreover, because the NRC Report makes it clear that the matching standards for each DNA typing system must be validated independently,[136] the debate is likely to arise anew as new systems are developed.

How should courts view disputes over matching? To address this question, we must consider two fundamental issues that underlie the matching debate: (1) whether objective match criteria are even necessary, and (2) what kind of empirical research, if any, is needed to validate the match procedure.

---

131 People v. Halik, No. VA00843 (Los Angeles Co., Cal. Super. Ct. Sept. 26, 1991) (Ruling and Order of C. Robert Simpson, Jr., J.).

132 *Id.* at 16-17.

133 *Id.* at 2. The court took judicial notice of the complete records of the *Frye* hearings in several previous cases, including State v. Anderson, 853 P.2d 135 (N.M. Ct. App.), *cert. granted*, 848 P.2d 531 (N.M. March 11, 1993); United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991); and State v. Despain, No. 15589 (Yuma Co., Ariz. Super. Ct. Feb. 12, 1991), as well as hearing live testimony.

134 *Halik*, No. VA00843, at 39.

135 *Id.*

136 NRC REPORT, *supra* note 12, at 54 ("The match criterion must be based on the actual variability in measurement observed in appropriate test experiments conducted *in each testing laboratory*.") (emphasis added).

D. THE NEED FOR OBJECTIVE MATCH CRITERIA

Proponents of DNA testing often argue that a trained analyst can reliably determine whether DNA prints match by simple visual inspection.[137] They suggest that objective match criteria, although perhaps helpful, are not really necessary.[138] Hence, concerns about whether matching rules have been adequately validated or appropriately followed have little to do with the fundamental reliability or validity of forensic tests.[139]

But many scientists think objective matching standards are necessary, and their position has been strongly endorsed by the NRC Report. The NRC Report declared that objective standards are "essential" both for scoring DNA prints[140] and for determining matches.[141] It acknowledged that subjective matching rules (e.g., comparison by eye) and the failure to adhere to a stated matching rule had been a source of considerable controversy, and rejected these practices as inappropriate for forensic DNA testing.[142]

Objective standards are needed, in part, because the results of forensic DNA tests may turn on distinctions among DNA prints too subtle to be reliably detected by visual inspection.[143] Objective

137 *See, e.g.*, Caldwell v. State, 393 S.E.2d 436, 442 (Ga. 1990) ("Lifecodes contended that a visual observation was adequate to declare a match."); Smith v. Deppish, 807 P.2d 144, 157 (Ka. 1991) ("One method of declaring a match is simply by visual determination . . . .").

138 A common argument is that formal match criteria are generally not used by scientists who perform RFLP analysis for research and genetic diagnosis and therefore should not be necessary in forensics. *See Lander, supra* note 3, at 501; Thompson & Ford, *supra* note 10, at 94-97; NRC REPORT, *supra* note 12, at 53-54.

139 "Whether or not there is a match between patterns produced by DNA samples extracted from two or more sources is primarily a qualitative judgment by a knowledgeable investigator based on a careful review of all information pertinent to the tests undertaken." Statement of the FBI Working Group on Statistical Standards for DNA Analysis (Aug. 13, 1990) (on file with William C. Thompson).

140 "There must be an objective and quantitative procedure for identifying the pattern of a sample." NRC REPORT, *supra* note 12, at 53.

141 "The match criterion . . . must be objective, precise and uniformly applied." *Id.* at 54.

142 *Id.* The NRC Report seems to imply that the concerns about matching arose only "in early cases." *Id.* But such concerns continue to be raised in litigation and in scholarly articles. The NRC Committee did not attempt a thorough review of the procedures actually in use in particular forensic laboratories. Committee members may not have realized the extent to which the subjective standards, which the Committee condemned, continue to be used and to cause problems in actual casework.

143 In research and diagnostics, scientists generally need to distinguish only three or four possible DNA prints, each of which is quite distinct and known in advance. In forensics, by contrast, there are, in theory, billions of possible DNA prints and the reliability of the test may depend on making fine grade distinctions among patterns that are quite similar. *See* NRC REPORT, *supra* note 12, at 53-54; Lander, *supra* note 3, at 501; Thompson & Ford, *supra* note 10, at 94-95.

match criteria are also needed in order to estimate the frequency of a match in a reference population. One cannot determine accurately what percentage of the DNA prints in a data base would "match" the print of a given individual, unless one can specify with precision what constitutes a match.[144]

The most important reason for using objective standards, however, is to prevent analysts from engaging in inappropriate bootstrap logic when evaluating DNA evidence. Commentators have noted a disturbing tendency for forensic analysts to resolve ambiguities in DNA patterns in a manner consistent with the expected result.[145] The analyst may, for example, infer that a discrepancy between two DNA profiles on one autorad *must* be an artifact (rather than a true genetic difference) because there is a match on the other autorads or, worse yet, because other evidence in the case suggests the two profiles have a common source.[146] Professor Eric Lander has condemned this kind of bootstrap interpretation in forensics because "one runs the risk of discounting precisely those differences that would exonerate an innocent defendant."[147] An analyst who too readily dismisses discrepancies in a DNA test that do not fit with other evidence can mistakenly conclude that weak, equivocal evidence is quite powerful, and thereby mislead the trier of fact.[148]

---

144 "[W]ithout an objective definition of a match, there is no meaningful way to determine the probability that a declared match might have arisen by chance." Lander, *supra* note 3, at 502-03. "[T]he probative value of the evidence is best gauged not by the frequency of prints that would match perfectly, but by the frequency of prints that would match under the looser standards actually used by the analyst." Thompson & Ford, *supra* note 10, at 97.

145 For example, Lander has suggested that Lifecodes' analysts failed to score potentially exculpatory bands in the *Castro* case because they were "influenced by making direct comparisons between lanes containing different DNA samples, rather than by considering each lane in its own right." Lander, *supra* note 3, at 502. Recognizing this danger, the NRC Report declared that "[i]t is not permissible to decide which features of an evidence sample to count and which to discount on the basis of a comparison with a suspect sample, because this can bias one's interpretation." NRC REPORT, *supra* note 12, at 53.

146 "Analysts not only know which samples are expected to match, but often are familiar with details of the case wholly unrelated to the forensic samples, such as the past criminal history of the suspect and the opinions of the police regarding likelihood of guilt. Where samples are submitted by a law enforcement agency, direct communication between the analyst and the detective handling the case frequently occurs. Fears have been voiced that when faced with ambiguity (whatever its technical cause), it may be difficult for the analyst not to be influenced by this type of information." Thompson & Ford, *supra* note 10, at 141.

147 Lander, *supra* note 3, at 502.

148 An analyst who ignores an otherwise troubling discrepancy on one autorad because there are good matches on other autorads is not treating the various bands in DNA prints as independent data. This practice invites overvaluation of the evidence, because the trier of fact receives statistics in connection with the DNA evidence that are

Bootstrap interpretation, whether intentional or unintentional, can best be prevented by the use of objective standards that remove the subjective component of matching.

A key issue now facing courts is whether to exclude DNA evidence from laboratories that fail to comply with the NRC Report's call for "objective and quantitative rules for identifying the pattern of a sample" and "a precise and objective matching rule for declaring whether two samples match."[149] In light of the NRC Report, litigants challenging the admissibility of DNA evidence can make a powerful showing that subjective interpretation of DNA prints is not generally accepted in the scientific community. They may also be able to show that subjective interpretation undermines the accuracy of statistical estimates and creates a serious danger of prejudice by allowing analysts in some cases to engage in inferential bootstrapping.

Such challenges have generally been unsuccessful so far, but have been deflected by arguments that fail to withstand close analysis. For example, in *State v. Jobe*,[150] the Minnesota Supreme Court dismissed concerns about subjectivity in the FBI's matching procedures with the observation that "each sample is also examined by a second trained examiner and ultimately the 'match' is confirmed or rejected through computer analysis using wholly objective criteria."[151] By this analysis, defendants have no cause to complain about subjectivity in the determination of matches because a match is not called unless it meets the "objective" quantitative match standards *as well as* the analyst's subjective standards; hence, subjectivity can never work against the defendant. This analysis is misleading for two reasons.

First, the "wholly objective" quantitative standards are applied

computed based on the assumption that the bands provide independent data. An analyst who takes into account the apparent strength of the case against the suspect when making critical judgments, such as what to make of an "extra band," is not treating the DNA evidence as if it were independent of other evidence in the case. This practice also invites overvaluation of the evidence in the case, because the trier of fact is unlikely to appreciate that what the analyst says about the DNA evidence depends partly on the strength of other evidence in the case. *See generally* Thompson and Ford, *supra* note 95. Such practices are not only prejudicial; they invade the province of the jury as well. Whether a discrepancy between two DNA prints should be discounted in light of other evidence incriminating the defendant, for example, is a judgment appropriately made by by the trier of fact, not by an analyst who is providing a putatively objective, independent interpretation of the DNA test.

149 NRC REPORT, *supra* note 12, at 72. The NRC Report characterizes such objective standards as "requirements" for DNA testing. *Id.*

150 486 N.W.2d 407 (Minn. 1992).

151 *Id.* at 420.

by the computer to data (band sizings) that are themselves determined subjectively.[152] For example, whether samples 18 and 25 in Figure 2 would be declared a match or not by a laboratory's "wholly objective criteria" depends entirely on how an analyst chooses to score the bands—which is to say that it depends on the analyst's subjective discretion. To label this process "wholly objective" is to give objectivity a new and perverse meaning.

Second, the use of quantitative match criteria does not eliminate the need for subjective judgments regarding the reasons for discrepancies in the number of bands.[153] Such discrepancies are common and require careful review because they are potentially exculpatory. The use of quantitative match criteria will *not* prevent an analyst from mistakenly attributing to experimental artifact a discrepancy that actually reflects a genetic difference between samples, nor will it prevent the analyst from making such determinations through inappropriate and prejudicial bootstrap inferences. In such cases, the analyst's subjective judgment clearly can work against the defendant, notwithstanding the application of quantitative match criteria.

On the other hand, problematic subjective calls may play a pivotal role only in a minority of cases.[154] If the case before a court is one in which experts all agree about the "match," it seems unreasonable to exclude the evidence on grounds that the match criteria are subjective. The best resolution to this dilemma may be the approach recommended in *Castro*: holding a pretrial hearing to determine, as a foundational matter, whether correct procedures were

152 The FBI arguably does not have the "objective and quantitative rules for identifying the pattern of a sample" called for by the NRC. *See, e.g., supra* note 81 and accompanying text concerning State v. Despain, No. 15589 (Yuma Co., Ariz. Super. Ct. Feb. 12, 1991). *But see The FBI's Responses to Recommendations by the Committee on DNA Technology in Forensic Science of the National Research Council, National Academy of Sciences*, 19 CRIME LABORATORY DIG. 49, 51 (1992) ("The [FBI] protocol include [sic] matching guidelines which are both objective and quantitative. However, the experience and expertise of the examiner remains an integral part of the interpretive process."). Subjectivity in the scoring of bands creates the potential for "fudge factors" in the determination of matches. *See supra* note 86.

153 *See supra* notes 88-89 and accompanying text.

154 While testifying in People v. Axell, 1 Cal. Rptr. 2d 429 (1991), Professor George Sensabaugh opined that two independent analysts, interpreting forensic autorads, would reach the same conclusions "[m]ore than nine out of ten times." *Id.* at 865. Professor Simon Ford, who has reviewed autorads in over 100 forensic cases, estimates that substantive disagreements about interpretation would occur in one-quarter to one-third of cases, with a higher rate for some laboratories than others. Telephone Interview with Professor Simon Ford (March 16, 1993). Thus, in the majority of cases, the match/no match determination may be straightforward, requiring minimal subjective analysis.

followed in the case at hand.[155] This approach was recently endorsed by the California Court of Appeal in *People v. Barney*.[156] The court acknowledged the NRC panel's call for objective matching standards, but concluded that the NRC report "does not equate the absence of a standardized rule with a lack of general acceptance as to the matching step of DNA analysis"[157] and hence does not render subjective tests inadmissible *per se* under the *Frye* standard.[158] However, the court left open to litigants the opportunity to raise foundational challenges to the admissibility of DNA evidence in particular cases on grounds that the laboratory failed to use "correct scientific procedures" for declaring a match.[159] Under this rule, a defendant presumably could challenge the admissibility of a result that depended heavily on a contested subjective call by an analyst. In light of the significant danger of prejudice that may arise from subjective interpretation of DNA tests, such challenges deserve to be taken very seriously by the courts.

E. THE NEED FOR EMPIRICAL VALIDATION OF THE MATCH CRITERIA

A second fundamental issue for courts considering the admissibility of DNA profile evidence is what kind of research, if any, is necessary to validate the laboratory's quantitative match criteria. A number of commentators have argued that validation research is essential,[160] and the NRC Report strongly endorsed this position.

> The match criterion must be based on reproducibility studies that show the actual degree of variability observed when multiple samples from the same person are separately prepared and analyzed under typical forensic conditions. . . . Each testing laboratory must carry out its own reproducibility studies, because reproducibility varies among

---

155 *See supra* note 99 and accompanying text.

156 10 Cal. Rptr. 2d 731 (Cal. Ct. App. 1992).

157 *Id.* at 739.

158 This conclusion is, of course, debatable. It seems difficult to maintain that subjective matching procedures are acceptable when a distinguished national panel, appointed by the National Research Council to examine the issue, has stated that objective standards are "necessary" and "required". On the other hand, the NRC panel also recommended that DNA analysis "be continued while improvements and changes suggested in this report are being made." NRC REPORT, *supra* note 12, at *x*. The NRC Report can be read to suggest that the "improvements and changes" suggested therein are not necessary for general acceptance of forensic DNA tests at present but may become necessary in the near future: "[a]fter a sufficient time for implementation of quality-assurance programs has passed, courts should view quality control as necessary for general acceptance." *Id.* How soon that future will come, if ever, remains to be seen.

159 "The use of match criteria in a given case is properly addressed as part of the inquiry whether " 'correct scientific procedures were used in the particular case.' " *Barney*, 10 Cal. Rptr. 2d at 739-40.

160 *See, e.g.*, Shields, *supra* note 10; Lander, *supra* note 3; Lander, *supra* note 87.

laboratories.[161]

But some proponents of DNA evidence have dismissed such research as unimportant.[162] It is foolish to quibble about the precise size of the matching criterion, they say, because under any criterion now in use there is an extremely low likelihood of a false match. Consequently, a match provides powerful evidence that the samples are from the same person regardless of whether the match criterion has been "validated." According to this argument, validation of the match criteria has little to do with the fundamental reliability or validity of forensic DNA tests.

The problem with this argument is that it focuses on only one of two factors that affect the probative value of DNA evidence. The value of DNA evidence for proving that two samples came from the same person depends, in part, on the probability of a particular test result if the samples are from *different* people—a probability likely to be low if the test indicates that the samples "match." But its value also depends on the probability of a particular test result if the samples are from the *same* person; if this probability is equally low, then the DNA evidence deserves no weight.[163]

---

161 NRC REPORT, *supra* note 12, at 61-62.

162 *See supra* notes 127-28 and accompanying text.

163 This argument casts the evaluation of DNA evidence in a Bayesian framework. *See generally* RICHARD LEMPERT & STEVEN SALTZBURG, A MODERN APPROACH TO EVIDENCE 148-53 (1st ed. 1977); John Kaplan, *Decision Theory and the Factfinding Process*, 20 STAN. L. REV. 1065 (1969); Richard Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977); David H. Kaye, *What is Bayesianism? A Guide for the Perplexed*, 28 JURIMETRICS J. 161 (1988); David Schum & Anne Martin, *Formal and Empirical Research on Cascaded Inference in Jurisprudence*, 17 LAW & SOC'Y REV. 105 (1982). It posits two mutually exclusive and exhaustive hypotheses about the relationship between two samples: *H1*—the samples are from the same person; *H2*—the samples are from different people. In accordance with standard Bayesian analysis, it is proposed that the value of a particular DNA test result, *D*, depends on the likelihood ratio $p(D/H1)/p(D/H2)$. The higher this ratio, the greater the probative value of the evidence.

If *D* indicates that the two samples have DNA prints that are similar, then $p(D/H2)$ will probably be low because DNA prints are so variable that different people are unlikely to have similar ones. However, $p(D/H1)$ may be as low or lower than $p(D/H2)$ if the two prints, although similar, have discrepancies that are unlikely to be observed in prints of the same person. To evaluate DNA evidence, then, one must know whether the discrepancies between two DNA patterns are within the normal range of variation (and therefore likely if the samples are from the same source), or beyond the normal range of variation (and therefore unlikely if the samples are from the same source).

