

Strategy and Art in Automated Death Searches

The mathematical theory of record linkage is well developed.¹ But the would-be user will find only a little guidance in the literature concerning the strategies that have been employed in applying the theory, and the degrees of success that may be expected using alternative strategies. It is for this reason that the article on "The California Automated Mortality Linkage System" in this issue of the Journal² is particularly timely. In this editorial, I will summarize some recent experiences in Canada that bear on the same topic. The aim, of course, is to make the automated death searches as accurate and cost-effective as possible, using whatever personal identifiers are available for linkage.

The strategies one may use to increase the efficiency of linkage are not dictated in detail by the theory. But a perceptive human searcher performing such a task will intuitively seek clues with which to help resolve uncertain matches in a variety of ways. The designer of an automated linkage procedure is therefore well advised to do considerable manual searching and resolution and to take careful note of the things that most frequently influence subjective judgments—perhaps almost unconsciously, especially in difficult cases. With the computer linkage system employed at Statistics Canada^{3,4} for searching the Canadian Mortality Data Base,⁵ the flexibility is such that one is free to incorporate virtually any refinement of logic that the human mind finds profitable to employ in a manual search. The principles and practice of this approach have been documented in journal article form⁶ and in an in-house publication which is available on request.*

For present purposes it will be sufficient to indicate the nature of some of the refinements as viewed by a manual searcher, and not to belabor here the conversion into a machine procedure. As specific examples, a human searcher is likely to notice, after a little experience:

- that he is recognizing various degrees of similarity or dissimilarity, and not just simple agreements and disagreements (e.g., as with SAMUEL versus SAM, or perhaps versus SAMPSON, and as with BIRTHDATE discrepancies of one or two or even three years, months, or days);
- that it is sometimes profitable to cross-compare first with second INITIALS or GIVEN NAMES, and vice versa, when the straight comparisons show no similarity;
- that comparisons of identifiers which are not of quite the same kind may often be revealing, as in the case of PLACE OF DEATH versus PLACE OF WORK; in this case, one would not look for agreement as such, but would consider the degree to which any implied migration was likely or unlikely for workers at that place, given the known places of death for other workers from there.

In addition, a human searcher will soon become aware: that a young person is usually less likely to have died in a given period than is an older person; and that searches of a very large file are inherently more likely to yield false positive linkages than are the corresponding searches of much smaller files.

All of these considerations may correctly influence the subjective judgment of a human searcher, because they are intuitively obvious. More importantly for present

* Newcombe HB, Abbatt JD: Probabilistic Record Linkage in Epidemiology: Principles Employed, 1983. Available (price \$5.00) from Mr. S. E. Frost, Eldorado Resources Limited, Suite 400, 255 Albert Street, Ottawa, Ontario, Canada K1P 6A9.

EDITOR

Alfred Yankauer, MD, MPH

ASSISTANT EDITOR

Kenneth J. Rothman, DrPH

EDITORIAL BOARD

Mary F. Arnold, DrPH

(1984), *Chairperson*

Susan P. Baker, MPH (1986)

Paul B. Cornely, MD, DrPH (1985)

Sandra J. Eyres, RN, PhD (1986)

Martin S. Favero, PhD (1985)

Joseph L. Fleiss, MS, PhD (1985)

David Hayes-Bautista, PhD (1984)

Lorraine V. Klerman, DrPH (1986)

Charlotte Muller, PhD (1984)

Kenneth D. Rogers, MD, MPH (1984)

Conrad P. Straub, PhD (1986)

Hugh H. Tilson, MD (1984)

Philip G. Weiler, MD (1985)

Joseph Westermeyer, MD, PhD (1986)

M. Donald Whorton, MD, MPH (1985)

STAFF

William H. McBeath, MD, MPH
Executive Director/Managing Editor

Adrienne Ash, PhD
Publications Director

Doyne Bailey
Assistant Managing Editor

Michelle Horton
Production Editor

Suzanne Morgan
Production/Advertising Assistant

Anne E. Pace
Editorial Assistant

CONTRIBUTING EDITORS

George J. Annas, JD, MPH
Public Health and the Law

Barbara G. Rosenkrantz, PhD
Public Health Then and Now

Jean Conelley, MLS
Book Corner

purposes, they are capable of being quantified. Appropriate refinements based on all of the above factors have, in fact, been incorporated into the automated procedures for mortality follow-up of Canadian uranium miners and other study cohorts in this country. To assess the importance for linkage of a given identifier comparison "outcome" (agreement, disagreement, specified similarity, etc.), one asks a very simple question, "How common, or 'fashionable' is that 'outcome' in genuinely linked pairs of records as compared with unlinked pairs?" "Outcomes" that are more likely to occur in linked pairs will obviously argue for linkage, and those that are more common in unlinked pairs will argue against linkage. The strength of the argument is proportional to the magnitude of the ratio. It is that simple, provided the relevant data are captured, from the files themselves and from the matched pairs of records out of a preliminary linkage operation.

