

Automated Follow-up Facilities in Canada for Monitoring Delayed Health Effects

MARTHA E. SMITH, MSc, AND HOWARD B. NEWCOMBE, PhD

Abstract: Increased public awareness of the possible presence of carcinogens and other potentially harmful agents in the workplace and in other areas of the environment has created a demand for studies to determine the extent of the risks associated with exposure to such agents. These studies require large numbers of individuals in various "exposed" and "control" populations to be followed-up over long periods of time. Such large-scale follow-up can be greatly facilitated where information on all deaths and on cases of serious morbidity is accumulated centrally, for a

whole country, in a form that permits rapid searching by computer and in which individuals are well enough identified to minimize the possibility of mistaken identity.

The Canadian Mortality Data Base and the National Cancer Incidence Reporting System are two such centralized follow-up facilities which have been developed in Canada. We describe here the manner in which these files are used, the problems encountered, and their solutions. (*Am J Public Health* 1980; 70:1261-1268.)

Introduction

Increased public awareness of the presence of carcinogens and other harmful agents with possible delayed effects, in the workplace and in other areas of the environment, has led to demands for information on the magnitude of associated risks and the adequacy of current protection standards. To obtain the required data, substantial numbers of "exposed" individuals along with suitable "controls" normally have to be followed over a period of decades to determine whether there are differences in the proportions of individuals that subsequently develop cancer or some other serious condition. In general, the smaller the risks one seeks to detect and measure in this way, the larger the study populations that must be followed.

Traditionally, epidemiological follow-up has been carried out manually and clerically using a multitude of sources, local, regional and national. A list of such potential sources, prepared for the purpose of assisting those working on an American study of toxic goiter and its treatment will serve to emphasize this point. The local resources listed include: city directories, telephone company, post office, board of education, voter registration lists, social service exchange, marriage and divorce records, and group hospitalization. The state and federal resources included: bureau of motor vehicles, retail credit bureau, Veterans Administration, Department of Defense Locator System, Social Security Administration, Internal Revenue Service, Census Bureau, and

Editor's Note: See related editorial, p. 1246, this issue.

Address reprint requests to Ms. Martha E. Smith, Research Analysis Officer, Vital Statistics and Disease Registries Section, Health Division, Statistics Canada, R. H. Coats Building, Tunney's Pasture, Ottawa, Ontario, Canada K1A 0Z5. Dr. H. B. Newcombe is retired from Atomic Energy of Canada Limited, and now serves as a part-time consultant to the Vital Statistics and Disease Registries Section of Statistics Canada.

death certificates. Details of how to tap each of these sources are given in a mimeograph document.*

When carried out by manual procedures, such large follow-up studies tend to be exceedingly expensive. Economical versions of such studies are possible, however, where nationally centralized records of mortality and morbidity exist in computerized form and the searches can be automated, and provided that the machine-readable versions of the records contain sufficient personal identifying particulars so that instances of mistaken identity are rare.

The purpose of the present paper is to describe two such centralized follow-up facilities in Canada, and the manner in which they are being used to monitor very large study populations for delayed effects on health, possibly arising out of various occupational histories as well as medical, diagnostic, and therapeutic procedures. These two facilities are the Canadian Mortality Data Base and the National Cancer Incidence Reporting System.

Both facilities consist of provincial records that have been converted into a standardized format on magnetic tape, under the custody of Canada's central statistical agency, Statistics Canada. Both are used by Statistics Canada under the strict confidentiality rules of the federal Statistics Act.**

*Resources for Locating Patients. Prepared by the staff of the Cooperative Thyrotoxicoses Therapy Follow-Up Study, Washington, DC: US Public Health Service, Draft of Nov. 2, 1962.

**Statistics Canada became the new name for the Dominion Bureau of Statistics with the proclamation of a new Statistics Act (SC 1971, c.15) on May 1, 1971. The bureau was initially set up by statute in 1918 as the central statistical agency for Canada (SC 1918, c.43). The functions of Statistics Canada are to compile, analyze and publish statistical information relative to the commercial, industrial, financial, social, and general condition of the people, and to conduct regularly a census of population and agriculture of Canada as required under the Act.