The need to vary the likelihood ratio to take into account the quality of the match has been recognized by a few experts working within a Bayesian framework. *See* Evett et al., *supra* note 76, at 314-315 (1992); Donald Berry et al., *Statistical Inference in Crime Investigations Using DNA Profiling*, 41 APPLIED STAT. 499 (1992); Donald Berry, *DNA Fingerprinting: What Does It Prove*, *in* 3 CHANCE: NEW DIRECTIONS IN COMPUTING AND STAT. 15 (1990) ("If all the suspect's band weights are very close to the criminal's, then the numerator of the Bayes factor will be large. One discrepant band weight will serve to

FIGURE 3
DNA PROFILES OF VICTIM, DEFENDANT AND EVIDENTIARY SAMPLE IN *STATE v. BRUNO*

Probe	Victim	Defendant (Bruno)	Evidentiary Sample	% Difference (Bruno v. Evid)
D2S44	3020	2352	2328	+ 1.03
	1613	1643	1619	+ 1.47
D4S139	7268	11948	12622	− 5.49
	2789	5602	5457	+ 2.62
D17S79	1532	1539	1525	+ 0.91
	1429	1352	1337	+ 1.11
D1S7	2461	11072	11139	− 0.60
	1532	5342	5296	+ 0.86

Suppose, for example, that the FBI declares a match between a band in defendant's DNA profile and a band in an evidentiary sample and that the two bands differ in measured size by 4% (putting them well within the agency's quantitative matching criteria). Suppose further that only one person in twenty in a suitable reference population has a band that would "match" the band in the evidentiary sample by the FBI's criteria.[164] Is this match incriminating? That depends on how common a 4% discrepancy is between bands in two profiles from the same source. If discrepancies as large as 4% are common, then the match is certainly incriminating. If, however, discrepancies as large as 4% are rare, the "match" may have little value, or even be exculpatory.

A similar analysis is necessary to evaluate multi-band matches. Let us consider the DNA profiles represented in Figure 3, two of which were declared to match by the FBI in a North Carolina case, *State v. Bruno*.[165] The profiles show the estimated "size," in number of base-pairs, of the bands in the defendant's DNA print, in a print of an evidentiary stain, and in the victim's DNA print. Four probes were used, each of which produced two of bands in each sample. It

decrease the numerator . . . by an amount that depends on the degree of discrepancy."). However, these experts have, thus far, focused their Bayesian modeling solely on discrepancies in the estimated size of bands; they have not yet dealt with the potentially more severe problems arising from inconsistencies in number of bands and uncertainties in the initial scoring of bands.

164 For the sake of simplicity, let us assume that only one band can be visualized in the evidence.

165 Nos. 89 CR 7443, 7444 & 7445 (New Hanover Co., N.C., Super. Ct., 1990). Band sizes are those indicated in the FBI's report on the case, dated August 30, 1989 (FBI File No. 95-288752).

is undoubtedly true that the profile of a person who was not the source of the evidentiary sample would be unlikely to "match" as well as did Mr. Bruno's profile.[166] But the analysis cannot stop there. To know what to make of the evidence, one must also know how likely is it that the two "matching" patterns would be so discrepant if Mr. Bruno *is* the source of the evidence. If this probability is also low, the evidence may have little value.

Forensic laboratories do not attempt to make such probability estimates in each case. Instead, they trust the quantitative match rules to assure that a match is not called between samples unlikely to be from the same person. One cannot determine whether the quantitative rules perform this function adequately, however, unless one knows the results of empirical reproducibility studies of the kind demanded by the NRC Report.

Consequently, courts should take very seriously challenges to DNA evidence that allege that a laboratory's match criterion is inadequately validated. Such challenges go to the fundamental value of the evidence. Without reproducibility studies of the type specified in the NRC Report, one cannot determine the likelihood of a given "match" if the samples are from the *same* person, and one therefore has no scientific foundation for assessing the value of the evidence in a given case. In light of the NRC Report, a strong argument can be made that such evidence fails to meet the standards of the scientific community and is inadmissible under the *Frye* standard. Such evidence arguably should be inadmissible under the *Daubert* standard as well because it is inadequately validated.[167]

A somewhat different issue is raised by the claim that a laboratory's match criteria is too broad. This claim is usually raised against a laboratory that has done some reproducibility studies but has failed to draw its match criteria as narrowly as critics think is appropriate in light of those studies. Courts should consider carefully whether such claims go to the admissibility or merely the

---

166 Just how unlikely such a coincidental match might be is an issue on which experts may well disagree. *See infra* Part III for a discussion of this issue.

167 "Proposed testimony must be supported by appropriate validation—i.e., 'good grounds' based on what is known." Daubert v. Merrell Dow Pharmaceuticals, Inc. 113 S. Ct. 2786 (1993). *See also supra* notes 46-54 and accompanying text. Suppose, for example, that an expert testified that the evidence in the *Bruno* case was powerfully incriminating. If the expert had no scientific basis for evaluating the likelihood of the observed discrepancies between the evidentiary DNA profile and Mr. Bruno's profile (if Bruno was the source of the evidence), then such testimony would be based on speculation rather than scientific knowledge. Such testimony should be inadmissible because it is unscientific and it may greatly overstate the value of the DNA evidence, thereby proving prejudicial.

weight of the evidence. The proper analysis may vary depending on the facts of the case.

In some cases, a defendant may argue that the match criteria should have been drawn in a narrower manner that excludes him. If the defendant makes a strong showing prior to trial that the observed discrepancies are unlikely to have arisen if the "matching" prints have a common source,[168] then exclusion of the DNA evidence might well be appropriate. In such cases, the DNA evidence is likely to deserve considerably less weight than a jury would give it.[169] To challenge the DNA evidence before the jury would be a formidable task, and arguably would impose an unfair burden on a criminal defendant.[170] The evidence could properly be excluded under Federal Rule 403.[171]

---

168 In formal terms, the defendant should make a preliminary showing that $p(D/H_1)$ is quite small.

169 *See supra* note 163 and accompanying text.

170 The defendant would need to present complex testimony concerning the operating characteristics of the DNA typing system, the variability of band measurements in the system, and the consequent statistical likelihood of the observed data assuming a common source. The defendant would also need to educate the jury concerning the implications of such data for the value of the DNA evidence—a challenging task in itself. To judge the probative value of the DNA evidence, the jury would need to estimate the probability of the observed data, assuming a common source, and then balance that probability against estimates of the probability of a "match" if the samples have a different source. Whether lawyers and lay jurors are up to such a task is an important issue to consider when evaluating the admissibility of such evidence.

171 A contrary position was taken by the court in United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991), which concluded that such arguments go to weight rather than admissibility: "defendants who would be outside a smaller window but are within the FBI's larger window can make that point at trial." *Id.* at 208. Professor David Kaye goes even farther. He has argued, in a widely circulated but as yet unpublished manuscript, that such testimony should be excluded altogether because it is irrelevant. "If the match, as determined by the prosecution's rule, would be rare among innocent people, then the evidence proves something, and its probative value is unaffected by the tautology that the same measurements would not match under some even more stringent rule." David H. Kaye, DNA Evidence: Probability, Population Genetics and the Courts, 12 (1993) (on file with William C. Thompson).

Whether the *Yee* approach is correct is an issue on which reasonable people may differ, depending on how confident they are in the ability of lawyers and jurors to deal with complex scientific and statistical arguments. *See supra* note 166. But Professor Kaye's argument is wrong. If the defendant can show that the discrepancies that place him outside the "more stringent rule" are so large as to indicate a low likelihood that the samples are from the same source, then the value of the DNA evidence is significantly reduced. *See supra* note 163. In a personal conversation, Professor Kaye recently acknowledged this point and promised to rethink his position. Telephone interview with Professor David Kaye (Sept. 20, 1993).

Professor Kaye's equivocation on this issue is instructive. He is a leading authority on scientific and statistical evidence and has written a number of influential articles about the use of genetic evidence in court. That someone of his intellect and training can be uncertain about a question going to the fundamental value of DNA evidence

In other cases, defendant may be able to show that there is a controversy about the appropriate size of the matching window but be unable to show that the narrower window advocated by critics would have excluded him.[172] Such a showing should not, by itself, preclude admissibility of the DNA evidence. Informing the jury that the defendant is a member of an overly broad class is not unfair, so long as the size of the class is accurately reported.

III. THE DEBATE OVER THE VALIDITY OF STATISTICAL ESTIMATION METHODS

A. OVERVIEW

The third step of DNA profiling is statistical estimation. After determining that two DNA samples match, forensic analysts estimate the statistical frequency of such matches in a reference population. The purpose of the statistical estimates is to provide meaning to the match by showing the likelihood that an unrelated person in the reference population would match by chance.

Statistical estimation has been the most controversial aspect of DNA testing. The great majority of courts that have held DNA evidence inadmissible have done so based on the existence of a scientific dispute over the validity of the forensic laboratories' frequency estimation methods.

To estimate the frequency of a DNA profile in a reference population, forensic analysts first estimate the frequency of each allele (band) in the DNA profile by determining its frequency in a data base containing DNA profiles of a number of individuals.[173] These data bases consist of convenience samples drawn primarily from blood banks, with separate data bases for major racial and ethnic groups (Hispanics, non-Hispanic Caucasians, African-Americans, Asians). Analysts then combine the estimated frequencies of the individual alleles to determine the overall frequency of the DNA profile, using formulae that assume the alleles are statistically independent.[174]

---

demonstrates just how difficult it is to evaluate such evidence. It also demonstrates the strong potential for prejudice that exists when problematic DNA matches are allowed to go to the jury. A juror who adopted the seductive but erroneous reasoning in Kaye's article might greatly overvalue DNA evidence in a case like State v. Bruno, Nos. 89 CR 7443, 7444 & 7445 (New Hanover Co., N.C. Super. Ct., 1990), *aff'd*, 424 S.E.2d 440 (N.C., 1993), where the "quality" of the match is poor.

172 *See* discussion of United States v. Jakobetz, 747 F. Supp. 250 (D. Vt. 1990), *aff'd*, 955 F.2d 786 (2d. Cir. 1992), *supra* note 126.

173 *See generally* NRC REPORT, *supra* note 12, at ch. 3.

174 Statistical independence means that the likelihood of a person having a particular

Critics have challenged these procedures on several grounds. One problem, say critics, is that the forensic laboratories tend to underestimate the frequency of matching alleles in their databases, and thereby greatly underestimate the overall frequency of DNA profiles.[175] A more fundamental concern is that the procedure fails to take into account the possibility that there is significant variability among population subgroups in the frequency of alleles. Critics suggest that within major groups, such as Caucasians, Hispanics and Blacks, the frequency of alleles may differ among various ethnic, religious or geographic subgroups. If such variability exists, there are two important implications.[176] First, the convenience samples used by the forensic laboratories may be unrepresentative of the population in particular locales. Second, the assumption that the frequency of alleles is statistically independent would be invalid.

Concerns about the accuracy of statistical estimates were raised in early litigation on the admissibility of forensic DNA evidence,[177] but were voiced by only a few experts in a handful of cases and had little impact. Before 1991, statistical estimates of forensic laboratories were routinely ruled admissible in most cases;[178] typically, the defense failed to present a single expert to challenge them. In cases where the defense did muster experts, courts often found them less persuasive than the supporting experts presented by the prosecution, and concluded that the critics represented the viewpoint of only a small minority.

---

allele is not affected by what other alleles the person has. The probability of a series of independent events is the product of their frequencies, and hence will be quite low when all the frequencies are low. For example, the probability of rolling "one" eight times in a row with a fair die is $(1/6)^8 = .000000595$, or approximately one in 1.6 million. The probability might be considerably higher if the probability of rolling "one" on each throw is not independent of the other throws, as would be the case, for example, it it were not a fair die. Whether the alleles in DNA profiles are in fact statistically independent is a central issue.

175 *See infra* notes 196-206 and accompanying text.

176 *See generally* NRC REPORT, *supra* note 12.

177 *See* People v. Wesley, 533 N.Y.S.2d 643 (N.Y. Sup. Ct. 1988) (Lifecodes' test); State v. Schwartz, 447 N.W.2d 422 (Minn. 1989) (Cellmark's test); *see also* Christopher Joyce, *High Profile: DNA in Court Again*, NEW SCIENTIST, July 21, 1990, at 24 (describing the population genetics issues raised in early 1990 in State v. Anderson, 853 P.2d 135 (N.M. Ct. App. 1991), regarding the FBI's test).

178 One exception was *Schwartz*, 447 N.W.2d at 422, in which the Minnesota Supreme Court held statistical estimates inadmissible in connection with a Cellmark test. The defense in *Schwartz* mounted a serious challenge to the statistical estimates. However, this ruling may stem primarily from Minnesota's longstanding rule limiting the use of statistical evidence in connection with genetic evidence based on fears that such evidence will be misused by the jury. *See generally* William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence?* 52 LAW & CONTEMP. PROBS. 9, 26-28 (1989) (reviewing Minnesota caselaw on this issue).

In *People v. Axell*,[179] for example, the court chose to credit three experts who testified that Cellmark's statistical procedures are generally accepted over three experts who testified that they are not, suggesting that the prosecution experts had convincingly responded to the defense experts' concerns.[180] In *United States v. Yee*, a federal magistrate took a similar tact:

> [D]espite the prestige, standing, and expertise of the witnesses who share the view that the scientific community could not and would not find the FBI's database and resulting probability estimates acceptable, the view of the government's witnesses about the level of acceptance, when all factors are taken into account, is more likely to be the accurate view.[181]

Over time, however, the ranks of the scientific critics grew[182] and their concerns began to be taken more seriously by courts. In 1991, a number of trial courts[183] and one state supreme court[184] ruled DNA evidence inadmissible based on concerns about the statistical procedures.

This trend gained momentum after two leading population geneticists, Richard Lewontin and Daniel Hartl, published an article in

---

179 1 Cal. Rptr. 2d 411 (Cal. Ct. App. 1991).

180 It is ironic that the court, in reaching this conclusion, cited Dr. Kidd's unpublished (and at that time unavailable) data on the "very small differences" in allele frequencies among various subgroups—the very data that, when later reanalyzed by Dr. Mueller, showed rather large differences among American Indian groups. *See infra* note 253 and accompanying text. Defendant Axell was partly American Indian. The court also relied on testimony by Professor Conneally who "opined that the probes used by Cellmark are in linkage equilibrium because they are on different chromosomes . . . ." *Id.* at 852. This position, however, is no longer widely accepted. *See* NRC REPORT, *supra* note 12, at 79-80.

181 United States v. Yee, 134 F.R.D. 161, 206 (N.D. Ohio 1991).

182 In July 1991, Professor Charles Taylor, a population geneticist at UCLA, conducted a survey of others in his field to determine their positions on the acceptability of the statistical estimation methods used by forensic laboratories. Charles Taylor, Survey of Population Geneticists Concerning Methods for Calculating Matches in Forensic Applications of VNTR Loci (July 9, 1991) (unpublished manuscript, on file with William C. Thompson). Taylor identified 30 academic scientists with expertise in population genetics who had taken a position on the issue. By his count, 11 supported the forensic labs and 19 did not. Moreover, the ratio of critics to supporters was higher among those whom Taylor rated as better known in the field (based on their work having been cited in major textbooks). *Id.* The results of the survey were introduced in evidence in People v. Halik, No. VA 00843 (Los Angeles Co., Cal. Dist. Ct. May 13, 1991), and later were made part of the record before the California Court of Appeal in People v. Barney, 10 Cal. Rptr. 2d 731 (Cal. Ct. App. 1992).

183 State v. Despain, No. 15589 (Yuma Co., Ariz. Super. Ct. Feb. 12, 1991); State v. Fleming, Nos. 90-CR-2716 & 90-CR-5546 (Cook Co., Ill. Cir. Ct. Mar. 12, 1991); State v. Hummert, Nos. CR 90-05559 & CR 90-03684 (Maricopa Co., Ariz. Super. Ct. Apr. 16, 1991); State v. Passino, No. 185-1-90 (Franklin Co., Vt. Dist. Ct. May 13, 1991); United States v. Porter, 120 DAILY WASH. L. RPTR. 44 (D.C. Super. Ct. Sept. 20, 1991).

184 Commonwealth v. Curnin, 565 N.E.2d 440 (Mass. 1991).

Science, in December 1991, that raised concerns about population structure and concluded, in light of those concerns, that the statistical estimation methods used by forensic laboratories are "unjustified and generally unreliable."[185] Because of the prominence of the authors, this article attracted a great deal of attention and ignited a furious academic debate over how much population structure exists and the extent to which population structure undermines the validity of the statistical estimation methods used by forensic laboratories.[186]

Further debate was inspired by the publication of the NRC Report in April 1992. The NRC panel did not take sides in the dispute over population structure; it found existing empirical data insufficient to resolve the question. But the report concluded that the concerns raised by Lewontin, Hartl and others were sufficiently serious that the statistical estimation methods developed by the forensic laboratories should not continue to be used because these methods might greatly underestimate the frequency of DNA profiles. Instead, the NRC proposed a more conservative method that it dubbed "the ceiling principle."[187]

A number of courts thereafter held DNA evidence inadmissible on grounds that the statistical methods of the forensic laboratories were not generally accepted.[188] Some expressed hope that a scientific consensus would soon emerge in favor of methods based on the NRC's ceiling principle.[189] Those hopes, however, now seem to be fading as it "appears that the level of debate has only increased as a result of [the NRC] report."[190]

---

185 Lewontin & Hartl, *supra* note 10, at 1745.

186 *See* Roberts, *supra* note 3, at 1721.