The art of record linkage, as distinct from the theory and general strategy, lies in the choice of the specifics of any such refinements. The designer of a linkage procedure wishing to introduce refinements of the above kinds must inevitably explore the options empirically, and quantitatively in some detail using actual data, before deciding upon some optimum simplification. Although the practice of this art is laborious, experience has shown that substantial improvements in overall discriminating power can, in fact, be achieved as a result of the fine tuning that is possible when the intuitive insights gained from dealing manually with the more difficult linkages are tested quantitatively and used to modify the machine procedures.

From the Arellano, *et al.*, paper in the present issue of the Journal² and from others,⁷ one is encouraged to believe

that the art of automated death searching will continue to be refined at various centers in the future. In essence what is being refined is the judgment of a machine in the light of the experience which, in large part, the machine has made possible. The practical benefit to people will be increased awareness of some of their more important mortality differences.

HOWARD B. NEWCOMBE, PHD, DSC

Address reprint requests to Howard B. Newcombe, PhD, DSc, P.O. Box 135, Deep River, Ontario K0J 1P0 Canada. Dr. Newcombe is a consultant with Vital Statistics and Disease Registries Section of Statistics Canada.
© 1984 American Journal of Public Health 0090-0036/84 \$1.50

REFERENCES

1. Fellegi IP, Sunter AB: A theory of record linkage. *J Am Stat Assoc* 1969; 64:1183-1210.
2. Arellano MG, Petersen GR, Petitti DB, Smith RE: The California automated mortality linkage system. *Am J Public Health* 1984; 74:1324-1330.
3. Hill T: Generalized Iterative Record Linkage System. Special Resources Sub-division, Systems Development Division, Statistics Canada, 12-K R.H. Coats Building, Tunney's Pasture, Ottawa, Ontario K1A 0T6, 1981.
4. Howe GR, Lindsay J: A generalized iterative record linkage computer system for use in medical follow-up studies. *Computers Biomed Res* 1981; 14:327-340.
5. Smith ME, Newcombe HB: Automated follow-up facilities in Canada for monitoring delayed health effects. *Am J Public Health* 1980; 70:1261-1268.
6. Newcombe HB, Smith ME, Howe GR, Mingay J, Strugnell A, Abbatt JD: Reliability of computerized versus manual death searches in a study of Eldorado uranium workers. *Computers Biol Med* 1983; 13:157-169.
7. Rogot E, Feinleib M, Ockay KA, Schwartz SH, Bilgrad R, Patterson JE: On the feasibility of linking census samples to the National Death Index for epidemiologic studies: a progress report. *Am J Public Health* 1983; 73:1265-1269.

The 'Trickle-Down' Theory—Is That Any Way to Make Policy?

Whether there are enough physicians, and whether they are practicing where people need them, have been difficult public policy questions for well over a decade now. During those years, the answers to these questions have shifted several times.

In the middle and late 1960s a consensus emerged that there was a physician shortage in the United States. The Congress responded with enthusiasm and created programs that helped private and state medical schools drastically increase their enrollments. These included capitation support to institutions, traineeships for biomedical research, and scholarships and loans for students.

As the problem was studied more carefully, it was clear that the principal shortages were of primary care physicians and of any physicians in many rural and inner-city areas. Even the remarkable improvements in access brought about by Medicare and Medicaid had not made these problems go away. Again, the Congress responded with funds to help train family physicians, general internists, and general pediatricians, and with programs to identify health manpower shortage areas and to encourage or even require physicians to work in those areas. During this time there was rapid growth in the National Health Service Corps, and in the Migrant and Community Health Centers.

The past few years have seen another major shift in the perception of the physician shortage problem and the appro-

priate solution to that problem. Now, there is general agreement that the total number of physicians is more than adequate to meet the needs of the people of this country. Some have argued that this large supply will create market forces that will eliminate geographic and specialty shortages. Most recently, an optimistic interpretation of a few studies^{1,2} has created a new "conventional wisdom" that such shortages are already or will very soon be a problem of the past. In turn, the need for continuing many of the programs mentioned above has now come into question.

In this issue of the Journal, the article by Hicks³ points to quite a different conclusion. Together with another recent study⁴ Hicks' article makes it clear that even huge overall increases in physician supply in a state leave some areas still in great need. A degree of progress was made, yes—but at what cost? While some of the new physicians went to counties with shortages, nearly eighty per cent settled in areas that already had more than enough physicians to meet local needs. Twenty-four counties, largely rural ones, experienced a decline in their ability to meet the needs of their residents, while the number of physicians practicing in the state grew by one-third. This situation may well get worse as older physicians are not replaced by younger ones in those areas.

No studies are perfect, and these issues are complex⁵ and may never be fully resolved. But it is clear that the