the free atmosphere, away from the influences of the earth's surface. From all of these measurements will come an understanding of the processes that control hydroxyl and determine the oxidizing capacity of the atmosphere.

REFERENCES

1. G. H. Mount and F. L. Eisele, *Science* **256**, 1187 (1992).
2. S. M. Beck *et al.*, *J. Geophys. Res.* **92**, 1977 (1987); D. R. Crosley and J. M. Hoell, *Future Directions for H_xO_y Detection, NASA Conference Publication 2448* (NASA, Washington, DC, 1986).
3. G. H. Mount, *J. Geophys. Res.* **97**, 2427 (1992).
4. F. L. Eisele and D. J. Tanner, *ibid.* **96**, 9295 (1991).
5. J. A. Logan, M. J. Prather, S. C. Wofsy, M. B. McElroy, *ibid.* **86**, 7210 (1981); D. Perner *et al.*, *J. Atmos. Chem.* **5**, 185 (1987).
6. M. Trainer *et al.*, *J. Geophys. Res.* **92**, 11,879 (1987).
7. C. M. Spivakovsky *et al.*, *ibid.* **95**, 18,441 (1990).
8. A. M. Thompson, *Science* **256**, 1157 (1992).

The Continuing Case of the Florida Dentist

Temple F. Smith and Michael S. Waterman

Applications of DNA fingerprinting or typing are becoming increasingly common and diverse. A suspect can be convicted or set free on the basis of a "fingerprint" of DNA found at a crime scene. In a recent dramatic case, the Nazi war criminal Josef Mengele of the notorious Auschwitz concentration camp was declared to have been dead for more than a decade on the basis of a DNA fingerprint match between a set of Brazilian bones and a sample from his living son. This is an example of the use of DNA fingerprinting to establish relationships, in this case by paternity testing. General applications of DNA fingerprinting in pediatric practice range from paternal identification through genetic disease diagnoses to child abuse verification (*1*). In this issue of *Science* Ou *et al.* (*2*) present an analysis of the viral DNAs associated with the alleged human immunodeficiency virus type 1 (HIV-1) infection of a number of patients by their Florida dentist.

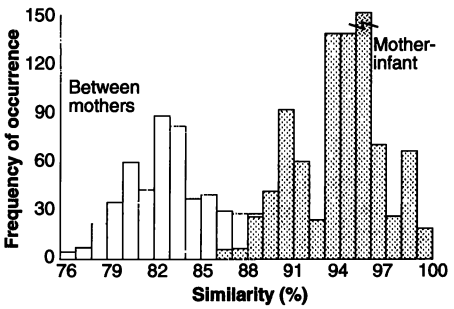
These uses of molecular data raise a multitude of important issues, both scientific and ethical. There are also major unresolved issues as to the levels of statistical and scientific reliability appropriate for deciding public policy or for use in our judicial system. In 1989 Lander (*3*) argued that new experimental standards must be defined for DNA data use in judicial cases but did not directly address the statistical issues. A recent report of the National Academy of Sciences Panel on DNA and Forensic Science, headed by Victor McKusick of Johns Hopkins Hospital (*4*), addresses these and related issues. These questions must be addressed by the scientific community because science is and will continue to be used to formulate public policy and to form judicial decisions. Quite often scientific data and conclusions are used by those unfamiliar with its limitations.

The majority of comparative DNA sequence studies have as their conclusion the generation of scientific inferences. It is ideal if such inferences are supported by overwhelming statistical evidence. However, under the assumption of eventual experimental testing of these inferences, those supported by evidence of even marginal statistical significance can be quite valuable, particularly when deciding which experiment to perform next among the myriad possible experiments. In other situations, statistics play an important but less well-defined role. For example, in the public policy arena, estimates of statistical significance are folded into risk-cost-benefit analyses, cost-effectiveness studies, and policy analyses. Here the links between science, statistics, and conclusions can become very obscure. In these areas, there is not always a guarantee of timely rejection of incorrect inferences as a result of experimentation nor is there an accepted criterion for statistical significance. The statistical significance used by the public and many agencies to recommend life-style changes often is not extremely strong, particularly as it may relate to cancer. One should recall that the great statistician and biologist R. A. Fisher never accepted even the evidence linking smoking and lung cancer; the public certainty has come from evidence accumulated since Fisher's death. However, the analyses of HIV-1 transmission have provided sufficient statistical evidence for most nations to spend considerable money on research, to recommend the wide use of condoms, and to support the use of gloves by all medical personnel.