187 NRC REPORT, *supra* note 12, at 82-85.

188 Opinions citing the NRC Report as evidence of a dispute over the validity of forensic statistical estimation methods include Commonwealth v. Lanigan, 596 N.E.2d 311 (Mass. 1992); People v. Barney, 10 Cal. Rptr. 2d 731 (Cal. Ct. App. 1992); State v. Vandebogart, 616 A.2d 843 (N.H. 1992); State v. Anderson, No. 12899, 1993 WL 135835 (N.M. Ct. App. Jan. 28, 1993); United States v. Porter, 618 A.2d 629 (D.C. 1992); People v. Atoigue, No. CR 91-95A (Guam Dist. Ct. App. Div. Sept. 11, 1992); People v. Wallace, 17 Cal. Rptr. 2d 721 (Cal. Ct. App. 1993); State v. Bible, No. CR900167AP, 1993 WL 306544 (Ariz. filed Aug. 12, 1993).

189 *See Barney*, 10 Cal. Rptr. 2d 731, 745 ("The NRC report on DNA analysis appears to point the way to . . . common ground."). In *Lanigan*, 596 N.E.2d at 311; *Vandebogart*, 616 A.2d at 843; and *Porter*, 618 A.2d 629; the courts remanded cases in which defendants had been convicted based on standard statistical estimates to determine whether the ceiling principle is generally accepted and, if so, whether it would have produced different numbers than those used against the defendant.

190 Laurence Mueller, *The Use of DNA Typing in Forensic Science*, ACCOUNTABILITY RESEARCH 2 (1993); *see also* People v. Wallace, 17 Cal. Rptr. 2d 721, 725 ("recent developments have shown that general acceptance may not be easily achieved").

In retrospect, the judgment of the court in *Axell* court and the magistrate in *Yee* appear premature. Over time, as increasing numbers of critics emerged, and as their views gave rise to serious debate in the academic community, the conclusion that the prosecution experts speak for the scientific community, and that the defense experts do not, has become more difficult to maintain.[191] Recently, courts in *Frye* jurisdictions have been more hesitant about deciding which scientific views to credit in the face of this dispute. The views of the District of Columbia Court of Appeal reflect this new, and overdue, trend toward judicial humility:

> We specifically decline the government's invitation to hold that the position of one group of distinguished scientists (those favoring the government's position) is more persuasive, as a matter of molecular biology or population genetics, than the position of an apparently equally distinguished group of scholars who have reached an opposite conclusion; indeed, we view the government's position on this issue as contrary to *Frye*.[192]

Courts in relevancy and *Daubert* jurisdictions do not have the option of waiting for scientific consensus to emerge. They must shoulder the burden of assessing the validity and helpfulness of proferred statistical estimate in the midst of the current scientific turmoil. The following sections lay out the major scientific and legal issues.

B. DETERMINING THE FREQUENCY OF INDIVIDUAL BANDS

To determine the frequency of each band (allele) in the DNA print, forensic analysts estimate the percentage of bands in a data base that would "match" the band in question. Typically, the laboratory counts all bands in the data base that fall within a range of sizes; this range is designated a "bin." Some laboratories use "floating bins" keyed to the band in question. For example, to estimate

---

191 The issues raised by the defense have been quite consistent over time. The change in judicial attitude would appear to stem solely from the growing number and prestige of the scientists voicing critical views. *See, e.g., Barney*, 10 Cal. Rptr. 2d at 731 (noting that "the challenges asserted by Howard and Barney [based on the testimony of Laurence Mueller in 1989] are essentially the same as the points raised by Lewontin and Hartl").

192 United States v. Porter, 618 A.2d 629 (D.C. 1992). In *Porter*, the government argued that the weight of scientific authority favored current statistical procedures, because over 18 peer-reviewed articles and 10 published letters, written by 45 scientists, supported the methodology used by the government, while only six articles and seven letters, authored by only 12 scientists, questioned the approach. *Id.* at 638. The court properly responded that "the government asks this court to choose between scientists on the basis of rather unimpressive numbers, and thus to make precisely the kinds of determinations as to which *Frye* requires a consensus of experts." *Id.*

the frequency of a band of 1000 base pairs, Lifecodes counts all bands that fall within ± 1.8% of its size—that is, all bands in the data base which have an estimated size between 982 and 1018 base pairs.[193] Because all of these bands fall within Lifecodes' match criteria, they all are bands that potentially could be "matched" with the band in question. Other laboratories use "fixed bins" which are established in advance and used in each case. For example, the FBI divides the full range of band sizes into 31 fixed bins.[194] A band of 1000 base pairs would fall into bin number five, which includes bands from 964 to 1077 base pairs.[195] To estimate the frequency of the 1000 base pair band, the FBI would count all bands falling in bin number five.

A major concern of critics is that forensic laboratories underestimate the frequency of matching bands. Some forensic laboratories have been criticized for using bins that are narrower than their match criteria,[196] a practice the NRC Report called "unacceptable."[197] For example, Lander reported that in *People v. Castro*, Lifecodes used a bin for allele frequency estimates of only ±2/3 standard deviation, while using a matching criteria of ±3 standard deviations.[198] Similarly, the FBI has computed statistics in some cases based on floating bins of ±2.5%, while employing a matching criteria of ±5%.[199] Because the frequency of each allele is multi-

193 Shields, *supra* note 10, at 19.

194 The FBI championed the fixed bin method. *See* Budowle et al., *supra* note 86, at 845-48. However, more recent FBI publications indicate that the Bureau will abandon the fixed bin approach in favor of smaller floating bins (±2.5%) if it is required to comply with the "conservative" statistical estimation procedure recommended by the NRC. Bruce Budowle & Keith Monson, *The Approach Used by the FBI for Calculating Ceiling Frequencies*, 19 CRIME LABORATORY DIG. 84 (1992). The floating bins have been used to compute statistical frequencies in some recent cases.

195 The FBI's fixed bins do not correspond exactly to the FBI's quantitative match criteria. For example, the 1000 base pair bands could potentially be called a match with bands from 950 to 1050 base pairs under the FBI's 5% criterion.

196 Mueller, *supra* note 190, at 3 ("The size of the floating bins that have been used by some labs are only about half the appropriate size, thereby producing allele frequency estimates that are too small.").

197 NRC REPORT, *supra* note 12, at 78.

198 Lander, *supra* note 3, at 504. Following People v. Castro, 545 N.Y.S.2d 985 (N.Y. Sup. Ct. 1989), Lifecodes abandoned this practice and adopted a floating bin the same size as its quantitative match standard.

199 If a band of 1000 base pairs is found in an evidentiary sample, for example, the FBI's quantitative match rule allows a match to be declared to bands falling within the range 951-1050 base pairs. But in some cases the FBI has estimated allele frequencies using floating bins of ±2.5% of the average value of the bands being compared. *See supra* note 175. Thus, if the 1000 base pair band were matched to a band at 1050 base pairs, the FBI's bin would range approximately from 999-1051, a range only half as large as the range within which a match would be called. By one account, the FBI has recently acknowledged this problem and has doubled its bin size to avoid repeating this

plied to obtain the frequency of the overall DNA print, systematic underestimates of the frequency of each allele can cause severe errors. With regard to Lifecodes' procedure in *Castro*, Lander states: "[f]or a three-locus genotype, the error may thus be about 8000 fold."[200] Lander analogized the practice to "catching a match with a 10-foot-wide butterfly net, but then attempting to prove the difficulty of the feat by showing how hard it is to catch matches with a 6-inch-wide butterfly net."[201]

Although most laboratories have now adopted bins that are the same size as their match criteria, some critics question whether even that is sufficient. To count all bands in the data base that would "match" a given band, say critics, it may be necessary to use bins larger than the match criteria. They base this claim on evidence that the variability of band measurement (measurement error) is significantly greater among bands in data bases than among bands examined in forensic casework. Two samples that would produce bands within 5% of each other in case work may produce bands that are more than 5% apart if the samples were part of the data base because measurements taken on samples in the data base are more variable.[202] Hence, to capture all bands that would be within a lab's ±5% match criteria, it may be necessary to use a bin larger than ±5% when searching the data base for matching bands.[203] In recognition of this problem, one laboratory, Cellmark Diagnostics, has recently begun using bins that are larger than its match standards. Other laboratories have not chosen to follow this example.

---

error in future casework. Interview with Professor Bruce Weir, Dept. of Statistics, N.C. State Univ., Tempe, Ariz. (Mar. 19, 1993).

200 Lander, *supra* note 3, at 504.

201 *Id.*

202 The greater variability is due in part to the fact that samples in the data bases were typed over time, often by different analysts, in a series of separate assays on separate gels. By contrast, in most forensic casework, all samples are run by the same analyst as part of the same assay on the same gel.

203 This claim appears to have been first raised in State v. Pennell, 584 A.2d 513 (Del. Super. Ct. 1989). Experts for the defense argued, and the court agreed, that Cellmark should use a larger bin for allele frequency estimation than it used for declaring a match between DNA prints run on the same gel, because "the data base was generated using different gels." *Id.* at 518. Cellmark had tacitly acknowledged that its measurement error is greater across samples run on different gels than among samples run within the same gel by adopting a broader matching criteria for between-gel than within-gel comparisons.

This claim was also raised in People v. Axell, 1 Cal. Rptr. 2d 411 (Cal. Ct. App. 1991). A Cellmark expert testified that using a larger bin in order to account for between gel variability in the data base would change the frequency from one in six billion to one in three or four million. The court stated that this matter "would appear to go to the weight of the evidence more than its admissibility." *Id.* at 429.

Finally, some critics argue that forensic laboratories may underestimate allele frequencies by failing to take into account sampling error—that is, the tendency for the allele frequency observed in a sample to differ from the true frequency due to the operation of chance in the selection of a sample. Lander's comments are particularly relevant:

> Sampling error poses a particularly serious problem when estimating the frequency of a very rare event, such as an allele at a hypervariable RFLP locus. In a sample size of 500, an allele whose observed frequency was 1 in 500 might have a true frequency (taking a 99% confidence interval) of nearly 7 in 500. If this correction were neglected, the odds of finding the allele would be underestimated by a factor of nearly 7. If this were neglected for both alleles at [three loci] the total chance of finding the observed genotype would be underestimated by a factor of $7^6 = 118{,}000$.
>
> Fortunately, statistical methods for incorporating sampling error are well developed. If conclusions must be proved beyond a reasonable doubt, it might be wise to use the 99% upper limit of the confidence interval for each allele.[204]

Whether such adjustments for sampling error are necessary and appropriate has been controversial. Some forensic scientists have argued that upper confidence limit estimates are not necessary because the procedure for allele estimation is sufficiently "conservative" to overestimate the frequency of matching alleles even without such a correction.[205] Furthermore, some statisticians disagree with Lander's suggestion that the upper confidence limit of the frequency of *each allele* be used in computing the frequency of the DNA profile, prefering to make the adjustment for sampling error after the allele frequency estimates are combined, rather than before,[206] a procedure likely to produce smaller adjustments.

C. DETERMINING THE FREQUENCY OF GENOTYPES

After determining the frequency of each band, the next step is to determine the frequency of genotypes. A genotype is the pair of

---

[204] Eric Lander, *Population Genetic Considerations in the Forensic Use of DNA Typing*, *in* DNA TECHNOLOGY AND FORENSIC SCIENCE 146-47 (Jack Ballantyne et al. eds., 1989); *see also* Shields, *supra* note 10, at 20.

[205] For example, the FBI has taken the position that its fixed bins overestimate allele frequencies by a sufficient margin that it is unnecessary to apply upper confidence limits to them. Defense experts point out, however, that some of the FBI's bins are in fact not much larger than its matching criteria. Shields, *supra* note 10, at 24. For example, the size of bin five differs little from the match criteria around a band of 1000 base-pairs.

[206] Weir, *supra* note 15, at 11657 ("Setting confidence limits on estimated profile frequencies is the best way to convey the effects of the sizes of the databases on which the estimates are based."). In other words, Weir would compute an upper confidence limit for the DNA profile, not for the constituent alleles in the profile.

alleles (bands) produced by a given probe. As noted earlier, one of these alleles is inherited from the mother and one from the father. To determine the frequency of heterozygous (two band) genotypes, forensic DNA laboratories use the formula $2pq$, where p and q are the frequency of the two alleles in the genotype.[207] For example, if the frequency of band A is .03 and the frequency of band B is .05, the laboratory will multiply $.03 \times .05 \times 2$ and conclude that the frequency of the genotype AB is .003 (three in 1000).[208] This formula assumes the frequencies of band A and band B are statistically independent, and may significantly underestimate the frequency of genotypes if the allele frequencies are not independent.

When alleles at any genotype are statistically independent in a particular population, the population is said to be in Hardy-Weinberg equilibrium.[209] Whether U.S. populations are in Hardy-Weinberg equilibrium has been a major issue in the debate over DNA statistics. Critics suggest that Hardy-Weinberg equilibrium may not hold due to endogamous mating patterns—that is, a tendency for people to mate with others having the same subset of possible alleles. For example, if people with allele A are more likely to mate with those having allele B, than with those having alleles C or D, then genotypes AA and AB would be more common (and genotypes AC and AD less common) than suggested by the formula $2pq$. Endogamous mating might occur without people having any awareness of it. If alleles A and B were common in a particular population subgroup, and if people in that subgroup tended to mate with each other, genotype AB would be more common in the general population, and much more common among members of that subgroup, than predicted by applying the formula 2pq to allele frequencies for the general population.[210]

The final step in the statistical procedures is to determine the frequency of the entire DNA profile, which is sometimes called a multi-locus genotype. The forensic DNA testing laboratories do this by multiplying together the frequencies of the genotypes. If

---

207 *See generally* NRC REPORT, *supra* note 12, at 76-85.

208 The product of the individual allele frequencies is multiplied by two because there are two ways a person can get a given genotype. A person may have genotype *AB* as a result of receiving *A* from his father and *B* from his mother, or vice versa. By analogy, there are two ways to roll number eleven with a pair of dice: a five on the first die and a six on the second, or vice-versa. Hence, the probability of rolling eleven is $2 \times 1/6 \times 1/6 = 1/18$.

209 NRC REPORT, *supra* note 12, at 78.

210 Genotypes AA, AB, and BB would also be more common in the general population than predicted by formula, although the effect might be difficult to detect if the subgroup was small relative to the general population.

four probes were used, the laboratory would, during Step 2, have computed four genotype frequencies. The product of these frequencies would be presented as the frequency of the entire DNA print. The use of the product rule (i.e., multiplication) to compute the frequency of multi-locus genotypes assumes that the frequencies of the genotypes are statistically independent, and may significantly underestimate the frequency of the multi-locus genotype if the individual genotypes are not independent.

When the genotypes at different loci are statistically independent in a given population, the population is said to be in linkage equilibrium.[211] Whether the major racial groups in the U.S. population are in linkage equilibrium is another major issue. Linkage equilibrium and Hardy-Weinberg equilibrium are closely related issues because endogamous mating patterns among heterogeneous groups could undermine both. "Once a population is known to be heterogeneous, one also cannot assume linkage equilibrium even for loci on different chromosomes; if an individual possesses an allele common among Puerto Ricans at one locus, it is more likely that he will do so at a second locus as well."[212] Hence, the possibility of endogamous mating among heterogeneous groups, which is also called population structure, is a key underlying issue in the debate over the validity of the forensic laboratories' statistical estimation methods.[213]

D. THE COMPETING VIEWS ON POPULATION STRUCTURE

The scientific community has split into three camps on the issue of population structure. One school of thought holds that concerns about population structure can be dismissed on theoretical grounds. Members of this "theoretical school" argue that deviations from Hardy-Weinberg equilibrium and linkage equilibrium in the general U.S. population are so implausible that empirical proof of the absence of such deviations is simply unnecessary.[214] A second group,

---

211 NRC REPORT, *supra* note 12, at 78-79.

212 Lander, *supra* note 3, at 504.

213 If there is such structuring, the multiplication of band frequencies could produce severe errors. By analogy, if a population survey showed that 10% of Europeans have blond hair, 10% have blue eyes, and 10% have fair skin, it would be a mistake to multiply these frequencies to conclude that the frequency of Europeans with all three traits is one in 1000. Because these traits tend to co-occur in Nordics, the actual frequency is much higher, particularly if one happens to be in Scandinavia. NRC REPORT, *supra* note 12, at 76.