Materials and Methods

Design Principles

To minimize the cost and maximize the efficiency of both reporting systems, certain simple design principles were followed:

First, existing routine records (the provincial death registrations and cancer registrations) were used. Where information from these was already being keyed to cards or tape routinely for purposes other than epidemiological interest, the resulting machine-readable versions of the records were adapted to form the basis of the follow-up facility.

Second, special attention was paid to the adequacy and accuracy of the personal identifying information. Rather than adding further identifying particulars to the tapes of the older records as a special operation, attempts are being made to have the more important additional items entered at the time of the initial key operation for all current records.

Third, information keyed from the provincial records in the various locations, and to various formats, was rearranged by computer into a standardized magnetic tape format.

Finally, for efficiency of searching, the magnetic tape records were sequenced by sex and a phonetically coded form of the surname subdivided by alphabetic spelling of the surname and the first given name. Such sequenced files were consolidated to cover all provinces, and sometimes to cover a span of registration years. The purpose of the surname coding is to get around the effects of common spelling variations which could otherwise interfere with finding a record; the consolidation serves to reduce the number of files to be searched. Where a special search is required for records which might be in unexpected locations due to discrepancies in the surnames or birth dates, the files are sometimes rearranged in an alternative sequence which depends on a different combination of identifying particulars. The manner in which these strategies have been applied will be described in greater detail in the sections which follow.

Mortality Data Base

As in most developed countries, information from the provincial death registrations in Canada is routinely keyed to cards or tape for two purposes: to serve as a source of annual mortality statistics, and to provide alphabetic indexes with which to access the bound volumes of death registration forms and their microfilm images. These registration forms are maintained in numeric sequence by a death registration number which is assigned serially when the documents reach the provincial registry offices. The causes of all deaths, names of the decedents, and other pertinent information are converted into machine-readable form for statistical and administrative purposes. These existing, routinely produced, computer records of all deaths are used to produce a centralized Mortality Data Base that can be employed for searching out the dates and causes of death of named individuals on the large scale required for epidemiological follow-up.

Although the amount of personal identifying information entered in machine-readable form for indexing purposes was less for the earlier death records than at present, the

format of the computer records used for the Mortality Data Base is standard and allows for the inclusion of almost all identifiers currently being keyed. Details of the numbers of characters of information provided for in the machine-readable record, including the coded underlying cause of death, are given in Table 1. The degree of redundancy in this identifying information constitutes the only safeguard against possible mistaken identity, and the likelihood of being misled by errors and discrepancies in the manner in which the various identifying items are sometimes reported on separate occasions.

The files of the Mortality Data Base, for most purposes, are sequenced by a phonetic New York State Identification and Intelligence System (NYSIIS)¹ code of the surname of the deceased, and sex. The Mortality Data Base includes deaths back to 1950. The accumulated total in death registrations represented in the Mortality Data Base (1950-1977), is approximately 4,000,000.

Each entry on the magnetic tape files is supported by the relevant microfilm copy of the completed death registration form. Where the summary information contained in a magnetic tape death record is insufficient for some special purpose, it is possible to locate the microfilm image of the original death registration for visual inspection. These images are arranged in order of the death registration number, under year and province of occurrence, so that manual retrieval is reasonably rapid when the registration number is known. Thus the microfilm image provides a complete back-up for the Mortality Data Base file.

For recent years, the level of reporting of the various kinds of personal identifying information on the death registration forms is high. In practice, it is often deficiencies of the identifiers contained in the "exposure" records used to initiate the searches of the Mortality Data Base that limit the success of the searching.

