Am. J. Hum. Genet. 59:497–500, 1996

INVITED EDITORIAL
The Second National Research Council Report on Forensic DNA Evidence

Bruce S. Weir

Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh

The National Research Council has issued an update of its 1992 report on the forensic uses of DNA. This 1996 report endorses the use of DNA profiling for human identification and acknowledges that "some of the statements in the 1992 report have been misinterpreted or misapplied in the courts" (National Research Council 1996, pp. ES-1 and O-1). In what is likely to become their most important finding, the 1996 report's authors "foresee a time when each person can be identified uniquely (except for identical twins)," and they outline some calculations to show when this expectation reasonably can be expected to be fulfilled.

DNA profiling has been used for individual identification since 1985 and has become a routine tool in forensic and paternity-testing laboratories. When the profile from a biological stain does not match that of a specific individual, that person is excluded as a possible contributor to the stain. When there is a match, however, there has been debate over the most appropriate way to express its evidentiary strength. The debate had reached such a level by 1991 in the 1996 report that the National Research Council set up a committee to examine the issues. The resulting 1992 report also endorsed the forensic use of DNA profiling, and it provided a useful handbook for explaining the nature and utility of the profiles (National Research Council 1992). However, the sections concerning the population-genetic and statistical issues involved in the quantification of matching profiles were rather weak.

Even in 1991, forensic DNA profiles were often based on so many VNTR loci, at each of which so many alleles could be distinguished, that the resulting number of possible genotypes far exceeded the world population size. Estimating multilocus genotype frequencies for specific profiles could not rest on "counting" methods, which simply count how many times the profiles were seen in population samples. An appreciation for the rarity of specific profiles could instead be obtained by multiplying together the frequencies of the constituent alleles, and the resulting extremely small products were often attacked. Neither the 1992 report nor the 1996 report pointed out that these products do not estimate the frequencies of profiles in the current population. Instead these products provide the probabilities of genotypes that could be formed by the random union of extant alleles. In essence, these frequencies refer to infinite populations and need not be limited to values such as the reciprocal of the population size.

The 1992 report sought to settle arguments about the validity of the product rule by saying that it ought to be applied to "ceiling" allelic frequencies rather than to estimated frequencies. This was intended to allow for variation in allelic frequencies among groups within populations. This ad-hoc rule quickly proved to be unworkable. The 1996 report is unequivocal: "We share the view of those who criticize [the ceiling principle] on practical and statistical grounds and who see no scientific justification for its use" (National Research Council 1996, pp. 0–27, 5–31). Similar short shrift is given to the "counting method" advocated by the 1992 report. So many loci can be used in current DNA profiles that a specific profile will not be seen in forensic databases that do not include the profile donor.

The 1996 report offers a scientific way of accommodating population structure. The worry has been that the people who might be considered contributors to biological samples in a particular crime may belong to a group whose allele frequencies are higher than those in the general population. Using estimates of the general frequencies could, therefore, lead to an underestimate of the profile frequency and be prejudicial to people who match simply by chance. The 1996 National Research Council report points to the several published studies of forensic marker frequencies that show less variation between subgroups within racial groups than between racial groups. This means that quoting frequencies from different racial groups provides an upper limit on the variation within racial groups.

The 1996 report uses published accounts of VNTR

Received May 22, 1996; accepted for publication June 25, 1996.

Address for correspondence and reprints: Dr. Bruce S. Weir, Department of Statistics, North Carolina State University, Box 8203, Raleigh, NC 27695-8203. E-mail: weir@stat.ncsu.edu

This article represents the opinion of the author and has not been peer reviewed.

0002-9297/96/5903-0003$02.00

data to recommend that profile frequencies be assigned a "confidence interval" formed by both multiplying and dividing the frequency by 10. The data that suggested this factor were giving profile frequencies of the order of 10^{-8}. Unlike rules based on statistical procedures such as bootstrapping, this ad-hoc factor-10 rule does not give uniformly sensible results. As the profile frequency becomes larger, the limits of the confidence intervals should become smaller multiples of the estimate, and the factor of 10 is not appropriate for PCR-marker frequencies of the order of 10^{-2} or 10^{-3}. Nor does the factor of 10 take account of sample size.

Nevertheless, substructure can increase homozygote frequencies in the whole population, over those expected from the Hardy-Weinberg law. This effect is known as the Wahlund principle and is a consequence of the average value, over subpopulations, of the square of an allele frequency being greater than the square of the average allele frequency. Estimating a one-locus homozygote frequency as the square of the whole-population allele frequency would therefore give an underestimate. The 1996 report recommends use of a correction factor for homozygotes when the object is to provide profile frequencies, where the correction factor reflects realistic levels of human population structure. If homozygote frequencies are increased by population structure, it follows that heterozygote frequencies will, on average, be decreased. With this logic, the 1996 report recommends that no correction be made for heterozygotes. This is even though, as they acknowledge, some heterozygotes may also have increased frequencies. Of course, it will not be known whether any particular heterozygote is one that has a frequency over- or underestimated by the product rule, and a prosecutor who follows the 1996 report and does not also adjust heterozygote frequencies is likely to be challenged. The law is concerned about particular cases, rather than about what happens on average.