214 Members of the theoretical camp concede exceptions for such distinctive subgroups as the Amish, Hutterites and Mennonites, which may be "inbred." *See* People v.

which has been labeled the statistical school,[215] "feels that the issue [of population structure] can be resolved by studying population samples of broad racial groups and applying statistical tests of Hardy-Weinberg equilibrium ("HWE") and linkage equilibrium ("LE") to detect substructure."[216]

In other words, they believe the issue of population structure must be addressed through empirical (rather than theoretical) analysis, but that statistical analysis of existing data is sufficient to resolve the issue.[217] The third group, which has been labeled the empiricist school, "feels that the only way to resolve the issue is to sample particular ethnic groups to ascertain the actual extent of variation—i.e., how high or low allele frequencies might range."[218] The empiricists[219] believe that no amount of statistical analysis on existing data bases can resolve the issue because such analyses "have insufficient statistical power to detect deviations, if present"[220] and because existing data bases may not include members of some endogamous mating groups. They favor collecting additional data on a number of distinctive subgroups.

E. THE DEBATE OVER METHODS OF TESTING FOR POPULATION STRUCTURE

Concerns about population structure were raised from the be-

Axell, 1 Cal. Rptr. 2d 411, 431 (Cal. Ct. App. 1991) (discussing testimony by prosecution expert Michael Conneally).

215 *See Lander Reply*, *supra* note 110, at 901. Lander sees three schools of thought on the issue of population structure, which he labels the "keep it simple school," the "statistical school" and the "empirical school." The present article follows Lander in distinguishing the statistical and empirical schools, but ignores the "keep it simple school" (the existence of which Lander appears to infer from the comments of a single scientist), and adds a "theoretical school" (which Lander fails to mention, but which has had an influential presence in litigation).

216 *Id.*

217 *See generally* Ranajit Chakraborty & Kenneth K. Kidd, *The Utility of DNA Typing in Forensic Work*, 254 SCIENCE 1735 (1991); Bernard Devlin et al., *Estimation of Allele Frequencies for VNTR Loci*, 48 AM. J. HUM. GENETICS 662 (1991); Weir, *supra* note 15 at 11656-58.

218 *Lander Reply*, *supra* note 110, at 901.

219 The most prominent exponents of the empiricist school are Richard Lewontin and Daniel Hartl, both Harvard geneticists. *See* Lewontin & Hartl, *supra* note 10, at 1745. Other articles supporting this viewpoint include Laurence Mueller, *Population Genetics of Hypervariable Human DNA*, *in* FORENSIC DNA TECHNOLOGY 51 (Mark Farley & James Harrington eds., 1991); Mueller, *supra* note 190, at 2; Lander, *supra* note 87, at 819; Dan Krane et al., *Genetic Differences at Four DNA Typing Loci in Finnish, Italian, and Mixed Caucasian Populations*, 89 PROC. NAT'L ACAD. SCI. 10583 (1992); Jennifer R. Slimowitz & Joel E. Cohen, *Violations of the Ceiling Principle: Exact Conditions and Statistical Independence*, 53 AM. J. HUM. GENETICS 314 (1993); Shields, *supra* note 10, at 1.

220 *Lander Reply*, *supra* note 110, at 901.

ginning of litigation on the admissibility of DNA evidence.[221] At first, however, the discussion was dominated by theoretical arguments.[222] The prosecution experts were generally members of the theoretical school, who found analysis of data unnecessary. The defense experts at first had difficulty gaining access to the data bases.[223]

> Much time in court was wasted by a failure of both the prosecution to establish that Hardy-Weinberg. . .held in their databases and the defense to show it did not. Early trials did not consider any empirical evidence on the issue. . . .[224]

As the data bases became available, the range of issues being debated in the courtroom broadened. Defense experts began presenting statistical analyses of the data bases which, they claimed, supported the existence of population structure.[225] Prosecution experts challenged these claims, and a debate erupted over the appropriate method for testing for substructure. According to Weir, "[t]he problem was that population geneticists had little experience

---

221 *See* People v. Wesley, 533 N.Y.S.2d 643 (N.Y. Sup. Ct. 1988), *aff'd* 589 N.Y.S.2d 197 (N.Y. App. Div. 1992) (Lifecodes' test); State v. Schwartz, 447 N.W.2d 422 (Minn. 1989) (Cellmark's test); *see also* Joyce, *supra* note 177, at 24 (describing the population genetics issues raised in State v. Anderson, 853 P.2d 135 (N.M. Ct. App.), *cert. granted*, 848 P.2d 531 (N.M. March 11 1993) regarding the FBI's test).

222 For example, in Caldwell v. State, 393 S.E.2d 436, 443 (Ga. 1990), prosecution experts characterized assumptions underlying Lifecodes' statistical procedures as "not unreasonable." But, the court noted that "none of the state's witnesses had studied Lifecodes' database to determine whether the relevant population is in Hardy-Weinberg equilibrium. Only a defense witness attempted to make that determination. [He] testified that he had analyzed Lifecodes' databases and concluded that the populations were not in Hardy-Weinberg equilibrium." *Id.* at 443.

223 In early cases, forensic laboratories provided tables of allele frequencies, but sometimes refused to provide the complete data bases in a form that would allow tests of Hardy-Weinberg equilibrium or linkage equilibrium. *See* State v. Schwartz, 427 N.W.2d 422, 427 (Minn. 1989) ("The defense request for more specific information regarding its methodology and population data base was denied by Cellmark."); *see also* Seymour Geisser, *Some Remarks on DNA Fingerprinting*, 3 CHANCE: NEW DIRECTIONS FOR STAT. AND COMPUTING 8 (1990) ("Cellmark and Lifecodes use proprietary excuses for shielding their data."). Cellmark first made its data bases available in 1989 under a court order issued in pretrial procedings in State v. Cauthron, 846 P.2d 502 (Wash. 1993), but only after extensive litigation during which Cellmark hired a private law firm to assist the district attorney in efforts to quash the discovery order.

The FBI also refused such requests at first, claiming that the data simply were not recorded in a manner allowing tests for independence across loci. Some defense experts found this claim suspicious, but it effectively thwarted their efforts to obtain the data. *See* Geisser, *supra*, at 9.

224 Weir, *supra* note 15, at 11655.

225 At first, these analyses tested only for Hardy-Weinberg equilibrium—that is, for independence of the alleles produced by a given probe. Tests for independence across loci were impossible for defense experts, because the necessary data were unavailable. Geisser, *supra* note 223, at 9.

in testing for Hardy-Weinberg for loci with many alleles."[226] Although tests for Hardy-Weinberg equilibrium and linkage equilibrium were well established,[227] they had, to that point, been applied only to genetic systems with few alleles. Extending these techniques to VNTRs, with many alleles, posed problems that had not previously been discussed in the scientific literature, and that were first addressed only recently.

One test for substructure compares the total frequency of homozygotes observed in a sample (data base) with the frequency expected if the sample is in Hardy-Weinberg equilibrium.[228] Because substructure entails endogamous mating within subgroups, it increases the likelihood that mating pairs will share the same allele (band) and thereby produce homozygous offspring (who have only one band, rather than two at a given loci). Hence, if the number of homozygotes observed in a data base exceeds the number expected to occur by random mating (by an amount unlikely to occur by chance), it is evidence of substructure. Tests for "excess homozygosity" were first performed on forensic data bases by experts retained by defendants in criminal cases. They reported "spectacular deviations from Hardy-Weinberg equilibrium"[229] and argued that these findings raised serious concerns about the validity of the statistical procedures of the forensic laboratories.

Prosecution experts responded that the "excess" of homozygotes was more apparent than real. Technical problems in the assay of samples in the data bases may have caused some individuals who are in fact heterozygous to appear to have a single band. Hence, they argued, the true frequency of homozygotes is undoubtedly lower than indicated by the data. Some experts have argued that the

---

226 Weir, *supra* note 15, at 11655; *see also* Lander, *supra* note 87, at 821 ("The task is made especially difficult by the large number of alleles and by the measurement error inherent in hypervariable genetic systems.").

227 *See generally* BRUCE S. WEIR, GENETIC DATA ANALYSIS (1990).

228 If a population is in Hardy-Weinberg equilibrium, the expected frequency of homozygotes is the sum of the squared frequency of each allele. For example, if a population has four equally common alleles—*A*, *B*, *C* and *D*—the frequency of homozygotes (i.e., genotypes *AA*, *BB*, *CC* and *DD*), if mating is random, is expected to be $.25^2+.25^2+.25^2+.25^2 = .25$, because a person with allele *A*, for example, is no more likely to mate with another *A* than to mate with a *B*, *C* or *D*. Population structure means that people are more likely to mate with others in their subgroup, hence people are more likely to mate with someone sharing their same allele and the overall frequency of homozygotes is higher.

229 *See* Lander, *supra* note 3, at 504; *see also* Mueller, *supra* note 219, at 57 ("My own analysis [of forensic data bases] have shown consistent deviations from the Hardy-Weinberg law."); Geisser, *supra* note 223, at 9 ("an excess of homozygotes over Hardy-Weinberg expectations has often been found by defense experts").

"excess" homozygosity disappears entirely when adequate corrections are made for technical problems in the assays,[230] although this conclusion is controversial.[231]

Debate about total homozygosity has faded over time as population geneticists on both sides of the dispute have come to realize that these analyses cannot adequately address the issue of substructure. Measurement error and uncertainty about the quality of the data undermines confidence in any analysis suggesting substructure is present;[232] the limited sensitivity of these procedures undermines confidence in any analysis suggesting it is absent.[233]

Another way to test for substructure is to compare the distribution of allele frequencies in various subgroups. However, there has been controversy about which subgroups allow relevant comparisons. For example, FBI scientists argued against the possibility of substructure based on data that showed that, within major groups (Caucasians, Blacks, Hispanics), similar allele frequencies were found in samples drawn in Texas and Florida.[234] Lander responded: "One might analogously conclude that blond hair, blue eyes, and fair skin are not correlated because such traits show similar frequencies in Florida and Texas; examining average frequencies in mixed populations sheds no light on substructure."[235] According to Lander and others of the empiricist school, what is needed is direct comparison of distinct ethnic subgroups; differences among such groups are difficult to detect in mixed populations.[236] The

---

230 Bernard Devlin et al., *No Excess of Homozygosity at Loci Used for DNA Fingerprinting*, 149 SCIENCE 1416 (1990).

231 *See Lander Reply*, *supra* note 110, at 901 (arguing that the conclusions of Devlin et al. have "been disproved"); *see also*, Phillip Green & Eric Lander, *Technical Comment*, 253 SCIENCE 1038 (1991) and Joel Cohen et al., *Technical Comment*, 253 SCIENCE 1037 (1991) (both challenging the conclusions of Devlin et al.).

232 Devlin et al., *supra* note 230, at 1418.

233 "Failure to reject the hypothesis that genotypes are well described by the Hardy-Weinberg law does not mean that large errors cannot be made. . . . [t]he data set and statistical tests have limited power to uncover deviations if they do exist." Mueller, *supra* note 219, at 57, 62; *see also* Lander, *supra* note 87, at 821 (calling the failure to find excess homozygosity "virtually meaningless because the tests have such low statistical power to detect substructure even if it is present"); Lewontin & Hartl, *supra* note 10, at 1747 ("Statistical tests for HWE are so lacking power that they are probably the worst way to look for genetic differentiation between subgroups in a population.").

234 Budowle et al., *supra* note 86, at 851.

235 Lander, *supra* note 87, at 821.

236 Another methodological issue surrounding the comparison of subgroups concerns the utility of "genetic distance" computations. Genetic distance is a statistic "designed to express the genetic difference between two populations as a single number." WEIR, *supra* note 227, at 163. Although some experts believe the genetic distance between two subgroups provides "[a] useful quantification of the structure in a population," *Id.* at

NRC Report adopted this position.[237]

Recently, courtroom debate has shifted as new analyses of the data bases, using more sophisticated methods, have appeared. These new analyses purport to find reassuring evidence of the statistical independence of VNTR alleles, and thus to show that substructure is not a significant problem. Now it is primarily experts for the prosecution who are relying on statistical analyses of data bases, while experts for the defense are challenging the accuracy of the data and the appropriateness of the statistical methods. The theoretical school, on which prosecutors initially relied, has largely been abandoned and the main debate is now between proponents of the statistical school, who testify on behalf of the prosecution, and proponents of the empiricist school, who testify on behalf of the defense.

One of the "new" approaches has been simply to check within data bases for matches between different individuals over several loci. In 1991, the first published study of this type revealed extremely low frequencies of such matches in the data bases used by the FBI.[238] Although this study has been faulted for employing a more stringent criteria for a match than the FBI uses in casework,[239] subsequent research, employing more lenient match criteria, continues to show very low frequencies of multi-locus matches in forensic data bases[240] and, more importantly, to show that the frequency of such matches does not signficantly exceed the number that would be expected if the alleles are statistically independent, an important in-

171, the application of this statistical method to the analysis of VNTR data has been controversial.

237 NRC REPORT, *supra* note 12, at 80-82.

238 Neil Risch & Bernard Devlin, *On the Probability of Matching DNA Fingerprints*, 255 SCIENCE 717 (1991) ("an innocent defendant has little to fear from DNA fingerprinting unless he has an evil twin").

239 Patrick J. Sullivan, Letter, *DNA Fingerprint Matches*, 256 SCIENCE 1743 (1992).

240 *See* George Herrin, *Probability of Matching RFLP Patterns from Unrelated Individuals*, 52 AM. J. HUM. GENETICS 491 (1993). Herrin used four different quantitative match criteria when searching for multi-locus matches in data bases gathered by eight crime laboratories in the Southeastern United States. Applying a 5% match criterion (approximating the FBI's matching standard) in pairwise comparisons in which each of several hundred individuals was compared to all others, he found relatively few matches: the highest number of three-locus matches (for any combination of three probes) was three out of 377,146 comparisons, a match rate of approximately one in 126,000. Arguably, a 5% match criteria may be too small for detecting potential matches in data bases (where variability of band measurement is greater than in case work). Using a 10% match criterion, the highest number of three-locus matches (for any combination of three probes) increases to 83 out of 377,146 comparisons, a match rate of approximately one in 4500, although the rate of matches is much lower for some three-probe combinations.

dication of the absence of substructure, at least among the individuals in existing data bases.

Another "new" approach to analyzing existing data bases uses sophisticated techniques for Hardy-Weinberg and linkage equilibrium pioneered by Weir. Although Weir first described these techniques in 1978[241], methods for applying them to VNTR data were not worked out until 1991, and analyses of VNTR data using these technique were not published until 1992.[242] These tests purport to show that "substructuring, at the geographic level, exists but is unlikely to have a meaningful impact on multilocus genotypic frequencies."[243]

Critics from the empiricist school respond that analyses of existing data bases cannot rule out the possibility that there is substructure within broad American populations. There may be significant genetic variation among ethnic subgroups that goes undetected because members of discrepant subgroups are not included in the data bases, or because they appear in numbers too small for their differences to be noticed.[244] Most data bases consist of "blood bank or hospital data from a narrow region"[245] and therefore may fail to capture the genetic diversity of the total population.[246]

The leading proponents of this viewpoint, Lewontin and Hartl, argue that the large racial/ethnic groups for which laboratories maintain data bases are in fact "conglomerates" of ethnic subgroups

---

241 Bruce Weir & C. C. Cockerham, *Testing Hypotheses About Linkage Disequilibrium with Multiple Alleles*, 42 HEREDITY 105 (1978).

242 Bruce Weir, *Independence of VNTR Alleles Defined as Fixed Bins*, 130 GENETICS 873, 887 (1992) [hereinafter *Fixed Bins*]; Bruce Weir, *Independence of VNTR Alleles Defined as Floating Bins*, 51 AM. J. HUM. GENETICS 992, 997 (1992) [hereinafter *Floating Bins*].

243 *See Fixed Bins*, *supra* note 242, at 887; *see also Floating Bins*, *supra* note 242, at 997 ("whatever levels of dependence do exist are unlikely to have a meaningful impact on forensic calculations").

244 *See, e.g.*, R.A. Nichols & D.J. Balding, *Effects of Population Structure on DNA Fingerprint Analysis in Forensic Science*, 66 HEREDITY 297, 299-300 (1991) ("[H]uman populations are known to be composed of large relatively outbred populations, typically in cities, and smaller inbred populations. In a random sample from both, the effects of inbreeding in the small populations will be swamped. We must examine studies of inbred populations for estimates [of how much they vary].").

