APPENDIX

Critique of
Chapter 3: Statistical Basis for Interpretation

The following is a critique of Chapter 3 which, though not exhaustive, emphasizes the FBI's primary concern that the majority opinion of the scientific community regarding the statistical basis for interpreting forensic DNA test results is not adequately presented. Other issues raised here, if left unaddressed, will serve to weaken the credibility of the final report.

Lack of Balance

Chapter 3's discussion of approaches to statistical interpretation is limited to a minority viewpoint within the scientific community. Only scientists whose views agree with the author have been referenced in Chapter 3, while the majority view of scientists (including paternity testers) who advocate the product rule is not addressed. The majority viewpoint and relevant comments from reviewers should be given greater attention to produce a more balanced perspective.

Chapter 3 concludes that allelic data from VNTR's can only be "assumed" to be independent, and that the product rule is inappropriately applied until independence can be demonstrated. (3-4-9) However, pertinent information on statistical independence from Drs. Chakraborty, Devlin, Evett, Budowle, and Weir, which was available to the committee, was not cited.

Some of the information at issue was presented at a forensic meeting sponsored by the University of California, Riverside, in March 1991. Dr. Oskar Zaborsky, project officer for this NRC study, attended the meeting at which Budowle, Devlin, Evett and Weir presented data on statistical independence.

At the meeting, Dr. Ian Evett provided Dr. Zaborsky with a packet of his analyses using a Bayesian approach to address the issues for arriving at a statistical estimate and the effect of population substructure (personal communication Evett to Budowle). Although the FBI does not advocate the Bayesian approach (3-25), it is another method for arriving at a statistical estimate that warrants discussion in Chapter 3. The Bayesian approach is currently used in some paternity laboratories in the United States. It should be noted that the Bayesian approach utilizes the product rule. None of this information was incorporated in Chapter 3.

Reviewers of Chapter 3 also provided the committee with references supporting use of the product rule and the minimal affect that subpopulations can be expected to have on the viability of the product rule. This information was not included in Chapter 3 (see Hernandez et al., Human Genetics 85:343-348, 1991 and Sokal et al., Nature 1991).

Only five references concerning VNTR population data and substructure issues are cited in Chapter 3 to support the author's opinions and conclusions (3-9-23). There are at least 25 additional papers, plus numerous scientific presentations and personal communications that should have been cited. (A bibliography is attached.)

Failure to Ask the Right Question

Chapter 3 does not ask the fundamental question forensic scientists must answer in interpreting DNA casework results. The question the author should have asked is: What is the chance of anyone else depositing the biological evidence at the crime scene? Suspects generally claim they did not contribute the evidentiary material. Legally, a suspect must be presumed innocent, even if his DNA profile matches the DNA evidence. The defendant's ethnic background is irrelevant unless one assumes the suspect is guilty. Some reviewers have suggested this as an issue to be addressed, but their comments were ignored.

Inconsistencies and Inaccuracies

The introduction to Chapter 3 states: "It is meaningless to say that two patterns match without providing a scientifically valid estimate ... of the frequency..." (3-1-24). However, the author's statement is inconsistent with a subsequent statement that "...a match occurring at each of the four loci by chance is probably quite rare..." (3-37-2).

The author should know that a match constitutes a failure to exclude the suspect as a potential source of the evidence, and does not establish an absolute link between the suspect and the evidence. When coupled with the fact that a match is rare, a failure to exclude, even without an estimate of the frequency, can be very informative to the trier of fact.

It is inconsistent for the author to suggest that the inquiry on issues pertaining to forensic application of VNTRs should be limited to published data while, at the same time, relying heavily on a personal communication from R.C. Lewontin (3-10-22, 3-12-9, 3-12-19, and 3-14-6). This is a curious point because pertinent unpublished data from scientists (e.g., Budowle, Chakraborty, Devlin, Evett, Kidd, and Weir) who have actually analyzed VNTR databases also could have been obtained by personal communication. In fact, some of the same information was presented at scientific meetings attended by some members of the committee. Such information, however, does not support the author's thesis.