The problem disappears if attention is shifted from profile frequencies to conditional frequencies. Instead of determining the chance of seeing a profile in some random person, the question becomes the chance of seeing a second occurrence of the profile, given that it has been seen once already—in the suspect, for example. Conditional frequencies are greater than products of allelic frequencies, for both homozygotes and heterozygotes, and they arise naturally in likelihood ratios.

The forensic-science literature is clear on the merits of phrasing the strength of trace evidence in terms of likelihood ratios, whereby the probability of the evidence under one explanation is compared to the probability under an alternative explanation. In the DNA context, this means the probability of finding a specific profile in an evidentiary stain when it either is supposed to have come from a specific person whose type is known or it comes from some unknown person. The cursory rejection of this approach by the 1992 report is remedied in the 1996 report. As part of the discussion of likelihood ratios, the 1996 report shows how the likelihood ratio reduces to the reciprocal of the profile frequency when suspect and perpetrator have independent frequencies. The 1996 report embraces the ratio of conditional frequencies, using results of Balding and Nichols (1994), when suspect and perpetrator (if different) are known to belong to the same subpopulation. Because such knowledge is likely to be lacking, it would seem better to use an approach that makes as few assumptions as possible and not to require membership of independent populations. The population-structure parameter values involved in the conditional frequencies could always be chosen to reflect the knowledge available for the case. The point is that the same general approach would lead to appropriate results in all cases. As it is, the forensic scientist is being told to decide whether to use conditional frequencies.

Giving central prominence to likelihood ratios would have made the 1996 report's calculations given for relatives, in addition to those given for members of the same subpopulation, appear more naturally. In its discussion on uniqueness, the 1996 report stresses the need to calculate the conditional probability of a profile, given that it has already been seen, and this view could well have been adopted throughout the 1996 report. Although the authors of the 1996 report have not adopted the general use of conditional frequencies, they at least have recommended that research be conducted into how best to present evidence to the trier of fact. Such research should be of value, but there can be little doubt that the purely population-genetic considerations are best accommodated by likelihood ratios expressed in terms of conditional probabilities.

The 1996 report does give a strong endorsement to the use of likelihood ratios in the interpretation of stains from more than one contributor. A careful discussion is given, and the resulting logical treatment employing conditional frequencies is emphasized. We have shown recently (Weir et al., in press) that the naïve method of the 1992 report, based on "random man not excluded" thinking, actually can be prejudicial to a defendant whose profile is not excluded from a mixed stain. The 1992 report called for reporting only the frequency with which a random individual would be excluded from being a contributor to a mixed stain, ignoring the fact that the stain is known to be from more than one contributor. The hurried treatment of mixed stains in that report was the cause of just one of the barriers to the presentation of sound analyses in court (Weir 1995), a barrier that has been removed by the 1996 report.

The 1996 report echoes the sentiments of forensic scientists such as Evett and Buckleton (in press), who

decry overemphasis on statistical testing for departures from the Hardy-Weinberg frequencies. Instead, the 1996 report assumes that departures do exist and considers whether such departures are large enough to be important. The proposed correction factor for homozygotes is to ensure that courts are not misled about the strength of DNA evidence. Readers will need to note the disclaimer in the 1996 report—that is, that the score test discussed in chapter 4 is not a test for Hardy-Weinberg frequencies and that studies on the use of exact tests (e.g., see Maiste and Weir 1995; Zaykin et al. 1995) show satisfactory power for detecting departures from Hardy-Weinberg frequencies, at multiallelic loci. The staunchest critics of the product rule have stressed the difficulty of testing for independence between alleles over several loci. However, two-locus tests have been found to behave satisfactorily, and, with its characteristic emphasis on scientific reasonableness, the 1996 report considers it unlikely that there will be dependence at larger numbers of loci when there is no evidence for dependence at one or two loci. Citations are given of studies that bolster this view by comparing observed and expected numbers of matches at several loci.

Although an overemphasis on independence testing is indeed to be avoided, the 1996 report's implication that testing be abandoned seems to err in the other direction. If the parameterization that the 1996 report gives for possible departures from Hardy-Weinberg frequencies is used, it can be shown that findings of significant departures are likely to indicate large values of these parameters—certainly larger than the values of 0.01 or 0.03 suggested as correction factors for homozygotes. It would surely be better to continue with testing. Although failure to reject the hypothesis of independence does not establish the existence of Hardy-Weinberg frequencies and does not obviate the need for corrections to the product rule, when the tests do reject the hypothesis other modifications may be necessary.