245 Weir, *supra* note 15, at 11656. Cellmark and the FBI draw their data bases from blood banks. Lifecodes uses samples from paternity test casework, which "tend to be from a wide geographic area," *Id.* Nevertheless, "[t]here have been arguments about the ethnic composition of people becoming involved in paternity disputes and donating blood, and these are unlikely to be resolved." *Id.*

246 Mueller, *supra* note 190, at 5 ("The people in the data base have not been sampled at random and thus there is no expectation that the mixture of subgroups within these samples is representative of any real population.").

which stem from genetically differentiated ancestral populations[247] and have not yet fully blended in the American "melting pot."[248] According to them and other empiricists, the way to resolve the issue of substructure is "to sample ethnically distinct subpopulations and to observe the actual degree of genetic differentiation."[249] Some data on the range of variation among distinctive population subgroups already exist, and more is emerging all the time.[250] Hence, a resolution to the issue of substructure may be in sight.[251]

In litigation, some experts have repeatedly asserted that allele frequencies *are* similar across distinctive ethnic groups.[252] Efforts to evaluate such claims have been impeded, however, by the failure of the prosecution experts to grant access to the data. For example, Professor Kenneth Kidd, a frequent prosecution witness, who began testifying about the lack of variation among subgroups in 1988, refused to allow access to some of the data on which he based his testimony until 1991.[253] At that point, critics challenged some of Kidd's assertions about the data. Lewontin and Hartl reported that there are striking differences in allele frequencies between some of the groups Kidd examined.[254] In one American Indian group, the Karitiana of Brazil, the entire sample of over fifty individuals had the same allele at one VNTR locus and had one of two alleles at another locus. Lewontin and Hartl note that a comparison of the Karitiana

247 Among Caucasians, for example, evidence from genetic markers other than VNTRs show "more genetic variation among Irish, Spanish, Italians, Slavs, Swedes, and other subpopulations, than there is on the average, between Europeans, Asians, Africans, Amerinds, and Oceanians." Lewontin & Hartl, *supra* note 10, at 1747.

248 *See* Kenneth Lange, *Match Probabilities in Racially Admixed Populations*, 52 AM. J. HUM. GENETICS 305, 306 (1993) ("What is lacking in these discussions [of match probabilities] is an agreed-on method of computing match probabilities in admixed populations typical of the United States.")

249 Lander, *supra* note 87, at 821; Philip Green, Letter, *Population Genetic Issues in DNA Fingerprinting*, 50 AM. J. HUM. GENETICS 440, 441 (1992) ("For VNTR loci, there is no obvious alternative to gathering community/ethnic-specific subpopulation data bases.").

250 *See* Krane et al., *supra* note 219, at 10583.

251 *But see* Weir, *supra* note 15, at 11656 (suggesting that "collecting such data may not be feasible").

252 "In *Jakobetz*, as here, Doctor Kidd testified that from looking at data from many subgroups, i.e., Irish, Swedes, Amish, all have 'very small differences' in allele frequencies." People v. Axell, 1 Cal. Rptr. 2d 411, 431 (Cal. Ct. App. 1992) (citing United States v. Jakobetz, 747 F. Supp. 250, 260 (D. Vt. 1990)).

253 Quite understandably, Kidd did not want to make public data he had collected at great effort until he had had the first opportunity to analyze and publish it. He made the data public after publishing an article on it in 1991. J.R. Kidd et al., *Studies of Three Amerindian Populations Using Nuclear DNA Polymorphisms*, 63 HUM. BIOL. 775 (1991). Before then, the data was produced through discovery in United States v. Yee, 129 F.R.D. 629 (N.D. Ohio 1990) but was subject to a protective order that prevented the experts for the defense from discussing it outside the courtroom.

254 Lewontin & Hartl, *supra* note 10 , at 1749.

with another Brazilian group, the Surai, show that "certain three-locus VNTR genotypes differ in frequency by a factor of more than 500 in these populations, even though they are separated by only 420 kilometers."[255] In response, Chakraborty and Kidd state that despite inbreeding among the Karitiana "there were no two individuals with identical VNTR profiles."[256] However, this assertion has proven false. An analysis of Kidd's Karitiana data by Professor Laurence Mueller found that over twenty percent of the individuals in the sample matched another individual over four or more probes. Two pairs of individuals matched over seven probes, although one of the seven-probe matches appears to have been a "false positive" caused by a laboratory sample handling error.[257] Mueller also discovered a previously undetected six-probe match between one of the Karitiana and a member of a different group, the Maya.[258]

Defense experts have also had difficulty obtaining data on population subgroups from the FBI. FBI scientists began testifying about data purportedly showing an absence of variation across subgroups in early 1990, and discussed the data (while presenting only selected examples) in a published article in 1991,[259] but have yet to produce all of the underlying data in discovery.[260] The problem of access may have been resolved by the recent publication by the FBI of a four-volume set of data compiled by the FBI.[261] However, this 1993 publication appeared over three years after the FBI first refused to provide access to data that have played a role in litigation and public debate[262] in the interim. Given the stakes, three years is

255 *Id.* (citing *Yee*, 129 F.R.D. at 629).

256 Chakraborty & Kidd, *supra* note 217, at 1738.

257 Mueller, *supra* note 190, at 4-5. A false positive is an event which is, itself, of considerable interest and importance.

258 *Id.* at 5. In describing the differences between Kidd's assessment of his data and the assessment of it by the defense experts, the author does not mean to imply that Dr. Kidd, a distinguished scientist, intentionally misrepresented his findings. What the episode illustrates is that different scientists can draw differing conclusions from the same data, and thus that it is very important that data be open to general scrutiny. It is dangerous to rely on the unexamined conclusions of a single scientist (or a single group of scientists), no matter how distinguished.

259 Budowle et al., *supra* note 86, at 851.

260 In one case with which the author is familiar, when defense lawyers requested the data underlying claims made by Budowle et al. (the claims were supported by citations to "authors' unpublished data," *see* Budowle et al., *supra* note 86, at 851), the FBI staff responded that the article is incorrect—the data cited did not in fact belong to the authors and therefore could not be produced.

261 VNTR POPULATION DATA: A WORLDWIDE STUDY, 1993. (available from the National Technical Information Service, Washington, D.C.)

262 *See, e.g.*, John Brookfield, *Law and Probabilities*, 355 NATURE 207 (1992) (concluding that "surveys of diverse populations" have not shown evidence of substructure).

a long time to wait.

F. THE NATIONAL RESEARCH COUNCIL'S CEILING APPROACH

The NRC report acknowledged the existence of the scientific dispute over population structure and proposed a compromise.[263] The danger of population structure is sufficiently serious, the NRC concluded, that current approaches should not be used. Additional empirical studies of ethnic subgroups are needed (immediately) to determine the extent of population structure. However, the NRC asserts that it is not necessary, as some critics had suggested, to abandon the use of the product rule altogether. Laboratories may continue multiplying allele frequencies so long as they employ a conservative approach for estimating those frequencies—an approach the NRC calls "the ceiling principle."[264] This principle is designed to take into account the possibility that undetected population structure may make allele frequencies far higher in some groups than others.

> [I]t is not enough to sample broad populations defined as "races" in the U.S. census (e.g., Hispanics), because of the possibility of substructure. On the other hand, it is not feasible or reasonable to sample every conceivable subpopulation in the world to obtain a guaranteed upper bound. The committee strongly recommends the following approach: Random samples of 100 persons should be drawn from each of 15-20 populations, each representing a group relatively homogeneous genetically; the largest frequency in any of these populations or 5%, whichever is larger, should be taken as the ceiling frequency.[265]

Pending the completion of the population studies, the NRC recommended the following approach to statistical calculation. First, the laboratory should check to determine whether the DNA print observed in casework matches any DNA prints in existing data bases. The frequency of such matches (and the size of the data bases) should be reported to the trier of fact. Second, the laboratory should estimate the frequency of the DNA print by applying the product rule to "modified ceiling frequencies," consisting of the 95% upper confidence limit of the highest frequency observed in an existing data base or 10%, whichever is higher.[266]

The NRC's modified ceiling approach was intended to provide a reasonable, broadly acceptable basis for making statistical esti-

---

263 NRC REPORT, *supra* note 12, at 79-80.

264 *Id.* at 82-85.

265 *Id.* at 83. "Ideally, the populations should span the range of ethnic groups that are represented in the United States—e.g., English, Germans, Italians, Russians, Navahos, Puerto Ricans, Chinese, Japanese, Vietnamese, and West Africans." *Id.* at 84.

266 *Id.* at 91-93, 95.

mates in the face of the current scientific uncertainty over population structure.

> In order to ensure the admissibility of this important technology, the NRC committee sought to define common ground, namely a standard of practice so conservative as to ensure that there would be no serious scientific argument that the evidence could be said to *overstate* the case against a defendant[267]

But the modified ceiling approach has not yet proven to be an acceptable compromise. Some scientists have attacked the ceiling principle as unnecessarily conservative,[268] some say it is unclear and subject to conflicting interpretations,[269] some have defended it as a workable compromise,[270] and some, confounding the expectations of the NRC panel, have argued seriously that it may not be conservative enough.[271] A major theme of critics from all perspectives is that the "ceiling principle" is not a principle of science. It is an arbitrary policy statement, and can be accepted or rejected only as such.[272]

Another problem is that a dispute has arisen over how to apply the NRC's modified ceiling approach.[273] Several points are in contention. One issue is which data bases (and how many data bases) should be consulted when estimating allele frequency. Some ex-

---

267 Eric S. Lander, letter, *DNA Fingerprinting: The NRC Report*, 260 SCIENCE 1221 (1993). Lander was a member of the NRC panel.

268 *See* Aldhous, *supra* note 15 at 755; Bernard Devlin et al., *Statistical Evaluation of DNA Fingerprinting: A Critique of the NRC's Report*, 259 SCIENCE 748 (1993).

269 Weir, *supra* note 15, at 11654; Bruce Weir, letter, *DNA Fingerprinting Report*, 260 SCIENCE 473 (1993).

270 Lander, *supra* note 289; Krane et al., *supra* note 219, at 10587 ("the interim ceiling principle is a significant improvement over the conventional product rule for estimating the probability of matching DNA profiles").

271 Slimowitz & Cohen, *supra* note 219, 315 ("Before the ceiling principle is implemented, more research should be done to determine . . . whether the deviation would cause the ceiling principle to be non-conservative in practice.")

272 In an affidavit filed recently in a case in Washington state, Professor Elizabeth Thompson, Chair of the Department of Statistics at the University of Washington, derided the ceiling principle as "data-driven, interest-ridden, voodoo, pseudo-statistical, ad hoc methodology to which no statistician (or scientist) should be a party." State v. DeFroe & Hollis, 92-1-03699-8 & 92-104-603-9 (King Co., Wash. Super. Ct. June 3, 1993)(Ruling and motion to exclude DNA evidence). Responding to the question of whether the ceiling principle is generally accepted, she declared it, and other ad hoc adjustment methods "are not principles and are not science. Hence it is useless to ask concerning acceptance of these 'principles' by the "scientific community."

Similarly, Richard Lewontin, in an affidavit filed in a case in California, declared:

> In my view, the "modified ceiling principle" has no rational basis and has been chosen by entirely arbitrary means . . . . It is clear to me that the "modified ceiling principle" was invented in the hope of maintaining the use of DNA pattern matching for forensic purposes, despite the lack of the necessary statistical information to allow a valid estimate of matching probabilities."

273 Mueller, *supra* note 190, at 2; Sherman, *supra* note 3, at 1.

perts read the NRC recommendations to mean that *all* existing data bases (for a given DNA typing system) should be consulted and that the allele frequency should be the higher of .10 or the 95% upper limit of the confidence interval above the highest frequency observed in any data base. Others have taken the position that only data bases for "major races" (i.e., the FBI's Caucasian, Hispanic and Black data bases) should be consulted. Another issue is whether, in return for adopting this "conservative" approach, labs may abandon other "conservative" features of their estimation procedure.[274] What is at stake, obviously, is just how conservative the modified ceiling numbers will be, and how much flexibility forensic laboratories will have in implementing the NRC's recommendations. These interpretive issues may have a dramatic effect on the results of the calculations.[275] Hence, for courts considering the admissibility of statistical estimates, the issue is not whether the NRC's modified ceiling estimates are admissible, but which versions, if any, of these numbers are admissible.

The National Research Council is currently planning a new study that will make a second attempt to resolve the statistical debate, drawing on new data that has emerged since the publication of the NRC Report.[276] The NRC's new effort was prompted by a request from FBI Director William Sessions, who declared that debate over the ceiling principle has "created a climate of confusion for the courts."[277] Sessions noted that since the release of the NRC Report, eleven of thirty appellate decisions on the admissibility of DNA evidence have ruled it inadmissible, and that "courts in Canada,

---

274 The FBI at one point took the position that the modified ceiling approach is acceptable only if used with floating bins (in the FBI system) of ±2.5% around each band. It appears, however, that such a bin is smaller than the interval within which the FBI will declare bands to match; hence it may not be broadly acceptable to other scientists. *See supra* notes 116-20.

275 *See* Sherman, *supra* note 3, at 1.

In one case, an FBI Agent-Examiner purported to follow the modified ceiling principle while computing the frequency of defendant's DNA profile six different ways. The results ranged from one in 1.26 million (obtained by using floating bins smaller than the FBI's matching criterion and four probes) to one in 877 (using fixed bins and three probes—one probe was dropped due to concerns about Hardy-Weinberg disequilibrium). FBI Computations produced in discovery in State v. Anderson, 853 P.2d 135 (N.M. Ct. App.), *cert. granted*, 848 P.2d 531 (N.M. Mar. 11, 1993). But even this did not represent the full possible range. A defense expert, using three probes and fixed bins, but consulting a larger number of data bases, computed the frequency to be one in 84. Affidavit of Laurence Mueller in *Anderson* (No. CR-46255).

276 Telephone Interview with Eric Fischer, Project Director, National Research Council, Sept. 15, 1993.

277 Letter from William S. Sessions, ex-Director of the FBI to Dr. Frank Press, President of the National Academy of Sciences (Apr. 16, 1993).

Australia, and the United Kingdom began hearing challenges to DNA evidence—citing the NRC report—immediately following its release.[278] Calling this situation "a crisis," Sessions urged the NRC to "act quickly to resolve the controversy. . . ."[279] The FBI, which will help fund the new study, has been working closely with the NRC to define the scope of the NRC's efforts.[280] Regardless of its conclusions, the new NRC report, when it is delivered, is likely to be influential.

G. DETERMINING THE RELEVANT REFERENCE POPULATION

Another area of scientific dispute, closely linked to the debate over population structure, concerns the relevant reference population to be used to estimate the frequency of a DNA profile. Forensic laboratories typically assume that the relevant reference population is the broad racial or ethnic group of which the defendant is a member. Thus, if the defendant is a Caucasian, the laboratory reports the frequency of the defendant's DNA profile among Caucasians (basing its estimate on the frequency of defendant's alleles in the laboratory's Caucasian data base). Much of the debate over population structure has been framed in terms of the difficulty of estimating the frequency of the defendant's DNA profile, given that the defendant may be a member of a distinctive subgroup.[281] Discussion of the defendant's ethnic background has been a common theme in judicial opinions.

Recently, some commentators have argued that focusing on the defendant's ethnic background miscasts the issue. In a typical case, the defendant's DNA profile is known and has been found to match the DNA profile of an evidentiary sample. By denying that he is the source of the evidentiary sample, the defendant implicitly raises the possibility that the true perpetrator, whose identity is unknown, also has the "matching" profile. Hence, "it is not the ethnicity of the *defendant* that is the directly relevant question but rather the ethnic composition of the pool of possible alternative suspects."[282]

---

278 *Id.* at 1.

279 *Id.* at 3.

280 At one point, the FBI's Assistant Director in Charge of the Laboratory Division, John Hicks, threatened to withhold funding unless the NRC limited the scope of its inquiry to procedures for estimating the frequency of DNA profiles in a reference population. Letter from John Hicks to Richard Rau, National Institute of Justice (May 27, 1993). Hicks was apparently reacting negatively to a May 11, 1993 NRC Prospectus that suggested the new committee would also "[a]ssess and describe the degree of certainty of DNA evidence in ways useful to the courts. . . ."

281 *See, e.g.*, People v. Mohit, 579 N.Y.S.2d 990, 997-99 (Co. Ct. 1992).

282 R.C. Lewontin, Letter, *Which Population?*, 52 AM. J. HUM. GENETICS 205 (1993).

Based on this conclusion, a few commentators have argued that concerns about population structure are largely irrelevant to forensic DNA testing.[283] In most criminal cases, they argue, the pool of alternative suspects includes individuals from a variety of ethnic subgroups, not just individuals in the defendant's subgroup. It is the aggregate frequency of the matching profile, across all such groups, rather than its frequency within a particular group, that determines the likelihood a person drawn at random from the pool of possible suspects would match. Uncertainties about the frequency of the matching profile in defendant's ethnic subgroup, and the failure of forensic laboratories to gather "community/ethnic specific data bases" are not serious problems because it is aggregate frequencies, rather than frequencies in specific subgroups, that matter.