The author's discussion in Chapter 3 of the "the time-honored way to estimate frequency [is by counting]..." is not accurate (3-3-9 to 3-3-10). While it is correct that for protein genetic markers (except for HLA) the frequency for each phenotype of each genetic marker was estimated by counting, a multi-locus estimate was achieved by multiplying the frequency of each phenotype, not by counting.

As was pointed out by a reviewer of the first draft of Chapter 3, an extensive discussion of the statistical issues was conducted previously by scientists working in the paternity testing field. Chapter 3 contains no discussion of issues addressed previously in the paternity field which are relevant to forensic DNA testing.

The author opines that human population samples have been collected traditionally in a random fashion (3-3-10). This statement is neither accurate nor reflective of the manner in which human population samples typically have been collected. Rather, the great majority of population samples collected for establishing frequency estimates have been obtained from blood banks and paternity testing laboratories. Special sampling regimes were not normally undertaken for the collection of samples, as suggested in Chapter 3.

Another point to consider is that most genetic marker data, which come from samples collected in no special manner, have met the standard for applying the product rule. Collecting population samples in a statistically random manner has been shown to be unnecessary.

The author suggests that a conservative approach applied to the forensic data presented in Caldwell v. Georgia was reduced from 24,000,000:1 to 250,000:1, and was attained by simply counting (3-19-13 through 14). Lifecode's database, however, does not contain 250,000 samples; thus, counting could not have been the approach used. It would be more insightful and accurate to describe how the estimate was derived and who suggested the approach.

Use of "some" and "others" to describe the number of courts that accept the product rule (3-19-3, and 3-32-18 through 23) are apparent misrepresentations. To say "some courts" may suggest to an uninformed reader that equal numbers of courts have admitted and rejected the product rule. The great majority of courts, in fact, have admitted statistics which are based on the product rule.

Endogamy and Propinquity

The author argues that endogamy and propinquity can affect the extent of population substructuring. Referring to Spuhler and Clark (1961) (3-13-4), he states that "one-third of the marriages are contracted between persons living less than 10 miles apart." That means, however, that two-thirds of marriages are between persons living more than 10 miles apart. It is important to note that the Spuhler and Clark study is based on data from 1900 to 1950 when U.S. population was less mobile than today, thereby suggesting that endogamy and propinquity have little effect on subpopulations.

Chapter 3 identifies South Boston and St. Paul as examples of ethnic enclaves (3-13-6 through 9). The author implies that ethnic groups in these locales may demonstrate VNTR frequencies which differ significantly from the general population. Accordingly, the author suggests that until proper population studies are done it is inappropriate to estimate

frequencies using the product rule. In a subsequent section, however, the author questions the effect that enclaves may be expected to have on the distribution of VNTR alleles, noting that "such regional studies are much less sensitive than ethnic group studies, because each region contains a mixture of persons from different groups" (3-20-13). VNTR population data collected by forensic scientists participating in the Technical Working Group on DNA Analysis Methods (TWGDAM) are regional data based on DNA samples collected from one blood bank. If ethnic or regional enclaves exist to the degree suggested, regional VNTR data collected by TWGDAM participants should be expected to reflect the "anticipated" differences. The data does not, however, reflect such differences.

Incorrect Analysis of Subpopulation Data

Citing data from R.C. Lewontin, the author compares Polish with Italian data to illustrate the issue of potential subpopulations (3-14). As noted below, these data were incorrect. However, this example is not illustrative of the effect of subpopulations found in the United States. Rather, the appropriate question is: How different is a given subgroup from the general population or pool? Left unaddressed is whether the ethnicity of the suspect is a proper concern.

The data on 3-11-9 and 3-14 are troubling. The 14.5% value for cDe in Poles, in fact, can be found in Mourant. Unfortunately, the validity of the data underlying the estimate is not discussed. The 14.5% value comes from a 1947 study from Wroclaw, Poland. However, a 1966 study from the same town observed a 4.7% value for Cde. The estimate resulting from the subsequent data further challenges the acceptability of comparing certain Poles to certain Italians.