The causes of death given in the death registration forms are known to under-represent and over-represent true cancer mortality by something like 11-12 per cent.² Since only the underlying cause of death is keyed to machine-readable form, under-ascertainment of cancer as a cause of death, when the Mortality Data Base tapes are used, will be somewhat greater. However, additional information concerning members of a study population who are known to have died may often be obtained through a computer search of the National Cancer Incidence Reporting System tapes. Such comparisons are known from manual studies to identify other deceased individuals as having had cancer. In addition, they may be expected to yield in some instances alternative and perhaps more precise cancer diagnoses, and serve as a source of information on living members of a study group who have cancer. Both cancer mortality and cancer incidence have already been studied in such a manner.³

The National Cancer Incidence Reporting System

The records of the centralized National Cancer Incidence Reporting System are derived from those of the provincial cancer registers, starting with the 1969 registrations. These records contain less personal identifying information than the death registrations for a number of possible reasons:

TABLE 1—Mortality Data Base Records—Provision for Storing Information on Magnetic Tape*

Information		Number of Characters
Personal Identification		
Deceased's:	surname	10
	given names	9 + 7
	birth date (year, month, day)	7
	birth province or country (coded)	2
	sex, marital status, racial origin	4
Father's:	surname (maiden surname of deceased woman)	10
	initials	2
	birth province or country	2
Mother's:	maiden surname	10
	initials	2
Spouse's:	birth province or country	2
	birth surname (maiden surname where spouse is wife)	10
	initials	2
Death Details		
Death:	cause (ICDA code)	4
	date (year, month, day)	6
	province	2
	place of occurrence	10
	age	4
	home or institution	1
Accident:	nature of injury (ICDA code)	4
	place of occurrence	1
	county	2
	locality	3
Miscellaneous:	pregnancy death	1
	operation	1
	autopsy and findings	2
	attendant	1
Other personal		
Residence:	province or country	2
	county or census division	2
	locality	3
Occupation:	kind of work done (coded)*	4
	kind of industry (coded)*	4
Housekeeping Information		
	control codes	2
	death registration number	7
	Social Insurance Number (in case available)*	9
	surname codes (for 5 surnames, including alternate)	30
	date of update	2

*Not all of the items for which spaces are provided are keyed to tape. For example, occupation is rarely coded and Social Insurance Numbers are rarely keypunched. However, availability of the various personal identifiers in records of the Mortality Data Base has tended to improve over the years. For 1973 deaths, the full birth date is given in 96 per cent of the records, and the province or country of birth in 98 per cent; parental initials and the provinces or countries of birth of the two parents are given in 82 per cent of the records, and mother's maiden surname is given in 68 per cent.

in most provinces cancer registrations are not legal documents; the source documents from which cancer cases are ascertained often contain only limited personal identifying information; and epidemiological use of the centralized files, for which more precise personal identification would be desirable, is quite recent. Nevertheless, provision has been made in the magnetic tape record format, used in the centralized National Cancer Incidence Reporting System, for adequate personal identifiers to ensure reasonably efficient searching (Table 2).

Much of this identifying information is actually supplied by a majority of the provinces so that perhaps 70 per cent of the recent records may be viewed as adequately identifying the affected persons for the purposes of a computerized search. With additional manual-clerical work, the remainder of the records could also be of use.

The number of records of cancer registration in the National Cancer Incidence Reporting System (1969-76) is approximately 330,000. This may include a few cases that have been registered in more than one province, but corrections are possible through computerized linkage, within the file, of records containing similar names, birth date, and such. For the purpose of epidemiological follow-up of named individuals, however, the presence of an admixture of duplicate registrations creates no special problem, although it is of some concern when the files are used for studies of cancer incidence in the population at large. When Ontario becomes a part of the system there will be almost 70,000 registrations annually.***

***Malhotra A: Statistics Canada, personal communication. Ontario is the only Canadian province not now included in the system.