This argument rests on the assumption that there is no correlation between the ethnic background of the defendant and that of the true perpetrator in cases where the defendant is falsely accused—an assumption that is clearly wrong in some cases, and may be questionable in most. In some cases, for example, police will have information about the appearance of the perpetrator which limits the pool of possible suspects to individuals of the defendant's ethnicity. Even if nothing is known about the appearance of the perpetrator, however, the tendency of people to cluster and live in close proximity with others of similar ethnic background will, in many cases, assure that the pool of alternative suspects is heavily weighted toward individuals of the defendant's ethnicity. If a crime occurs in a neighborhood with a heavy concentration of people of Puerto Rican ancestry, for example, the defendant is likely to be of Puerto Rican ancestry, as is the majority of the pool of alternative suspects.[284] Hence, the need to know whether the frequency of a particular DNA profile is higher among defendant's subgroup (the local community of Puerto Rican immigrants) than among a larger aggregate group for which the forensic laboratories has a data base (e.g., Hispanics), will be unavoidable in many cases.[285] Whether the defendant's

283 Bruce Weir & Ian Evett, Letter, *Whose DNA?*, 50 AM. J. HUM. GENETICS 869 (1992); Weir, *supra*, note 15, at 11656.

284 Consider, for example, United States v. Two Bulls, 918 F.2d 56 (8th Cir. 1990), vacated for rehearing en banc but appeal dismissed due to death of defendant, 925 F.2d 1127 (8th Cir. 1991), where defendant was an Oglala Sioux and the crime, a rape, occurred on an Indian reservation for Oglala Sioux.

285 There are some cases, however, in which the pool of alternative suspects is unlikely to include individuals of the defendant's ethnicity. For example, in *Mohit*, the defendant was a Shiite Muslim of Iranian origin who was accused of rape in Westchester County, New York. Although the court expressed concern about the frequency of defendant's DNA profile within his ethnic subgroup, noting that defendant's "ancestors

ethnicity is relevant or not in a given case will not always be apparent, and will often be debatable.[286]

H. WEIGHT OR ADMISSIBILITY?

A key issue for courts is whether the dispute over statistical methods goes to the admissibility or merely the weight of the resulting statistical estimates. Prosecutors often contend that the accuracy of statistical estimates is a matter of weight to be considered by the jury; defense lawyers respond that the statistical method is neither generally accepted nor valid and therefore that the resulting estimates are inadmissible.

A common argument by prosecutors is that the statistical method is not properly reviewable under *Frye* because it is not novel.[287] By their account, fears about population structure do not

---

over at least the past five generations were of Persian descent, and all from the same town or a town close by. . . . inbreeding was very common in his family," such concerns are appropriate only if other members of defendant's group were members of the pool of possible alternative suspects, which seems unlikely given that the crime occurred in New York rather than Iran. Whether cases like *Mohit*, where the defendant's ethnicity appears irrelevant, outnumber cases like *Two Bulls*, where the defendant's ethnicity appears critical, is an empirical question on which we currently have no data. In any case, the existence of cases like *Mohit*, even if they are a majority, will not eliminate the need to address the issue of substructure for the remaining cases.

286 *See e.g.*, Weir & Evett, *supra* note 283, and the response by Lewontin, *supra* note 283, at 205. This revealing exchange of letters concerns the relevance of the ethnicity of the defendant in State v. Passino, No.185-1-90 (Franklin Co., Vt. Dist. Ct. filed May 13, 1991), where the trial court excluded the FBI's DNA test based on concerns that the defendant "belongs to an ethnic group whose genotype frequencies may occur more frequently than the FBI's estimate." (p. 26 of memorandum opinion). Defendant Passino, who was part Caucasian and part Abanaki Indian, had challenged the FBI's use of its Caucasian data base. (p. 13, n. 5). Weir and Evett consider the result a mistake because the true perpetrator was not necessarily an individual who matched the defendant's unusual background. Lewontin responds that the county where the crime occurred "has the highest proportions of Indians, largely Abnaki, of any county in the state" and notes that "[t]he victim, who was herself half-Abnaki, was assaulted and killed in a trailer camp in which a large fraction of the other residents were also of Abnaki ancestry." Hence, "the defense could make the entirely reasonable claim that it is indeed the Abnaki who comprise 'the population' of potential suspects, of whom defendant is only one." Lewontin concedes that the prosecution could make an alternative argument that since the trailer park was near a state highway "the entire population of western Vermont and eastern New York is the appropriate reference groups," but suggests that this merely adds to uncertainty about the value of DNA evidence: "Rather than being faced with objective science, we are faced with subjective arguments about the patterns of people's lives."

287 Statistical evidence on the frequency of genetic characteristics has previously been held admissible in connection with serology tests. *See, e.g.*, Commonwealth v. Gomes, 526 N.E.2d 1270 (Mass. 1988); People v. Morris, 199 Cal. App. 3d 377 (1989). Prosecutors point out that the "method" used for computing serology statistics is similar to the statistical method used by forensic DNA laboratories—that is, the frequency of particular alleles is estimated with reference to a data base, and the joint frequency of alleles in

concern the reliability of "the method" per se, but merely the accuracy of the results obtained by applying the method to particular data bases; whether the correct data base was used is, they argue, a matter for the jury.[288] This characterization of the scientific criticism, however, has been repudiated by one of the key critics[289] and is inconsistent with the NRC Report as well.[290]

Of course, the underlying question in this battle of characterizations is whether courts should wait for a resolution of the dispute among population geneticists before admitting statistical estimates based on current methods, or should admit such estimates now and allow the dispute to be aired and argued in front of juries. This question raises important policy issues concerning the ability of lawyers and experts to present scientific evidence effectively and the ability of juries to understand and draw reasonable conclusions from it. Semantic disputes over whether or not scientific criticisms concern "the method," or whether it is or merely the application of the method that is novel, do little to illuminate these issues. Empirical research on these issues has only just begun.[291]

---

different marker systems is multiplied to determine the joint frequency of a set of characteristics. Defense lawyers respond that the application of this technique to alleles generated by RFLP analysis of VNTRs is indeed novel, and is the very issue that is subject to scientific dispute.

288 For example, in a brief filed with the California Supreme Court, requesting review of People v. Barney, 10 Cal. Rptr. 2d 731 (1992), deputy district attorney Rockne Harmon of Alameda County, California argued, *inter alia*, that *Barney* rested on the "mistaken belief that the SCIENCE article [by Profs. Lewontin and Hartl] questioned the procedures themselves, rather than the actual resulting estimates." Letter brief filed Sept. 15, 1992. Harmon was the prosecutor in both *Barney* and *Howard* and entered both appeals as amicus curiae; his letter brief was filed under local rules allowing interested parties to comment upon petitions for review filed with the California Supreme Court.

289 In an episode reminiscent of the scene from the movie ANNIE HALL in which Woody Allen produces Marshall McLuhan while arguing with a pedant about McLuhan's views, counsel for Barney and Howard asked Professor Lewontin himself to respond to Harmon's characterization of the Lewontin & Hartl article. He did:

> It is [Harmon who has] a 'mistaken belief' about the article in question. The content and stated intent of the article were to challenge the *method* used to calculate probabilities of matches, and we concluded that the procedures themselves were unreliable and invalid.

Letter brief of Richard C. Lewontin, dated Oct. 1, 1992 and filed with the California Supreme Court Oct. 5, 1992. The California Supreme Court denied the petition to review *Barney* on November 25, 1992, letting stand the ruling of the Court of Appeal.

290 The NRC Report also suggests that the statistical dispute concerns "the method." Noting that different methods of determining the frequency of DNA prints can produce strikingly different estimates, the NRC Report suggests that "[t]he discrepancy not only is a question of the weight to accord the evidence (which is traditionally left to the jury), but bears on the scientific validity of the alternative methods used for rendering estimates of the weight (which is a threshold question for admissibility)." NRC REPORT, *supra* note 12, at 75.

291 *See* Jonathan J. Koehler, *DNA Matches and Statistics: Important Questions, Surprising*

Appellate courts that have considered the issue have generally held that the statistical methods are properly reviewable under *Frye*, offering as a justification the crucial importance of the statistical estimates and the arcane nature of the scientific dispute:

> To end the Kelly-Frye inquiry at the matching step, and leave it to jurors to assess the current scientific debate on statistical calculation as a matter of weight rather than admissibility, would stand Kelly-Frye on its head. We would be asking jurors to do what judges carefully avoid—decide the substantive merits of competing scientific opinion as to the reliability of a novel method of scientific proof. We cannot reasonably ask the average juror to decide such arcane questions as whether genetic substructuring and linkage disequilibrium preclude use of the Hardy-Weinberg equation and the product rule, when we ourselves have struggled to grasp these concepts. The result would be predictable. The jury would simply skip to the bottom line—the only aspect of the process that is readily understood—and look at the ultimate expression of match probability, without competently assessing the reliability of the process by which the laboratory got to the bottom line. This is an instance in which the method of scientific proof is so impenetrable that it would " '. . . assume a posture of mystic infallibility in the eyes of a jury . . . .'"[292]

I. ARE STATISTICS NECESSARY?

A number of courts have concluded that the procedures for producing DNA prints and determining whether they match are generally accepted but that procedures for statistical computation are not. These courts face an interesting decision: whether to admit evidence of a DNA test match in the absence of statistical data. A few courts have allowed a forensic expert to testify that two DNA patterns "match" but have barred the expert from presenting statistics to quantify the rarity of the matching DNA pattern or the likelihood of a coincidental match.[293] Other courts have viewed DNA evidence as meaningless and potentially prejudicial in the absence of such statistics.[294] They therefore exclude the evidence altogether

---

Answers, 76 JUDICATURE 222, 226 (1993)(reporting groundbreaking research examining mock jurors' reactions to DNA evidence as a function of how the statistical estimates are presented). Empirical research of this type is potentially very helpful in evaluating juror competency, but no studies as yet have examined reactions to the complex sort of testimony likely to arise in a clash of statisticians or population geneticists over statistical methods. For general background on jurors' use of probabilistic evidence in criminal trials, *see* Thompson, *supra* note 178, at 9; David H. Kaye & Jonathan J. Koehler, *Can Jurors Understand Probabilistic Evidence*, 154 J. ROYAL STATIST. SOC'Y 75 (1991).

[292] People v. Barney, 10 Cal. Rptr. 2d 731, 742 (Cal. Ct. App. 1992) (citing People v. Kelly, 549 P.2d 1240, 1245 (Cal. 1976)).

[293] Rivera v. State, 340 P.2d 933 (Wyo. 1992); State v. Schwartz, 447 N.W.2d 422 (Minn. 1989)

[294] Commonwealth v. Curnin, 526 N.E.2d 440, 443 (Mass. 1991) ("It is apparent from

on finding the results of the statistical procedures inadmissible. This position was strongly supported by the NRC Report:

> To say that two patterns match, without providing any scientifically valid estimate (or, at least, an upper bound) of the frequency with which such matches might occur by chance, is meaningless.[295]

The major danger of admitting DNA evidence without statistics is that jurors may mistakenly assume that because each person's DNA is unique, any DNA test showing a "match" provides a unique identification. Unfortunately, this error may be encouraged by the tendency of some forensic experts to use loose language which implies, or even states outright, that a match between two DNA profiles is equivalent to a unique identification.[296] The NRC Report recommends that, "[r]egardless of the calculated frequency, an expert should—given the relatively small number of loci used and the available population data—avoid assertions in court that a particular genotype is unique in the population."[297]

J. WHO IS AN EXPERT?

One of the difficulties courts have faced in dealing with the statistical dispute is determining which scientists have relevant expertise. This issue is important in every jurisdiction, but has special significance under *Frye*, which requires that a novel technique be generally accepted "in the relevant community." Disputes over the

---

the basis on which we decide the DNA testing issue that we would not permit the admission of test results showing a DNA match (a positive result) without telling the jury anything about the likelihood of that match occurring."); *Ex Parte* Perry, 586 So. 2d 242, 254 (Ala. 1991); People v. Barney, 10 Cal. Rptr. 2d 731, 742 (Cal. Ct. App. 1992)("The statistical calculation step is the pivotal element of DNA analysis, for the evidence means nothing without a determination of the statistical significance of a match of DNA patterns.")

295 NRC REPORT, *supra* note 12, at 74.

296 For example, in State v. Pierce, No. 89-CA-30, 1990 WL 97596, (Ohio Ct. App. July 9, 1990), *aff'd*, 597 N.E.2d 107, 113 (Ohio 1992), Dr. Daniel Garner of Cellmark Diagnostics testified that in his "opinion, those two DNA banding patterns match. In other words, it is the same DNA banding pattern. The DNA is from the same individual." Similarly, in State v. Cauthron, 846 P.2d 502 (Wash. 1993), a case where the prosecution presented evidence of a "match" between DNA profiles but chose not to present any statistical estimates of the frequency of the matching profiles, Dr. Robin Cotton of Cellmark testified: "The defendant Richard Cauthron using our test is the source of the semen samples in the five cases that we got the result on. His pattern matches the pattern from the semen stain in those cases." *Id.* at 515. And Ms. Amy Corey, a Cellmark technician testified that Cauthron "was the donor of the semen in those five cases." *Id.* at 516. The Washington Supreme Court quite properly reversed Cauthron's conviction, finding that "[t]estimony of a match in DNA samples, without the statistical background or probability estimates, is neither based on a generally accepted scientific theory nor helpful to the trier of fact." *Id.* at 515.

297 NRC REPORT, *supra* note 12, at 92.

scope and boundaries of the relevant scientific community have been frequent as prosecutors have tried to define the community in a manner that excludes or marginalizes scientific critics, while defense lawyers have tried to make the critics central.

A number of the experts who have been called by the defense to testify about theoretical issues in population genetics, such as the implications of population structure for the independence of alleles, are evolutionary biologists who do experimental work with fruit flies.[298] Prosecutors have argued that these experts are not qualified to comment on the frequency of DNA prints because their expertise concerns flies rather than humans. This boundary drawing maneuver proved successful in some cases,[299] perhaps in part because it was supported by testimony from experts on the FBI's scientific staff, although it has been disavowed by most scientists.[300] To assess such a challenge to the expertise of a scientific witness obviously requires a court to delve rather deeply into the nature and foundations of scientific expertise in esoteric fields. Thus, the notion that the *Frye* rule is easy to apply and that it protects judges from the need to delve deeply into scientific issues has not proven to be true for forensic DNA evidence.

Other boundary disputes concern scientists' familiarity with

---

298 Because it breeds profusely, using diploid, sexual reproduction, with a very short time between generations, the fruit fly is an ideal organism for studying the transfer and distribution of genetic characteristics through large populations over multiple generations.

299 In two *Frye* hearings in California, a judge issued written findings indicating that "human population genetics" and "non-human population genetics" are distinct fields and that "Non-human Population Genetics is not within this relevant scientific community. Drosophila experts in the field of Evolutionary Biology are not within the relevant scientific community." People v. Martinez, No. C-82183 (Orange Co., Cal. Super. Ct. June 18, 1991) (findings of Theodore Millard, J., concerning the admissibility of a DNA test performed by the Orange County, California Sheriff's Crime Laboratory). The same findings were issued by Judge Millard in People v. Gross, No. C-75486 (Orange Co., Cal. Super. Ct. Apr. 4, 1991) (a case concerning the admissibility of the FBI DNA test).

300 For example, Professor Bruce Weir, who is perhaps the most distinguished population geneticist to testify in favor of current statistical procedures, has denied that there is any distinction between human and non-human population genetic. Record at 90, People v. Halik, No. VA-000843 (Los Angeles Co., Cal. Super. Ct. July 23, 1991). Weir stated that the techniques used for evaluating the presence of population structure are the same for all diploid sexual organisms and that it would be wrong to conclude that a population geneticist who studies fruit flies is not part of the relevant scientific community for evaluating the assumptions underlying forensic DNA statistics. *Id.* Eric Lander has gone even further, asserting that "there's a funny sense in which a drosophila population geneticist is better qualified to talk about theory than a human population geneticist can talk about theory." Record Vol. XVII at 191, United States v. Yee, 134 F.R.D. 161 (N.D. Ohio 1991).

pertinent facts. A common attack against scientists in an opposing camp is that they are unaware of certain key facts (often recent, unpublished research findings) that would (or should) alter their opinion. In theory, a court need not consider how well informed the disputing scientific camps are in order to apply the *Frye* rule. The court need only determine whether there is a dispute in the "relevant scientific community." However, a number of courts have considered scientific knowledge a factor in defining the relevant community, and have looked for general acceptance only among "knowledgeable scientists."