A statistical recommendation of concern is the procedure to be followed when a suspect is identified by a search of a DNA-profile database. Such databases are now being set up in many states, and already samples have been collected from several hundred thousand convicted offenders. If the profile in question has a population frequency of P, then NP copies are expected in a database of size N. If the estimated value of P is $\tilde{P}$, the 1996 report advocates reporting $N\tilde{P}$, instead of just $\tilde{P}$, as a match probability. This procedure is called for because NP is an upper bound on the probability that at least one of the database profiles matches the evidence profile. However, the likelihood ratio for assessing the strength of the match still involves the probability of the match if the person identified is not the source of the evidence profile, and this has not been changed by the fact of a database search. Indeed, as Balding and Donnelly (in press) have shown, there is information in the fact that $N - 1$ people have been excluded as possible contributors. Certainly there would be no need for an NP calculation if N was the size of the population and if P was $1/N$; finding a match would be certain, and there would be no need to reduce the evidentiary weight of the match. The 1996 report acknowledges that more-complicated calculations will be needed for large databases, and it could prevent almost certain argument in court by detailing the calculations to which it refers.

The 1996 report's treatment of how to quantify the extent of population structure is briefer than it could have been. There is a danger of confusion between the quantities w_k, defined as the proportion of a population in the kth subpopulation, and the sizes n_k of samples from these subpopulations. The 1996 report provides an explanation of general principles rather than details for estimating measures of population substructure. Use of subpopulation proportions w_k has generally implied a fixed population approach, instead of a random population approach where replication provides a basis for expectation of statistics. It is the expected values that are written in terms of the structure parameters θ. The distinction between fixed and random populations, along with estimation procedures, has been given by Cockerham and Weir (1986). Because the numerical consequences of these different analyses are not great, however, the issue should have no impact on forensic conclusions.

The 1996 report devotes a chapter to laboratory performance, reflecting the shift in the debate to issues of possible false declarations of a match between two profiles. The thought has been expressed that there is little merit in presenting profile frequencies of 1 in 1 million if there is a 1-in-100 chance of laboratory error. A careful discussion is given, in chapter 3 of the 1996 report, of the need for quality control and quality assurance and for the accreditation of forensic laboratories. These steps all require proficiency testing and laboratory audits, and the 1996 report endorses the continuation of these practices. A clear distinction is made between the results of proficiency tests and the results of laboratory performance on a specific case. The 1996 report reiterates the 1992 report's statement that proficiency-test error rates should not be combined with profile-frequency estimates. Instead, the 1996 report points to one way of reducing the possibility of error—by duplicate and independent testing of samples. Although this will not always be possible, the merits of such duplication are obvious. Looking further ahead, it may well be that the contamination issue will be resolved by technical advances, such as PCR amplification, that will allow typing of samples to be completed, at a crime scene, before a suspect is even identified.

At several points, the 1996 report mentions the "cre-

ative misapplication" of the 1992 report (National Research Council 1996, p. O-27). A specific citation is given to the case of State v. Guevara (1993) in which results were presented after "combing through VNTR data on many subgroups to find the largest allele frequencies, taking the upper end of a confidence interval for each such frequency, ignoring loci because large samples indicate that alleles for some other locus do not occur in Hardy-Weinberg proportions, and using fixed-bin frequencies with extremely wide bins" National Research Council 1996, pp. 6–21). Such procedures "were neither contemplated by the 1992 committee nor reflect reasonable scientific judgments," but they did cause profile statistics not to be allowed in that case. I find the call for scientific reasonableness to be most refreshing, and I would like to share the 1996 report's optimism that admissibility of DNA-profile statistics should no longer be in doubt. However, I note that the defense expert in the State v. Guevara case employed the same procedures in a Pennsylvania admissibility hearing after reading the 1996 report.

Apart from addressing population-genetic and statistical issues, the 1996 report opens with a useful description of the molecular basis of DNA typing and closes with a thoughtful discussion of the legal implications of the conclusions and recommendations made in the body of the 1996 report. Consequently, the 1996 report will be a valuable resource to the forensic and legal communities. The authors of the 1996 report are to be congratulated on their efforts to make recommendations on the basis of scientific arguments. They have done much to help the proper calculation and presentation of DNA-profile statistics, and the day on which DNA profiles are employed with the same trust as are fingerprints has surely been brought forward by their report.

References

Balding DJ, Donnelly P. Evaluating DNA profile evidence when the suspect is identified through a database search. J Forensic Sci (in press)

Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64:125–140

Cockerham CC, Weir BS (1986) Estimation of inbreeding parameters in stratified populations. Ann Hum Genet 50:271–281

Evett IW, Buckleton JS. Statistical analysis of STR data. Proc 16th Int Cong Forensic Haemogenet (in press)

Maiste PJ, Weir BS (1995) Comparison of tests for independence in the FBI RFLP databases. Genetica 96:125–138

National Research Council (1992) DNA technology in forensic science. National Academy Press, Washington, DC

——— (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington, DC

State v Guevara. K9-92-1873, Minn Dist Ct, Dakota Cty, Jan 26, 1993

Weir BS (1995) DNA statistics in the Simpson matter. Nat Genet 11:365–368

Weir BS, Triggs CM, Starling L, Stowell LI, Walsh KAJ, Buckleton J. Interpreting DNA mixtures. J Forensic Sci (in press)

Zaykin D, Zhivotovsky L, Weir BS (1995) Exact tests for disequilibria with arbitrary numbers of loci and alleles. Genetica 96:169–178