TABLE 2—National Cancer Incidence Reporting System Records—Provision for Storing Information on Magnetic Tape*

Information	Number of Characters
Personal Identification	
Case's: surname	21
given names	16
maiden surname or other previous name	10
birth date (year, month, day)	8
birth place (province and census division or continent and country)	4
sex	1
marital status (females)	1
Spouse's: given name	9
Cancer Details	
cause (ICD-0 Topography Code)	4
(ICD-0 Morphology Code)	5
method of diagnosis	1
date of diagnosis	8
age at diagnosis	3
reporting province	2
Death Details	
dead-or-alive status	1
date of death	8
death registration (year, province, registration number)	10
Other Personal	
Residence: at time of diagnosis (city, town, village)	10
postal code	6
Standard Geographical Code (province)	2
(county or census division)	2
(census subdivision)	3
Housekeeping Information	
register case number	7
previous register case number	7
Social Insurance Number	9
health insurance number	12
source of registration (vital statistics source)	1
primary site number (1st, 2nd, etc.)	1

*The Social Insurance Number is rarely available, but for 1975 the medical care insurance number was present on 61 per cent of the records. For the same year, the full birth date was present on 67 per cent, and the province or foreign country of birth on 31 per cent.

There is less experience to date in the use of the centralized National Cancer Incidence Reporting System for epidemiological follow-up than in the corresponding use of the Mortality Data Base. The National Cancer Incidence Reporting System could be developed into an even more useful searching facility than it is at the present, through a modest increase in the amount of personal identifying information entered into some of the provincial records.

Linkage Techniques

Two steps are involved in establishing that a particular named individual is represented by a record in one of these files, i.e., of "linking" a record which is used to initiate a file search, with a record contained in the file. The first or "searching" step is to bring these two records together for comparison, and the second, or "linking" step is to calculate, from agreements and disagreements of the various identifiers, the odds in favor of or against the possibility that the same person is represented on both records.

In the searching step, the chief problems arise where the sequencing information (e.g., the surname and the birth date) has been stated differently on the two potentially linkable records (e.g., the surname may be spelled differently or some part of the birth date may be reported differently on one of the records). The cost of the searching step will be least where enough sequencing information is used to ensure that the file is finely divided, so that detailed comparisons with a record that initiates a search are thereby limited to just a few records on the mortality or cancer file.

In practice, a phonetically coded form of the surname (its "NYSIIS" code), plus sex code, are found to subdivide the file finely enough for economical searching. The phonetic coding circumvents many of the difficulties created by the more common spelling variations. Where discrepancies in the surname code might have prevented the bringing together of potentially linkable pairs of records for detailed comparison, a second search may sometimes be carried out in an alternative sequence which ignores the surname code (e.g.,

in a version of the file sequenced in order of the birth date and given name). However, such additional searches add to the cost and, normally, a single search under the main sequence is considered to be adequate.

In the linking step, the chief problems arise where some of the various identifying particulars agree on the records brought together for detailed comparison, and some disagree. In similar situations, a filing clerk would use subjective judgment to decide whether or not the two records relate to the same person. Until recently it was generally thought that computers would be less effective than a human in carrying out this kind of operation but, in fact, the rules on which the human judgment is based are not too difficult to formalize for the computer. When the computer is given information on the discriminating powers of various agreements and disagreements, it is capable of calculating the odds in favor of, or against, a genuine linkage.

The procedure is simpler than one might suppose. For example, when comparing two records both of which have to do with a Mr. Brown, the odds against them relating to the same person, in any large file, may be quite high if this is all one knows. However, if both records show the first initial to be the letter J, this agreement will modestly increase the odds in favor of a correct match: e.g., where J has a frequency of 1 in 32 as a first initial among males, the agreement will increase the odds that the same person is involved, by 32-fold. The odds would be vastly greater if both records related to a Mr. Z. Q. Brown, because these initials are rare and, as a result, carry greater discriminating power: e.g., where Z and Q each have frequencies around 1 in 1,000, the combined agreements will increase the odds that the same person is involved, by about a million-fold. The full given names would carry even greater discriminating powers, determined by their frequencies. Similarly, agreement with respect to the birth month and birthday will increase the odds in favor of a correct match by 365-fold, the number of days in a year. Likewise, agreement with respect to the birthplace will carry greater discriminating power where this is a small community than where it is an urban center, and so on. Look-up tables of discriminating powers are supplied to the computer, based largely on the frequencies of the various values of a given identifying particular as derived from the files themselves. The approach is applicable equally to the negative discriminating powers associated with disagreements of identifying items, some of which carry little discriminating power (as for example a change of address) and others of which are highly discriminating (a disagreement with respect to sex is quite unlikely to be wrongly reported or keyed to tape). To derive the negative weights, some preliminary linkages must be carried out and checked clerically.