This application of *Frye* can enmesh the court in complex technical disputes about whether particular scientific findings do or do not resolve scientific concerns raised by critics. Moreover, carried to its logical extreme, it can lead to ludicrously narrow requirements for general acceptance.[301] Yet courts have found it difficult to resist the invitation to ignore ignorant voices, and have been all too willing to undertake the assessment task that follows from accepting such an invitation. The success of lawyers on both sides at discrediting opposing experts by exposing gaps in their knowledge has provided an incentive for lawyers to rely on seasoned experts—repeat players who stay on top of every new development.[302]

IV. WILL FREQUENCY-OF-MATCH STATISTICS ASSIST THE TRIER OF FACT?

It may seem self-evident that statistics on the frequency of matches in a reference population, if valid, will be helpful to a jury evaluating DNA evidence. But these statistics can be misleading because, in some cases, they are an inaccurate index of the probative value of the DNA evidence.[303] In such cases, courts should consider

[301] One population geneticist testified that the FBI's statistical technique was "universally accepted" among knowledgeable scientists. When confronted during cross-examination, with criticisms of the FBI's methods made by other population geneticists (Hartl, Lewontin, Lander, Mueller), he acknowledged their expertise but discounted their opinions because they were unfamiliar with an important set of data. Further cross examination revealed that the data in question had not been published, were not publicly available, and had been analyzed only by the witness himself. Under the witness' narrow definition of the relevant community, "universal acceptance" was tantamount to acceptance by a single individual, himself. Record at 99-120, People v. Halik, No. VA-00843 (Los Angeles Co., Cal. Super. Ct. July 23, 1991) (Testimony of Dr. Bruce Weir).

[302] It also has placed a premium on strategic use of discovery proceedings, as each side tries during the discovery phase to hold back information that can be used to ambush opposing experts in court, while trying to force disclosure of information that might be used in similar fashion by the the other side.

[303] *See* Richard Lempert, *Some Caveats Concerning DNA As Criminal Identification Evidence: With Thanks to the Reverend Bayes*, 13 Cardozo L. Rev. 303, 325 (1991); Koehler, *supra* 291,

carefully whether this statistic will "assist the trier of fact to understand the evidence or to determine a fact in issue"[304] and whether the probative value of this statistic is "outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury . . . ."[305]

A poor quality match is one circumstance in which a frequency estimate might be misleading. As discussed earlier, the liberal matching standards employed by some laboratories may allow a match to be called between two prints unlikely to be from the same person.[306] In such a case, the DNA evidence may deserve relatively little weight. Yet the statistic jurors typically receive in such cases is the same as if the match were perfect.

The danger of prejudice may also arise where there is reason to believe that the probability of a coincidental match in the case at hand is higher than the frequency of matches in a reference population. This would be the case whenever the defendant was selected in a manner that renders him more likely to match than an average member of the reference population. In a case where the true perpetrator is likely to be a close relative of the innocent defendant, for example, the frequency of matches in a reference population may greatly understate the probability of a coincidental match.[307] Courts should consider carefully whether population frequency statistics would even be helpful to the trier of fact in such cases.

The population frequency of a particular profile also understates the probability of a coincidental match in cases where more than one potential match could prove incriminating. In some cases, for example, a defendant is matched to an evidentiary sample that has more than one donor. If the evidence sample contains alleles A, B, C and D, and defendant is genotype AB, he clearly is a possible donor and so will be declared to match. (The additional alleles are consistent with a second donor of genotype CD). In such cases, forensic laboratories all too often present statistics on the frequency of the defendant's genotype. This practice is highly misleading because the defendant's is only one of a number of genotypes that could be said to match. Defendant would also have matched had he had genotypes AC, AD, BC, BD, CD, AA, BB, CC, or DD. Thus, the

---

at 222; Thompson & Ford, *supra* note 95, *passim*. *See also infra* notes 307-10 and accompanying text.

304 FED. R. EVID. 702. *See supra* notes 46-47 and accompanying text.

305 FED. R. EVID. 403. *See supra* note 54 and accompanying text.

306 *See supra* note 87 and accompanying text.

307 NRC REPORT, *supra* note 12, at 86-87; Lempert, *supra* note 303, at 308-14.

probability of a coincidental match in such cases may greatly exceed the frequency of defendant's genotype.

Finally, frequency statistics are a misleading index of the value of DNA evidence when there is a significant chance of a false positive. To evaluate evidence of a DNA match between two samples, one must know the likelihood of finding such a match if the samples are from different people—an event that could occur in two ways: there could be a *coincidental match* between different individuals who happen to have the same DNA profile, or there could be a *false positive* (false match) due to an error. Even if the frequency of matches in a reference population accurately reflects the likelihood or a coincidental match, it says nothing about the likelihood of a false positive. Consequently, the frequency is not an accurate index of the value of DNA evidence where a false positive is possible. Indeed, the frequency may not even be corrollated with the probability jurors need to know in order to assess the value of the DNA evidence. Hence, opponents of DNA evidence will argue that frequency statistics, by themselves, are unhelpful and therefore excludable under *Daubert*.

The probability of a false positive has received far less attention than the frequency of DNA profiles, although having an accurate estimate of the probability of a false positive may be *more* important than having an accurate estimate of the frequency of matching profiles, because the rate of false positive is likely to be much greater than the rate of coincidental matches.[308] Without an accurate estimate of the false positive rate, one cannot evaluate the probative value of DNA evidence.

In courtroom testimony, forensic scientists and their academic supporters often insist that the probability of a false positive is zero—that is, that false positives are impossible.[309] Although the

---

308 Koehler, *supra* note 291, at 22 ("based on the little evidence available to data, a reasonable estimate of the false positive error rate is 1-4 percent"); Thompson & Ford, *supra* note 10, at 141-46; Lempert, *supra* note 303, at 325 (the probability of a coincidental match "is usually dwarfed by the probability of a false positive error").

309 *See* People v. Shi Fu Huang, 546 N.Y.S.2d 920, 921 (N.Y. Sup. Ct. 1989) ("Dr. Baird testified that it is impossible to get a false positive."); People v. Wesley, 533 N.Y.S.2d 643, 652 (N.Y. Sup. Ct. 1988) ("[I]t is impossible under the scientific principles, technology and procedures of DNA Fingerprinting (outside of an identical twin), to get a 'false positive'—i.e., to identify the wrong individual as the contributor of the DNA being tested . . . . Under the undisputed testimony received at the hearing, no 'wrong' person, within the established powers of identity for the test, can be identified."); Hicks v. State, 860 S.W.2d 419, Tex. Crim. App. 1993), *aff'g* No. 88-134-CR (Freestone Co., Tex. 1989) ("According to Caskey, a false positive finding was impossible . . . ."); State v. Cobey, 559 A.2d 391, 392 (Md. Ct. Spec. App. 1989) ("An incorrect match is an impossible result."); *see also* Koehler, *supra* note 291, at 228 n.16 (quoting a number of similar

falsity of these claims has been demonstrated repeatedly,[310] the claims still continue to be made. Koehler suggests that such testimony stems from "definitions of false positive error that exclude consideration of human error . . . ."[311] Indeed, when confronted with evidence that false matches have occurred in proficiency tests and casework,[312] the standard response is that these events resulted from "human error" by a technician executing the test, not from a deficiency in the test itself. Be that as it may, the effort to distinguish "human error" from "test error" is pointless and potentially misleading when humans are necessarily involved in the execution of the test, when occasional human errors are inevitable, and when it is the overall rate of error (from whatever cause) that is at issue.

Currently, juries often hear nothing about false positives, other than broad assurances that they never occur. When such claims are challenged, forensic experts typically concede that false matches are possible but claim the likelihood of such an event is vanishingly small and that a false match has never occurred in their laboratory or, if it has, that it resulted from problems that have been corrected and will not reoccur. Experts for the defense sometimes testify that false positive are possible, but they typically do not attempt to quantify their frequency. As a result, jurors hear impressive numbers that appear to quantify with precision the frequency of the DNA profile, accompanied (when the issue is raised at all) by a vague, non-quantitative discussion of the chances of a false positive. The danger of the current approach is that the possibility of a false positive will simply be ignored.[313] Indeed, the considerable time devoted in some trials to discussion of the frequency of the DNA profile, the methods for computing that frequency, and the contro-

---

statements from transcripts of expert testimony). *Cf.* Dan L. Burk, *DNA Identification: Possibilities and Pitfalls Revisited*, 31 JURIMETICS J. 53, 80 ("Bald statements or broad hints that DNA testing is infallible . . . are not only irresponsible, they border on scientific fraud.").

310 Thompson & Ford, *supra* note 10, at 136, 143-44; Koehler, *supra* note 291, at 228; Lempert, *supra* note 303, at 323-328.

311 Koehler, *supra* note 291, at 228.

312 *See* Thompson & Ford, *supra* note 10, at 143 (discussing a case where Lifecodes found a perfect match between samples from the defendant and the victim!); NRC REPORT, *supra* note 12, at 88-89.

313 "The syndrome is a familiar one: If you can't count it, it doesn't exist. . . . Readily quantifiable factors are easier to process—and hence more likely to be recognized and then reflected in the outcome—than are factors that resist ready quantification. The result, despite what turns out to be a spurious appearance of accuracy and completeness, is likely to be significantly warped and hence highly suspect." Lawrence H. Tribe, *Trial By Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1361-62 (1971).

versy surrounding those methods may reinforce the powerfully prejudicial suggestion that false positives are a minor issue and that the frequency of the matching print is the issue on which the value of DNA evidence will turn.

The NRC Report calls for laboratory error rates to be determined based on proficiency testing and disclosed to juries.[314] According to the report, accurate estimates of error rates require proficiency tests that are externally administered, blind, and based on samples that are truly representative of case materials.[315] To date, however, relatively little proficiency testing has been done, and most of it has not been blind.[316] Hence, there currently is not a firm scientific basis for determining the rate of false positives.

V. LESSONS FROM THE DNA WAR

While debate has raged over DNA profiling, new genetic identification procedures are already being developed. Molecular biology is advancing rapidly, and forensic scientists have been eager to adapt the new DNA technology to their purposes. New forensic tests involving more sophisticated techniques are already beginning to appear.[317] What lessons have we learned from the "war" over the first round of DNA technology that will help in evaluating the new tests?

A. NEW FORENSIC TECHNIQUES MUST BE VALIDATED

A major source of uncertainty about DNA profiling was the failure of forensic laboratories to validate the techniques adequately at the outset.[318] Proponents assumed too readily that the acceptance of RFLP analysis for research and genetic diagnosis assured it would be reliable for forensic identification. They failed to anticipate all of the difficulties that would be raised by the transfer of RFLP technology to forensics. In retrospect, it is clear that some widely cited

---

314 NRC REPORT, *supra* note 12, at 88-89.

315 *Id.* at 55, 88. In an ideal proficiency test, the samples simulate those processed in actual casework and the analyst is "blind" both to the correct result and to the fact that he or she is being tested.

> Nonblind proficiency tests may not provide a good indicator of the error rate in actual case work because the technicians may be unusually diligent and cautious when they know they are being observed and tested. For example, the observed technicians may be reluctant to declare matches in ambiguous situations to avoid the stigma of having committed a false positive error.

Koehler, *supra* note 291, at 228.

316 *Id.* at 229; Thompson & Ford, *supra* note 10, at 141-45.

317 *See supra* notes 36-38 and accompanying text.

318 See NRC REPORT, *supra* note 12, at ch. 2.

claims about forensic DNA tests were simply false,[319] while other claims lacked an adequate scientific foundation.[320] Courts have been slow to catch and correct these errors.[321]

Whether all of the criticisms raised in court were valid remains to be seen. Only time and additional research will tell, for example, the extent of population structure or the rate of false positives in forensic DNA testing. But regardless of whether the situation turns out to be as good as supporters hope or as bad as critics fear, these concerns should have been raised and evaluated. Indeed, recent developments have been kind to the critics. Leading scientists have sided with the earliest critics and the NRC Report has acknowledged the seriousness of their concerns.

Perhaps the most valuable and important aspect of the NRC Report is that it sets forth, in considerable detail, the steps that a laboratory should take to validate a new forensic DNA typing procedure. The NRC's validation standards take into account not only the range of problems that have arisen in connection with DNA profiling, but also issues likely to arise in connection with newer technology. When evaluating the admissibility of current and future DNA technology, courts should be very attentive to whether the validation research recommended by the NRC Report has been done. If it has not, then a strong case can be made for excluding the laboratory's results either on grounds that they fail to meet the general acceptance standard of *Frye* or that they are insufficiently validated to constitute "scientific knowledge" under *Daubert*.

B. NEW FORENSIC TECHNIQUES MUST BE EXPOSED TO BROAD AND OPEN SCIENTIFIC SCRUTINY

Exaggerated claims about the new tests probably arose, in part, from proponents' enthusiasm about a promising new technique. The claims were then perpetuated by commentators who confused

319 *See, e.g.*, *supra* notes 303-04 and accompanying text for the claim that DNA tests are incapable of producing a false positive and for the claim that loci on different chromosomes are necessarily in linkage equilibrium. Compare testimony of Dr. Patrick Conneally, in People v. Axell, 1 Cal. Rptr. 2d 411, 420 (Cal. Ct. App. 1991) with NRC REPORT, *supra* note 12, at 77-82, and Lewontin & Hartl, *supra* note 10, at 1747.

320 *See, e.g.*, *supra* note 214 and accompanying text for the claim that VNTR alleles are statistically independent.

321 The erroneous statement that DNA profiling is incapable of producing a false positive, for example, has been remarkably persistent. *See, e.g.*, People v. Wesley, 589 N.Y.S.2d 197, 200 (N.Y. App. Div. 1992) ("[I]t is *impossible* under the RFLP procedure to obtain a false positive result, i.e., to identify the wrong individual as the contributor of the DNA being tested.").

the commercial hyperbole of laboratories marketing the new tests with scientific fact.

1. *Forensic Science has Inadequate Internal Mechanisms for Detecting Problems and Correcting Errors*

The persistence of inadequately validated assertions may also reflect flawed error-checking mechanisms in the field of forensic science. Forensic DNA tests were initially shielded from broad scrutiny by the scientific community. They were developed by commercial laboratories whose procedures were proprietary, whose work product was generally not available for examination by outsiders (except when used in a court hearing), and whose laboratory protocols were initially available only by court order. The forensic scientists who developed and validated the techniques in both commercial and government laboratories were well aware that they would be called upon to defend the new tests when their admissibility was challenged in litigation, and that they had a professional stake in the outcome of that litigation.[322] In that atmosphere, they had little incentive to air and discuss problems, and strong incentive to suppress doubt and uncertainty for fear that candid statements would be used against them in court.

It is not surprising, then, that the emergence of DNA profiling tests was accompanied by what one commentator called "scientific happy talk" rather than critical evaluation.[323] Until recently, there has been almost no discussion in the scientific literature of the novel aspects of forensic DNA tests (particularly the matching and statistical procedures) that have become controversial. Given the institutional structure of forensic science, future innovations are likely to emerge in the same manner.

2. *Scientific Critics May be Slow to Emerge*

At first, nearly all of the academic scientists who tesified about DNA evidence were supportive; some defense lawyers reported they were unable to find any scientists who were critics. Proponents of DNA evidence were undoubtedly more successful in recruiting experts, at first, because they were helped by the forensic laboratories, while opponents were left to their own devices. Many criminal de-

322 Burk, *supra* note 309, at 79-80 (Employees of forensic labs "have a clear pecuniary interest in the acceptance of DNA evidence by the courts. The success of their employers and the stability of their own employment depends upon continued use of DNA testing.").

323 Statement of Eric Lander at the meeting of the National Association of Criminal Defense Lawyers (March 16, 1990).

fense lawyers wasted their time in a fruitless search for scientists who would challenge the fundamental principles and theory of DNA testing, not realizing that the real issue is not the validity of the underlying theory but the reliability of the scientific protocols designed to implement the theory. Of course, most academic scientists also did not realize that this was an issue, because they were unaware of the details of the forensic procedures. In most of the hearings that occurred before 1989, the defense failed to produce a single witness to challenge the reliability or general acceptance of forensic DNA tests, and efforts of defense lawyers to raise concerns through cross-examination were notably ineffective.[324]

Defense counsel eventually recruited experts from the ranks of academic scientists. None of these defense experts had any previous experience in forensic testing. Some learned about forensic DNA testing from the published articles and conference presentations of those engaged in forensic DNA testing; most, however, first became familiar with the details of forensic DNA testing only upon being retained by criminal defense lawyers to examine DNA evidence and give a second opinion.

Some of these experts were slow to appreciate the full range of problems that might arise in forensic testing because they often saw only one or two cases. Their limited exposure to casework gave them little basis for assessing the range of possible results for a particular procedure and limited opportunity to observe and document problems that might occur in some cases but not others. Experts who speculated about problems that might occur (but had not occurred in the case at hand) were dismissed as having raised "hypothetical problems."[325] In fact, problems dismissed as "hypothetical" in early cases often appeared in later cases. This experience should alert courts, particularly appellate courts, to the danger of foreclosing challenge to a new procedure before casework experience, and critical scrutiny by independent scientists, reveal the full range of problems.

3. Evaluation of New Forensic Techniques is Impeded by Secrecy

Efforts of defense experts to evaluate the tests were also impeded by the secrecy of forensic laboratories. Protracted battles over discovery erupted in 1989 and 1990 as defense lawyers, guided by their new experts, sought more information. Information emerged piecemeal and fueled a running dispute that pitted groups

[324] Hoeffel, *supra* note 11, at 511-15.