This inaccuracy points to a more basic problem of applying older data. There is an apparent lack of knowledge regarding technical limitations of immunoassay data collected during the 1940's and early 1950's. When more complete, appropriate, and recent data are analyzed, the exaggerated differences are not found.

Also, ABO (3-11-1) and disease (3-11) genes are known to have forces of selection working upon them. There are no known effects of selection on VNTRs. The examples used are misleading and not informative, and should be changed. While an example of disease genes in Finns does not apply to VNTR genes, data available in two papers by Sajantila et al. (1991) demonstrate that two DNA markers, HLA-DQ alpha and D1S80, show no statistical differences between Finns and U.S. Caucasians [Intl J Leg Med, 104:181-184; Am J Hum Gen (in press)]. These data are not cited by the author.

Lack of Corroborative Support for Direct Sampling

The author suggests "...direct sampling of ethnic subgroups is required" (3-5-11 and 3-19-18). This is a minority viewpoint, but it is the only viewpoint discussed in Chapter 3.

Interestingly, Dr. Eric Lander recently recognized the existence of three schools of thought on the same issue (letter to the editor [Am J Hum Gen (1991), 49:90]. The latter position is a more balanced view.

Acton et al. (1990) are cited in Chapter 3 as evidence for the existence of four-fold differences in VNTR allele frequencies for some regions of the United States (3-20-15). However, it is essential to note that the study did not employ the conservative criteria referenced on 3-7-4. Rather, Acton et al., used allele frequency bins smaller than the measurement error for agarose gel electrophoresis, which is clearly not "a more conservative rule for counting population frequencies" (3-7-4). There are no four-fold differences using the FBI's fixed-bin approach to categorize the same data analyzed by Acton et al.,

No Scientific Basis for Ceiling Approach

The ceiling approach and the 10% minimum allele frequency advocated in Chapter 3 are not scientific, but rather ad hoc approaches for statistical estimation. A better approach would be to investigate the available data to provide a scientifically acceptable means of estimating the frequency of DNA profiles. Chapter 3 ignores or discounts relevant reviewers' comments on this issue which question the arbitrary 10% minimum frequency advocated by the author. More importantly, some reviewers point out that the ceiling principle described in Chapter 3 is misguided. Since an individual is not composed of one allele, a ceiling on allele frequencies is incorrect. The ceiling should be based on the product of the alleles which is rare regardless of the database employed.

Population Substructure and Genetic Diversity

Phenomena other than substructure can lead to deviations from expected results in applying RFLP technology (3-8-14), a point recognized by Dr. Eric Lander in a letter to the editor [Am J Hum Gen (1991), 49:92], in which Dr. Lander states that blank alleles (i.e., a technical limitation of agarose gel electrophoresis) must be taken into account. Chapter 3 does not address this issue.

The reference to a statement by R.C. Lewontin in 3-18-16 through 18 (i.e., studies show that genetic diversity within races is greater than the genetic variation between races) may have been misinterpreted by the author and, accordingly, warrants further review. Interestingly, Lewontin's conclusion in the final paragraph of his 1972 paper suggests that subgroups are amazingly similar to each other.

More thought should be given before recommending that laboratories undertake collection of samples from genetically homogeneous populations (3-21-16 and 3-24-18). Points to consider include: What constitutes a genetically homogeneous group?; Should sample collection be targeted at a country, state, city, or neighborhood?; and, more importantly, Is it relevant?

As one reviewer suggests, ultimately, all unrelated individuals are ethnically distinct.

Limitations of the Counting Method

The counting method also has limitations (3-36). With a database of 500 individuals, the estimate can be no lower that 1/500. If four loci are analyzed, it is 1/500; if ten loci are analyzed, it is 1/500; if 20 loci are analyzed, it is 1/500. In addition, the counting method does not account for any principles of genetics.