Such methods have been developed and refined over a period of two decades, and there is now an extensive literature describing them and the degree of success which they have achieved. The first published description was in 1959.⁴ A convenient bibliography exists of virtually all computer-assisted record linkage studies up to 1970, including those dealing with the techniques and with the various applications,⁵ and two books on the subject are worthy of special attention.^{6,7} More recent studies illustrate the extent to

which the linkage methods may be refined to ensure maximum efficiency in the use of the available identifying particulars,⁸ and show that the accuracy of the computer tends to be greater than that of human clerks carrying out the identical record linkage operation on a routine basis.⁹

Specific information on the accuracy of the computer searching, in terms of the numbers of "false positives" and "false negatives", would be relevant but depends heavily on the amount of personal identifying information present on the records used to initiate the searches. The same could be said for any corresponding manual searches by human clerks. However, statistical data are rarely sought in this context. Many of the files by which study populations are identified contain an admixture of records having only the surnames, initials, and incomplete birth dates. Computer searches based on these will yield possible links with death registrations, but the calculated degrees of certainty will be exceedingly low. Since clerical searches based on the same records will be equally unrewarding, the simplest procedure is to exclude them from the study until more identifying particulars can be obtained. At the other extreme, where there is ample identifying information, the calculated degree of assurance of a correct computer link will be high, and there will be little likelihood that a human would arrive at a different conclusion. It is in the intermediate situations, however, that an investigator has to exercise one of the various possible options. He may, for example, raise the "threshold" for acceptance, thereby reducing the false positives but increasing the false negatives; lower the "threshold", with the opposite consequences; or subject the computer linkages of borderline certainty to visual scrutiny, in the hope that a human may perhaps see some way to resolve the ambiguity by going back to the source documents for additional information not in machine-readable form. In the one large comparison of computer versus human linkages mentioned earlier,⁹ the computer was nearly 30-fold better at excluding false and doubtful linkages than were its human counterparts (i.e., only 36 false positives vs 1,010 for the human clerks, out of a total of 44,315 total acceptable links). The human clerks, however, were better by 1.7-fold at avoiding "missed" linkages (i.e., 737 false negatives for the computer versus 449 for the human clerks). In this exercise the computer linkages may be judged 99.9 per cent "pure" and 98.3 per cent "complete". These figures relate to the linkage of records pertaining to children. When searching for records of the deaths of adults, additional inaccuracies may result from the longer time spans involved. However, these should affect the accuracies of both the computer and the manual searches in a similar manner.

A generalized record linkage system has been developed at Statistics Canada in collaboration with the Epidemiology Unit of the National Cancer Institute of Canada, and this facility is being used to carry out a number of medical follow-up studies.‡

‡Howe GR and Lindsay J: A generalized iterative record linkage computer system for use in medical follow-up studies. Manuscript in preparation.

TABLE 3—A List of Items to be Included in Any “Employee Health Identity Questionnaire”*

-
1. Surname
 2. Alternate surname (if any)
 3. First given Name
 4. Second and other given names
 5. Usual name (or nickname)
 6. Sex
 7. Birth date (year, month, day)
 8. Birth province or country
 9. Birth city or place
 10. Father's surname
 11. Father's first given name
 12. Father's second given name
 13. Father's birth province or country
 14. Mother's maiden surname
 15. Mother's first given name
 16. Mother's second given name
 17. Mother's birth province or country
 18. Marital status
 19. Spouse's birth surname
 20. Spouse's first given name
 21. Spouse's second given name
 22. Spouse's birth province or country
 23. Social Insurance number
 24. Health insurance number
 25. Pension plan number
 26. Signature
 27. Date (year, month, day)
 28. Complete address including postal code
-

*This form is designed to be filled out when the employee is hired. Other information relating to work histories and updates to the questionnaire should be added later. Termination dates should be added. A “last known alive date” is of value in order to reduce the amount of searching required in the death file. The complete address should be updated wherever possible.