The statistical criteria used in scientific inference or in public policy generation are distinct from those for setting new legal precedents or for civil or criminal courts. Legal situations require the highest statistical significance to generate a verdict beyond a reasonable doubt! The obvious potential forensic utility of various molecular techniques could be lost if their early judical use is inconclusive or improper. The concern here is not academic as exemplified by Ou *et al.* (*2*) in which statistical analyses of the alleged HIV-1 infection of patients by their dentist were used to conclude that, for at least five of his patients, the dentist was the most likely source of the infection. The work can influence pending insurance cases and public health policies and has already been involved in a Florida court process. The court delayed certain actions until the publication decision of this journal and until all of the data were made public (*5*).

Not only were the authors likely to be quoted or called by the court, but the reviewers were under some additional pressure. A News & Comment piece (*5*) even appeared before the review process was complete. This was due, in part, to the fact that some of the data in the present analysis had been used in an earlier preliminary study (*6*) and that an early copy of Ou *et al.* (*2*) had been obtained through the Freedom of Information Act. In addition, there was some delay in making available the entire data set to all parties concerned.

Thus, it is clearly possible for data, a manuscript, and even the review process to become entangled in the legal process. When there is potential for legal implica-



Distributions of nucleotide similarity between mothers and infants. The intermother histogram is of the nucleotide similarities between all pairs of variants, each pair composed of one variant from two different mothers. The mother-infant histogram is of the nucleotide similarities between pairs of variants, each pair from a mother and her own infant. The sequences were obtained from the GenBank database, where Wolinsky and co-workers have deposited 12 isolates from mother 1, 19 from mother 2, and 23 from mother 3, along with 21 isolates for the infant of mother 1, 13 for that of mother 2, and 27 for that of mother 3.

T. F. Smith is at the Biomolecular Engineering Research Center, Boston University, Boston, MA 02215. M. S. Waterman is at the Department of Mathematics, University of Southern California, Los Angeles, CA 90089–1113.

tions or major public policy, the scientist has a responsibility to decide, in advance, what statistical criteria are most appropriate.

The fingerprint utilized in Ou *et al.* (2) was based on the C2-V3 domain of the HIV-1 envelope gene. Under the hypothesis that the dentist has infected a patient, viral sequences of this highly variable region that are isolated from the patient should be much closer to viral sequences from the dentist than to a sequence from a randomly chosen HIV-1–positive person. However, analysis of the relationships among HIV-1 sequences is more complicated than in situations such as paternity testing. In the more common applications of DNA typing the loci remain fixed for the lifetime of an individual, but HIV-1 is evolving so rapidly that each individual is host to a population of HIV-1 variants. In particular this holds for the C2-V3 domain. There is a large difference between tracing the history of a single-copy gene and tracing the histories of viral variants each evolving in an individual member of the infected population. Furthermore, each variant could have different tissue and transmission behaviors (7).

A relevant study of the C2-V3 region has been published by Wolinsky *et al.* (8). They studied cases of known transmission from infected women to their unborn children for three mother-infant pairs when the infants were 2 to 4 months old. Up to 27 variants were isolated from one of these patients. Unexpectedly, in two of the three pairs a proviral form infrequent in the mother was dominant in the infant. In addition, the infant's viral sequences were less variable than were their mother's. Histograms of the distribution of mother and infant nucleotide sequence similarities, based on data deposited in GenBank, are shown in the figure. There is a small positive overlap between these distributions, suggesting that mothers and infants (where the pattern of transmission is incontrovertible) had some sequences that were just as different from each other as if they had been taken from the unrelated mothers. Mother-infant transmission is one example of HIV-1 transmission modes, but these data illustrate some of the difficulties that arise in studies of the transmission of this virus.

Ou *et al.* (2) have data from the dentist, from seven patients infected with HIV-1, and from 35 local controls (LCs). Two of these patients were identified as having high-risk behavior and thus were not considered as valid tests of the dentist's virus being their infectious source. Additional epidemiological data would have been of interest, such as the randomness of the LC sample including their randomness of times since infection. Time since infection is an important variable, since HIV-1 evolves so rapidly. There is an additional factor that adds difficulty in evaluating the data. For analysis, DNA was first amplified by the polymerase chain reaction (PCR). For the dentist, the patients, and some LCs, this DNA was then cloned into M13 and sequences were obtained from individual clones. Five to 12 cloned sequences were obtained from each patient. Direct PCR sequencing without first cloning was used to determine sequences for all LCs. Direct PCR sequencing is a sampling method that at best generates a sequence composed of the most prevalent base at each position and at worst generates only the sequence of one of the more prevalent viral variants.

