

emerging, they would plausibly agree to be randomized even after that data safety and monitoring board would find randomization intolerable for the typical patient. They may reach their indifference point just when more typical patients would be abandoning it. By recognizing the idiosyncratic value profiles of potential subjects and the legitimacy of idiosyncratic preferences among treatment options, we

must conclude that not only is it rational for some patients to prefer one treatment in a justifiable randomized clinical trial, but that some other patients may rationally accept randomization when the typical patient or researcher would not. Thus some studies may continue even after abstract assessment would lead to the conclusion that they should stop. They can continue for the same reasons that some patients would not

rationally agree to randomization in cases where doctors cannot know which treatment arm is preferable.

*Levine, Robert J.: Should I Enroll in a Randomized Clinical Trial? Excerpts from *A Patient's Guide*. *IRB: A Review of Human Subjects Research* 1988; 10(2): 10-11.

AIDS and The Soundex Code

by Simson L. Garfinkel

Confidentiality has always been a primary concern in AIDS research and treatment. In the early days of the epidemic, when many lay people felt that AIDS was synonymous with homosexuality, confidentiality was necessary to prevent discrimination because of sexual orientation. When the HIV antibody test was developed, confidentiality was critical to protect the jobs, insurance, and privacy of people who had been exposed to the virus but who might not develop the disease for many years.

This spring, the Presidential Commission on the Human Immunodeficiency Virus Epidemic declared that "rigorous maintenance of confidentiality is considered critical to the success of the public health endeavor to prevent the transmission and spread of HIV infection." If the names of people who have AIDS are not kept with a high degree of confidentiality, the commission found, people who have AIDS, out of fear, will not seek counseling or treatment.

But also from the beginning of the AIDS crisis, the need to identify and accurately enumerate AIDS cases has also been significant. It is important to know whether cases reported by different states represent new outbreaks, direct transmission, or simply the same individual moving from one state to another.

In 1981 the Centers for Disease Control in Atlanta maintained a list of

the names of all persons diagnosed with AIDS. The purpose of the list was to prevent duplicate counting, according to Dr. Meade Morgan, chief of the Statistics and Data Management Branch of the AIDS Program at the Center for Infectious Diseases (CID).

By 1982 the list had grown to more than 200 names. At that time, Dr. Morgan said, the CID saw the tremendous potential for abuse that such a list created. Yet at the same time, the need to exchange identifiers between state health departments and the CID was also growing.

That year, Morgan said, the CID turned to a system called Soundex for coding names into encrypted 5-character identifiers. Today AIDS cases are reported not by name but by Soundex code. The Soundex code begins with a letter that is the same as the last letter in the person's name. Following the letter is a four digit number that is derived from the remaining letters in the person's name. The system therefore provides for 260,000 unique identifiers.

The Soundex code accompanying new reports of AIDS are compared against a file containing the Soundex codes of all known AIDS cases, Morgan said. If a match is found, the CID's computer compares the dates of birth on the two matching reports. If both the Soundex code and the dates of birth match, the computer assumes that the second report is the same individual as the first.

Out of the first 20,000 AIDS cases, Morgan said, there were only twenty individuals who had both identical Soundex codes and dates of birth. At

the time the CID's practice was to telephone the two state health departments involved in the match and ask them to confer privately to determine if the two cases were in fact the same person. The result of the conference was reported to the CID. Since then, Morgan said, the CID has discontinued this checking procedure because there are simply too many new AIDS cases to cope with.

Although other identifiers, such as blood type, race, and height, could be used to improve the reliability of the decision in cases in which both the Soundex number and the date of birth match, the CID does not use such information, Morgan said, because these traits are frequently miscoded or subject to change. Even gender occasionally changes because some people with AIDS are transsexuals, Morgan said.

Robert F. Boruch and Joe S. Cecil, authors of *Assuring the Confidentiality of Social Research Data* (University of Philadelphia Press, 1979) called the Soundex code "a numerical alias which can't be decrypted for practical purposes."

The only way to "decrypt" the Soundex code, Boruch said in an interview, is to make a list of every possible name that might be encrypted to produce the same code. But since many different names will encrypt to the same code, it is impossible to know which name the Soundex code actually stands for.

"My name, for example, using their encryption scheme produces an alias which, if you tried to decrypt it ... you would come up with about 8000 possible names at a minimum," he said.

Simson L. Garfinkel is a freelance writer living in Somerville, MA, who specializes in issues of technology policy.