The speed of the computer searching is exceedingly high, and the cost is small, provided that the magnetic tape files are all in order and the sequencing information is such as to divide the file finely. One of the tests carried out under these conditions has indicated that 14,000 searches could be made per minute of central processor time, representing a cost of about one-half a cent each.⁹ In actual practice, most of the expenditures go on preparation of the files, including the formatting and sequencing of the magnetic tape records, and on planning and testing the linkage procedures, which may differ from job to job. Thus unit costs in the vicinity of a dollar or two per search may be encountered, depending on the study. In general, the larger the number of searches to be carried out, the lower the cost per search. Where the records which initiate the search are kept routinely in a form suitable for the purpose, considerable savings are possible.

Studies Carried Out and in Progress

Availability of the National Cancer Incidence Reporting System as a follow-up facility is still too recent to have provided any considerable experience in its use, but the Mortality Data Base has been actively employed for some time, and the demands for its use are increasing. At a recent workshop sponsored by the National Cancer Institute of Canada,¹⁰ two studies were reported in which there had been extensive

searching for death records in the Mortality Data Base files. In one of them, an hypothesis that isoniazid is not carcinogenic was tested by linking the files of 64,000 people treated for tuberculosis since 1952 with the mortality and cancer incidence data.³ In another study, the scope of the system for environmental monitoring was shown by extracting and linking the occupational records of a 10 per cent sample of the Canadian work force, a test which involved over 700,000 records searched for in the Mortality Data Base. Other studies not yet completed will involve a similar kind of follow-up of approximately 100,000 tuberculosis patients (for carcinogenic effects of multiple fluoroscopies), and of some tens of thousands of uranium miners, hard rock miners, radiation workers, and nickel workers, including the nickel sinter workers.

Design of Employee Name Rosters

When a study of the delayed effects of a working environment is initiated, there are often major problems in establishing who had worked in the past for the industry involved, and in gathering identifying information with sufficient specificity to reduce to a low level the likelihood of mistaken identity when searching large files.

Such difficulties could be greatly reduced in the future if adequate personal identifying information is obtained routinely, in a compact form, from each employee already currently employed and from all new employees at the time of hiring. In the past, application forms for employment often contained most of the required identifiers, but this has ceased to be the case in Canada, and even age and sex are no longer requested on many application forms because such information could be regarded as forming the basis for possible discriminatory hiring practices.

A list of items to be included in any “Employee Health Identity Questionnaire” which would facilitate searching for death registrations using the Mortality Data Base files and the National Cancer Incidence Reporting System is given in Table 3. The individual information items overlap those contained in the completed death registration forms, so that a very high level of assurance of a correct match with a death registration will usually be possible even where some of the information is reported differently on the two sorts of records. Where the abbreviated versions of this information, contained in the magnetic tape death and employee records, leave doubt concerning the correctness of a match, the full versions of the information may be compared visually.

The use of the Social Insurance Numbers in compiling an employee name roster is particularly desirable, both because of the links these create with the payroll records of a company, and because it simplifies the removal of duplicate entries in such a roster. These numbers also facilitate the compilation of work histories involving more than one employer.

In an ongoing study of the employees of a company, a cumulative name roster would be maintained routinely. Much would depend on making such a routine simple and inexpensive, and on the precautions to ensure that the information is retained permanently in a compact, retrievable form.

Privacy and the Right to Know the Risks

Two fundamental human rights must be taken into consideration in connection with most kinds of epidemiological follow-up, regardless of how they are carried out. On the one hand, there is the right of the individual to protection against harm or embarrassment which could arise through the inadvertent or deliberate release of sensitive personal information. On the other, there is his right to know the risks he runs, in the work place and elsewhere, and to be assured that standards of protection against harmful agents take into account all obtainable information on any delayed effects such as cancer. The objective in follow-up studies of the kinds discussed should be to ensure the second of these two rights without creating a threat to the first.