[325] People v. Axell, 1 Cal. Rptr. 2d 411, 425 (Cal. Ct. App. 1991).

of critics and supporters against each other in one hearing after another.

Generally, defense experts have been allowed to see the laboratory reports, lab notes and autoradiographs in the case in question.[326] However, questions over access to laboratory protocols, population data bases, and the raw data underlying laboratory validation studies and proficiency tests have led to extensive controversy and protracted litigation.

The forensic laboratories have resisted disclosing these materials on a variety of grounds, including claims that the materials are legally irrelevant,[327] constitute trade secrets[328] and are privileged.[329] The fight against disclosure has been led by prosecutors with the active assistance of private lawyers hired by the commercial laboratories and FBI staff counsel.

Defense lawyers fought hard for disclosure and gradually gained ground. In several early cases, defense lawyers spent more time and money in efforts to obtain information which their experts had requested than they did in litigating the substantive issues. A ruling of the Minnesota Supreme Court holding the Cellmark DNA test inadmissible based, in part, on Cellmark's refusal to disclose information requested by the defense, marked an important early

---

326 However, this has not always been the case. In one capital case in Texas, which the author has studied, the prosecutor provided to the defense in advance of trial only his own expurgated (and incorrect) summary of what the laboratory report stated. Hicks v. State, 860 S.W.2d 419 (Tex. Crim. App. 1993), *aff'g* No. 88-134-CR (Freestone Co., Tex. 1989). The prosecutor's actions are discussed in Simon Ford & William Thompson, Letter, *Reply*, SCIENCES July/Aug. 1990, at 10.

327 A summary of some of the legal arguments may be found in the order of a federal magistrate, granting discovery to the defense in United States v. Yee, 129 F.R.D. 629 (N.D. Ohio 1990).

328 The trade secret argument lost its force after commentators pointed out the fundamental incompatibility of the claim that protocol is a trade secret and the claim that it is generally accepted by the scientific community within the meaning of the *Frye* standard:

> On the one hand, [forensic laboratories] seek admissibility for their tests, arguing that their procedures are sufficiently known and proven to be regarded as generally accepted by the scientific community. On the other hand, they seek to keep their tests confidential, arguing that they contain procedures or processes sufficiently unique and innovative to constitute trade secrets. Courts would be well advised to tell these companies that they cannot have it both ways. If they wish to assert that their procedures are accepted, they must open themselves to scrutiny by the scientific community so that their assertion can be put to the test.

Thompson & Ford, *supra* note 18, at 60.

329 The FBI claimed that some of the scientific studies it had conducted to check the accuracy of its DNA tests were privileged, arguing that forced disclosure of the results of such research would deter the agency from doing such research in the future and would therefore be a bad policy. *See Yee*, 129 F.R.D. at 629.

victory for defense lawyers.[330] An Iowa District Court helped defense lawyers gain more ground when it compelled the FBI to disclose results of its proficiency testing studies.[331] The most important discovery ruling, however, came in *United States v. Yee*, where a federal magistrate granted the defense access to the raw data underlying the FBI's validation research and population data bases.[332] The wisdom of this ruling quickly became apparent when defense experts found instances in which conclusions published in scientific journals by FBI employees were not supported by the underlying data. Nevertheless, efforts to obtain discovery of such data have met strong resistance and have only been partly successful.

Even when courts have ordered discovery, compliance has been grudging. Internal memoranda obtained through court-ordered discovery from the FBI show that the agency contemplated destroying its own scientific data concerning the performance of its DNA test in proficiency trials rather than turn the data over to defense lawyers.[333] The adversarial nature of the discovery process has engendered game-playing by the litigants. Some defense lawyers have made unnecessarily broad discovery requests in the hope of at least obtaining something; the forensic laboratories have stonewalled, provided information in misleading formats, exaggerated the time and expense required to produce materials,[334] and generally re-

330 "The fair trial and due process rights are implicated when data relied upon by a laboratory in performing tests are not available to the opposing party for review and cross examination . . . . The defense request for . . . specific information regarding its methodology and population data base was denied by Cellmark." State v. Schwartz, 447 N.W.2d 422, 427 (Minn. 1989).

331 Record Vol. II at 57, State v. Smith, No. 41733 (Polk Co., Iowa Dist. Ct. Dec. 18, 1989). When the judge in *Smith* held a hearing on a defense request for discovery of the results of the FBI's internal proficiency tests in late 1989, FBI administrator James Kearney, Section Chief of the FBI's Forensic Science Research and Training Center, flew to Des Moines, along with FBI Legal Counsel L.W. McFarland, to argue against such an order. In his testimony during a pretrial hearing, Kearney suggested that a court order compelling disclosure could spell the end of meaningful internal proficiency testing in forensic laboratories; labs would be reluctant to do such work because the results could be used in court "to pistol-whip us." *Id.* at 64. The Court nevertheless granted the discovery request, becoming the first court in the United States to do so. *Id.* at 57.

332 United States v. Yee, 134 F.R.D. 629 (N.D. Ohio 1991). Courts should be mindful of fears expressed by some that "retrospective validation by a laboratory already heavily invested in a particular technology or procedure will be half-hearted, goal-directed, self-congratulatory, or even deceitful." Thompson & Ford, *supra* note 10, at 151.

333 Memorandum from FBI Legal Counsel to Assistant Director, FBI Laboratory Division (April 20, 1990) (introduced in evidence in *Yee*).

334 In a recent civil paternity action, for example, one commercial laboratory estimated the cost of complying with a request for information about studies validating its procedures to be $684,925. After a lengthy hearing on the issue, the court found that the laboratory's response to discovery "lacked credibility" and that the laboratory had

leased data in a piecemeal manner that has made it difficult for experts in any particular case to become fully informed.

Scientific scrutiny of the forensic tests was also impeded by the use of protective orders. In cases where courts have ordered extensive disclosure of information by forensic laboratories, prosecutors have often requested that the court also issue a confidentiality order to protect alleged trade secrets and prevent defense experts from taking unfair advantage of information provided to them in discovery.[335] In some cases, these orders have placed severe limits on what use defense experts could make of the information by forbidding them from discussing it with anyone not involved in the case and specifically barring them from making reference to the information in any publication, lecture or conversation about forensic DNA testing.[336] Defense lawyers complain that forensic laboratories have used these protective orders to frustrate defendants' preparation of their cases, prevent necessary consultation among scientific experts and keep problems in their techniques secret.

Scientific discussion has also been impeded by concerns about who has the right to comment upon and publish data that is made public in court hearings. For example, in two instances the FBI has refused to allow defense experts to cite or rely upon unpublished FBI data in academic publications.[337] These refusals were apparently based on a desire to protect the FBI employees' right of first publication of data which they collected (at government expense). However, the refusal to allow defense experts to use the data prevented much-needed public scrutiny and commentary upon the

---

wanted "to charge defendant money to produce . . . information that did not exist." Rodick v. Cazale, No. 379-874 (Parish of Jefferson, La. Dist. Ct. Dec. 5, 1991).

335 Forensic scientists have expressed fears that their unpublished data may be "stolen" by defense experts or used in a manner or for purposes they disapprove. Fear of disclosure is also said to hinder efforts of forensic scientists to establish collaborations because "[o]ther scientists . . . are afraid we will be forced to turn [their data] over in discovery before they have had a chance to finish their studies, formulate valid and substantiated interpretations, and personally present the results." Affidavit of FBI employee Bruce Budowle in support of the Government's motion opposing the defendant's discovery request at 8, *Yee*, 129 F.R.D. at 629.

336 *See, e.g.*, State v. Ferguson, No. 591717 (St. Louis Co., Mo. Cir. Ct. June 25, 1991) (Order relating to discovery of DNA typing material).

337 One episode is described in Christopher Anderson, *FBI Attaches Strings to its DNA Database*, 357 NATURE 618 (1992). Professor Seymour Geisser was told that he might publish an analysis of the FBI's database only if he allowed an FBI scientist (Dr. Budowle) to be co-author and granted him right of final approval over the content of the article. Geisser rejected this arrangement and, as a result, was unable to publish the data. A nearly identical case had arisen earlier. Professor Laurence Mueller sought the FBI's permission to publish an analysis of some data from the FBI's database and received and rejected an identical offer from the FBI.

FBI's procedures. The data had repeatedly been relied upon in litigation but has not been (and may never be) published.

The recent report of the National Research Council argues powerfully for greater openness and public scrutiny of the techniques used in DNA testing. Noting that some laboratories had used methods that "were not experimentally supported," the NRC Report traced the problem to the absence of "open scientific scrutiny" and to "the failure to publish a detailed explanation and justification of methods."[338]

> Presenting scientific conclusions in a criminal court is at least as serious as presenting scientific conclusions in an academic paper. According to long-standing and wise scientific tradition, the data underlying an important scientific conclusion must be freely available, so that others can evaluate the results and publish their own findings, whether in support or in disagreement. There is no excuse for secrecy concerning the raw data. Protective orders are inappropriate, except for those protecting individual's names and other identifying information, even for data that have not yet been published or for data claimed to be proprietary. If scientific evidence is not yet ready for both scientific scrutiny and public re-evaluation by others, it is not yet ready for court.[339]

The DNA war would have been shorter and more productive had the NRC's recommendations been followed from the beginning.

4. *Evaluation of New Forensic Techniques may be Impeded by Active Efforts to Suppress Criticism and Discredit Critics*

The acrimony that has made litigation over DNA evidence a "war" stems, in part, from proponents' efforts to silence or discredit critics[340] in order to prevail under the *Frye* standard, and from critics' resentment at being treated in such a fashion. The nasty, personal nature of exchanges between prosecutors and defense experts, both inside and outside the courtroom, has been widely noted.[341] For example, prosecutors and FBI personnel have been accused of making threatening comments and spreading false rumors about scientists who have testified for the defense. Perhaps the most disturbing allegations, however, are that proponents of DNA testing have attempted to corrupt the peer-review process in

---

338 NRC REPORT, *supra* note 12, at 56.

339 *Id.* at 93-94.

340 *See generally* Gina Kolata, *Critic of 'Genetic Fingerprint' Tests Tells of Pressure to Withdraw Paper*, N.Y. TIMES, Dec. 20, 1991, at A1; Leslie Roberts, *Prosecutor v. Scientist: A Cat-and-Mouse Game*, 257 SCIENCE 733 (1992); Anderson, *supra* note 14, at 753.

341 Roberts, *supra* note 9, at 753.

order to prevent the publication of articles critical of the forensic tests.

In one episode, a federal prosecutor twice called Professor Daniel Hartl and allegedly pressured him to withdraw an article that had already been accepted for publication by *Science*.[342] According to Hartl, who found the calls "chilling," the prosecutor cited the "political consequences" of publishing the article and "possible disasterous consequences for future DNA fingerprint-based prosecutions."[343] After Hartl refused to withdraw the paper, a prominent scientific supporter of DNA testing, Dr. Thomas Caskey, who has been a frequent prosecution witness, persuaded the editor of *Science* to reconsider publication of the article.[344] It was eventually published, but only after Hartl and his co-author, Richard Lewontin agreed to "tone down" their criticisms.[345] Furthermore, at Caskey's recommendation, the editor took the unusual step of soliciting a rebuttal article from two proponents of DNA testing, which was published in the same issue. Caskey told *Nature* that he had intervened in the editorial process of *Science* because "he was concerned that 'publishing defence testimony in a scientific journal' gives it such weight that courts might reopen, perhaps to overturn convictions obtained on the basis of DNA evidence."[346]

Another defense witness, Professor Laurence Mueller, has been the target of numerous public attacks by a California prosecutor.[347] The prosecutor has sent a series of letters criticizing Mueller to editors considering publication of Mueller's scientific articles and to Mueller's department chair and University Chancellor. The letters make disparaging statements about the quality of Mueller's scientific testimony and suggest that he is an unethical individual whose work warrants special scrutiny. The prosecutor characterizes Mueller's ideas as "knuckle-headed" and claims that they "could conceivably result in a vicious, violent criminal being freed to continue to prey on society."[348]

In yet another episode, a Minneapolis prosecutor was tipped off that a defense witness, Professor Seymour Geisser, had submitted an article to *Genetics*, and that the article had been rejected based on

342 Kolata, *supra* note 340, at A1; Roberts, *supra* note 7, at 732; Anderson, *supra* note 7, at 753.

343 Anderson, *supra* note 7, at 500.

344 Roberts, *supra* note 7, at 735.

345 Anderson, *supra* note 7, at 500.

346 *Id.*

347 Roberts, *supra* note 340, at 753.

348 *Id.*

negative reviews.[349] The prosecutor apparently learned about the bad reviews before Geisser himself. Fifteen minutes before Geisser received, by fax, the comments of the editor and reviewers on his article, he received a fax from the prosecutor requesting that he bring the reviews with him to court when he came to testify the following week. Subsequent inquiry revealed that the author of the most negative review was Dr. Ranajit Chakraborty, a frequent prosecution witness, who often is at loggerheads with Geisser in the courtroom. Chakraborty has collaborated with FBI scientists on research designed to validate the FBI's DNA test, of which Geisser has been critical, and was co-recipient of a $200,000 grant from the Department of Justice to support additional validation research. He did not inform the editor of *Genetics* of any potential conflict when he was asked to review Geisser's article.[350]

These efforts to suppress criticism probably stemmed from proponents' attempts to meet the general acceptance standard imposed by *Frye*. The *Frye* standard is designed to block the admissibility of evidence about which there is a serious scientific controversy. As the California Supreme Court once put it, if "scientists significant either in number or expertise publicly oppose . . . [a scientific technique] as unreliable," a court may safely conclude that it should not be presented to the jury.[351] Consequently, the admissibility of the technique in court is threatened when scientists begin speaking out in opposition to it.

The only effective way to deal with critics is to silence them or discredit them. Proponents cannot simply argue that the critics are wrong (although they may sincerely believe this). Under the general acceptance standard, courts are not supposed to decide which scientists are right and which are wrong in a scientific dispute; the mere existence of a serious dispute over the fundamental value of scientific evidence renders it inadmissible. If proponents debate the critics in a respectful manner they implicitly admit the existence of the dispute, which is fatal to their cause. Therefore, name-calling and personal assault, rather than respectful debate, persists in this area. Courts should recognize this unfortunate reality and not assume that experts targeted by disparaging attacks are necessarily unworthy of notice.

---

349 Christopher Anderson, *Coincidence or Conspiracy?*, 355 NATURE 753 (1992); Roberts, *supra* note 7, at 736.

350 *Id.*

351 People v. Shirley, 723 P.2d 1354, 1377 (Cal. 1982).

VI. CONCLUSION

It is hard to disagree with Eric Lander's observation that, "when you try to manage the quality of scientific evidence, the legal system is a very, very blunt instrument."[352] The legal debate polarized the scientific community and created perverse incentives for secrecy, posturing, personal attacks and other behavior contrary to the norms and values of science. The great majority of pretrial hearings have been poorly organized, poorly funded affairs in which scientists (often recruited at the last minute and placed on the stand after only minimal consultation) were questioned by lawyers with a fuzzy understanding of the scientific issues. Important advances in the litigation have come in only a relative handful of hearings in which defense counsel managed to retain experts in advance, obtain information through pretrial discovery, and present it in an organized manner. As the inefficiencies of the current approach become more apparent, so will the need for a better approach to evaluating novel scientific evidence. Like most wars, the DNA war evokes a powerful feeling in retrospect that there must have been a better way.

Nevertheless, the courtroom debates have served a useful purpose. Forensic laboratories have been forced to open themselves to much-needed scientific scrutiny. For all of its disadvantages, the adversarial system at least assured that someone looked at the new techniques with critical eyes and tried to explore their limitations.

Perhaps the real problem is that we rely too heavily on admissibility hearings for screening new scientific innovations. These hearings would be less costly and more manageable if new forensic procedures received broader scrutiny before they came to court. The NRC Report makes a number of helpful suggestions to this end. First, the report condemns the sort of secrecy that has surrounded RFLP-based tests and urges that laboratory protocols, data bases and validation studies be open and available for public scrutiny.[353] Second, the report urges the creation of a National Committee on Forensic DNA Typing, consisting of distinguished independent scientists and legal experts, that would provide advice to forensic laboratories and review new technologies before they went to court.[354] The new committee would have no formal regulatory authority, but would be a body to which courts might look for authoritative guidance when evaluating the admissibility of a new

352 Roberts, *supra* note 7, at 736.
353 NRC REPORT, *supra* note 12, at 56, 93-94.
354 *Id.* at 70-72.

technique. Third, the report suggests that a formal system be developed for accreditation of laboratories doing forensic DNA testing, that admissibility of test results in court be contingent on the laboratory being accredited, and that accreditation depend on complying with rigorous standards for quality assurance, such as participation in externally administered proficiency tests.[355]

These proposals have great merit. They would not eliminate the need for judicial review of new innovations, but would help assure that future DNA wars are more limited than the first.

355 *Id.* at 97-110.