The task of using this molecular data to test the hypothesis that the five non–high-risk patients contacted HIV-1 from the dentist is statistical in nature. The Wilcoxon rank-sum test was used by Ou *et al.* (2) and has the following form. We have two samples of numerical values $X_1X_2 \cdots X_5$ and $Y_1Y_2 \cdots Y_{30}$ and wish to test if the X (patient) distribution coincides with the Y (LC) distribution. The X samples and the Y samples are pooled and the 5 + 30 values are ranked, smallest to largest. The Wilcoxon rank-sum statistic S is the sum of the ranks corresponding to the X sample. The hypothesis of identical X and Y distribution is rejected if S is too small or too large. With the molecular data at hand, the ideal test would be to use one dentist DNA to compare with one DNA from each of the 5 patients and the 30 LCs and reject the hypothesis if S is too small. There were at least two DNAs cloned from the dentist and varying numbers of cloned patient DNAs. These unequal sample sizes were handled by using the average distances between the dentist DNAs and those of the patients and LCs. Since averages with different samples sizes have different distributions, there is a problem with using the Wilcoxon statistic with averaged Xs and Ys. Another problem is the distinction between the cloned sequences and the direct sequences, which again starts us off with different Xs and Ys. The authors turned to bootstrapping to get around these problems. Bootstrapping is a statistical procedure that resamples the original data with replacement, obtaining statistical replicates of the original sample. However, bootstrapping a flawed procedure does not correct it, although this cautionary note is not likely to be widely heeded. At the end of reference 23 in Ou *et al.* (2) two other DNA analyses were mentioned that overcome the above objections. One analysis averages the dental cloned sequences from the dentist with direct sequences for all patients and LCs, and the other compares a direct dentist sequence with the other direct sequences. These improved procedures resulted in values of slightly less significance than the value reported in the text.

It is natural to ask the more general and fascinating question of the phylogenetic relationships among these variants. Various modes of transmission, such as blood transfusion or mother to infant, can imply very different phylogenetic trees among the variants. In Ou *et al.* (2) some nucleotide sequences were cloned sequences and others were direct sequences, and the authors preferred to use cloned sequences. Not all LC sequences could be satisfactorily represented in a tree. They chose those LC sequences most closely related to the patient sequences, in keeping with their philosophy of selecting the data most likely to conflict with the hypothesis that the patients contracted HIV-1 from the dentist. In addition, they took the two most divergent clones from each sampled individual. This resulted in use of cloned sequences from the dentist, all seven patients, and four LCs. In addition, two LCs from whom only direct sequences had been obtained were included, as was an LC consensus sequence. Tree building is a difficult art and science and, as in this case, the data are seldom ideal. An indication of the difficulties is given by the analysis in which the authors resampled the 146 informative sequences sites by bootstrapping and observed a grouping of the dentist with the five patients in only 79 of 100 replications.

Ou *et al.* (2) is a pioneering study in the use of DNA typing to resolve individual viral transmission events. This study has opened an important area of inquiry, that of determining the molecular domains and data along with the statistical procedures most appropriate to test such hypotheses. The intersections of scientific and legal issues ensure a controversial and lively future for these discussions. Inferring evolutionary history from molecular data is difficult, and it is easy to come to questionable conclusions. The arguments supporting the "African Eve" hypothesis of the late Alan Wilson and colleagues have been shown to be statistically flawed (9). The biology and statistics of HIV-1 transmission are unlikely to be any less complex.

REFERENCES

1. E. R. McCabe, *J. Pediatr.* **120**, 499 (1992).
2. C. Y. Ou *et al.*, *Science* **256**, 1165 (1992).
3. E. S. Lander, *Nature* **339**, 501 (1989).
4. National Research Council, "DNA technology in forensic science," (National Academy Press, Washington, DC, 1992).
5. J. Palca, *Science* **255**, 392 (1991).
6. *Morb. Mortal. Wkly. Rep.* **39**, 489 (1990).
7. S. S. Hwang *et al.*, *Science* **253**, 71 (1991).
8. S. M. Wolinsky *et al.*, *ibid.* **255**, 1134 (1992).
9. M. Barinaga, *ibid.*, p. 686.