In the case of the two data bases described, no new records have been created, and the information remains with the current custodians. The same personal information as rearranged in the Mortality Data Base file and the National Cancer Incidence Reporting System stays wholly in the hands of those agencies that are already using it. For example, when an investigator provides the work histories of a study population to be linked with the records of subsequent deaths and cancers, the results of the searches will be released to the investigator only as bulk statistics. Thus, personal information about a particular individual is not revealed. However, the bulk statistics may be broken down to permit appropriate comparisons and analyses.

Confidentiality of the personal information contained in the vital and health records from which the data bases are derived is protected under the federal Statistics Act and the corresponding provincial acts.†† Those working with the records federally have to swear an oath of secrecy under the federal Statistics Act, and severe penalties are provided by the Act for any breaches of confidentiality.

Public uneasiness about epidemiological follow-up in general, and government participation in particular, arises for a number of reasons. Most important is a widespread misimpression that there must be some other adequate way to detect and measure risks without having to follow the exposed people to see what eventually happens to them. There is also a widespread failure to understand that, if one is concerned with detecting small increases in the frequency of a common disease like cancer, the number of persons to be followed up must be large. Finally, there is a feeling on the part of many that the government should not be involved in such matters. Nevertheless, when a special hazard is newly suspected by the public it is usually government agencies that are called upon to ensure that appropriate investigations be carried out. The alternative is by no means clear. At one time, when epidemiologists in the United States were pressing for a national death index, it was suggested by members of the staff of the National Center for Health Statistics that such work might best be handled by a private agency¹¹. As would be expected, however, this option was not pursued, presumably because state registrars would have been far less

willing to release confidential vital statistics records to a commercial concern than to a federal agency.

Support for the idea of facilitating epidemiological follow-up for the detection and measurement of delayed risks, has come, in part, from technical people concerned with making the follow-up process efficient; from the unions whose membership may be at risk in the workplace; and from the occasional non-technical author who seeks to warn the public about the delays in identifying industrial and other risks. Examples include:

- Reports to the Medical Research Council of Canada,¹² the Ontario Council of Health,¹³ and the Government of Ontario.¹⁴

- A resolution adopted unanimously by the 1975 National (Canadian) Policy Conference of the United Steelworkers of America, calling upon federal and provincial governments to

"enact legislation providing for . . . A complete work history of every industrial worker which, coupled with a complete medical profile, will enable us to quickly identify and eliminate potentially hazardous working conditions";¹⁵ and

- Various recent books, some written in a popular and provocative style which tend to put the blame for failure to monitor industry (including epidemiological follow-up of workers) partly on industry itself and partly on government.¹⁶⁻¹⁸

There are indications that the above point of view is now more widely understood than it was in the past. One such indication is the decision in the United States to create their National Death Index, beginning with the deaths in 1979.¹⁹ Another is a statement from an American source to the effect that "pressure was growing in Congress to remove some of the obstacles that the Privacy Act of 1974 had placed in the way of the solution of environmental health problems by epidemiological means."¹⁰

For the future, the efficiency with which the National Cancer Incidence Reporting System files may be searched will depend on the consistency with which a certain minimum amount of personal identifying information is supplied through the provincial cancer registers. Whether or not the provincial health insurance records come to be used in a similar fashion will depend largely on the manner in which people are identified on these records. However, at the provincial level, Saskatchewan's health insurance records have already been used to study the risks of various occupational diseases among such groups as farmers and grain buyers.²⁰

Legislation governing the use and availability of data, information, and records pertaining to individuals has profound implications for the feasibility of epidemiological studies and surveillance systems for monitoring delayed health effects. In planning legislation to further protect the privacy of individuals, there is a need to ensure that this aim will be achieved by efficient and economical means, without preventing the detection of delayed environmental and occupational risks.