Failure to Recognize Fixed-Bin Approach

The author states that some laboratories use conservative approaches to determine statistical estimates (3-7-7). In fact, almost all forensic laboratories in North America apply the fixed-bin approach for estimating the frequency a DNA profile, a point not mentioned at 3-26-16. Nowhere does Chapter 3 discuss the viability of this approach. See 3-2 25, 10, 11 (AM)

Poor Choice of Analogies

The Porsche and Nordic analogies cited in Chapter 3 are weak to the point of being irrelevant (3-4 and 3-5). More realistic analogies for VNTRs exist and should be used instead. The same criticism applies to the comparison of two subpopulations having allele frequencies of 1% and 20%, respectively. There is no example of such dramatic differences among the VNTR markers employed by forensic laboratories. The 20% example seems especially inappropriate given the 10% minimum ceiling frequency advocated by the author.

Meddling in Legal and Funding Issues

Whether or not a court compels an individual to provide a blood sample for DNA testing in the course of a criminal investigation (3-28-11) is a legal, not a statistical, issue. Furthermore, the suggestion to re-examine serology cases (3-30-11) is a legal issue and should not be addressed in a chapter on statistical interpretation. Finally, the author should not suggest or recommend the FBI as a source of funds for followup studies (3-36-8). References to these issues are inappropriate and should be removed.

BIBLIOGRAPHY

(1) Waye et al - 1990 Promega Proceedings

(2) Budowle et al (1991) Am J Hum Genet 48:841-855

(3) Budowle et al (1991) Am J Hum Genet 48:137-144

(4) Chakraborty et al (1991) In: DNA Fingerprinting Approaches and Applications (Burke et al, eds)

(5) Edwards et al (1991) Am J Hum Genet 49:746-756

(6) Edwards et al (1991) Genomics (in press)

(7) Edwards et al 1991 Promega Proceedings

(8) Odelberg et al (1989) Genomics, 5:915-924

(9) Evett and Gill (1991) Electrophoresis 12:226-230

(10) Chakraborty et al (1991) Am J Hum Genet (in press)

(11) Chakraborty and Jin (1991) Mol Biol and Evol (in press)

(12) Gill et al (1991) Int J Leg Med 104:221-228

(13) Akane et al (1990) J for Sci 35:1217-1225

(14) Chakraborty et al (1991) Crime Lab Dig (in press)

(15) Devlin et al (1991) Am J Hum Genet 48:662-676

(16) Devlin et al (1990) Science 249:1416-1420

(17) Devlin et al (1991) Science 253:1039-1041

(18) Devlin et al (1991) J Am Stat Assoc (in press)

(19) Brinkmann et al (1991) Int J Leg Med 104:81-86

(20) Henke et al (1991) Int J Leg Med 104:33-38

(21) Evett et al (1990) J For Sci Soc 31:41-47

(22) Baird et al (1986) Am J Hum Genet 39:489-501

(23) Yokoi et al (1990) Jpn J Hum Genet 35:235-244

(24) Yokoi et al (1990) Jpn J Hum Genet 35:179-188

(25) Wong et al (1986) Nuc Acids Res 14:4605-4616

(26) Gasparini et al (1990) Hum Hered 40:61-68

(27) Budowle et al (1991) Crime Lab Dig 18:9-26

(28) Wiegand et al (1991) In: DNA - Technology and Its Forensic Application, Berghaus et al (eds) pp 121-127

(29) Rothamel et al (1991) In: DNA - Technology and Its Forensic Application, Berghaus et al (eds) pp 128-133

(30) Fattorini et al (1991) In: DNA - Technology and Its Forensic Application, Berghass et al (eds) pp 134-140

(31) Morris and Brenner (1991) In: DNA - Technology and Its Forensic Application, Berghaus et al (eds) pp 199-202

(32) Morris and Brenner - 1990 Promega proceedings

(33) Hummel and Fukshansky (1991) In: DNA - Technology and Its Forensic Application, Berghas et al (eds) 28-31, pp 203-207

(34) Hernandez et al. Human Genetics 85: 343-348, 1991.

(35) Sobal et al. Nature 1991. vol 351, 143-5