REFERENCES

1. Lynch BT and Arends WL: Selection of a Surname Coding Pro-

††Rowebottom LE: Statistics Canada: The legal basis of health record linkage. A paper presented at a workshop on Computerized Record Linkage in Cancer Epidemiology, Ottawa, August 1979.

- cedure for the SRS Record Linkage System. Washington, DC: US Department of Agriculture, 1977.
2. Barclay THC and Phillips AJ: The accuracy of cancer diagnoses on death certificates. *Cancer* 1962; 15:5-9.
 3. Howe GR, Lindsay J, Coppock E, *et al*: Isoniazid exposure in relation to cancer incidence and mortality in a cohort of tuberculosis patients. *Int J Epidemiol* 1979; 8:305-312.
 4. Newcombe HB, Kennedy JM, Axford SJ, *et al*: Automatic linkage of vital records. *Science* 1959; 130:954-959.
 5. Wagner G and Newcombe HB: Record linkage—Its methodology and application in data processing. *Methods Inf Med* 1970; 9:121-138.
 6. Acheson ED: *Medical Record Linkage*. London: Oxford University Press, 1967.
 7. Acheson ED (ed): *Record Linkage in Medicine*. Edinburgh: E and S Livingstone, 1968.
 8. Smith ME: Methods for computer linkage of hospital admission-separation records into cumulative health histories. *Methods Inf Med* 1975; 14:118-125.
 9. Smith ME and Newcombe HB: Accuracies of computer versus manual linkages of routine health records. *Methods Inf Med* 1979; 18:89-97.
 10. Privacy, epidemiology, and record linkage (editorial). *Br Med J* 1979; 2:1018.
 11. Chase HC: Report on a national death index—pros and cons. *Am J Public Health* 1972; 62:719-723.
 12. *Health Research Uses of Record Linkage*. MRCC Report No. 3, Ottawa: Medical Research Council of Canada, 1968.
 13. *Report on the Activities of the Ontario Council of Health, June 1966 to December 1969*. Toronto: Ontario Department of Health, 1970. (Note especially Recommendations 10, 23, 24 and 25, pp. 104-108).
 14. *A Study of the Implications of Using a Personal Identifier in the Ministry of Health*. Project No. 446 of the Management Consulting Division. Toronto: Government of Ontario, January 22, 1974.
 15. Valentine K: Worker health and record linkage. In: *Hazards at Work—Proceedings*, Fairclough C (ed) Toronto: Corpus Information Services Limited, 1978, pp. 44-48.
 16. Agran L: *The Cancer Connection—And What We Can Do About It*. New York: St. Martin's Press, 1977.
 17. Epstein SS: *The Politics of Cancer*. Garden City: Anchor Press/Doubleday, 1979.
 18. Tataryn L: *Dying for a Living*. Ottawa: Deneau and Greenberg Publishers Ltd., 1979.
 19. *Human Health and the Environment—Some Research Needs*. Report of the Second Task Force for Research Planning in Environmental Health Science. DHEW Pub. No. NIH77-1277, Washington, DC U.S. Govt Printing Office, 1976.
 20. Dennis CAR, Drope R and Matz D: *Medicare Data: Its Use in Defining the Effects of the Environment on Health*. NRCC Report No. 15387 of the Environmental Secretariat, Ottawa: National Research Council of Canada, 1977.

ACKNOWLEDGMENTS

The authors thank Anna Malhotra and Christine Poliquin for their assistance in supplying detailed information for the preparation of the tables regarding the National Cancer Incidence Reporting System.

The opinions expressed are those of the authors and do not necessarily reflect those of Statistics Canada.

1981 Symposium on Environmental Epidemiology Scheduled

The 2nd annual Symposium for the Center for Environmental Epidemiology, University of Pittsburgh, will be April 27-29, 1981 at the David Lawrence Auditorium at the University. The theme of the seminar is "Epidemiologic Basis for Current Environmental Standards." and will be concerned with the epidemiologic evidence on which current or proposed environmental standards are based.

Invited speakers include nationally recognized scientists, on such topics as: general methodological issues including acceptability of risk; ionizing radiation; various chemical agents including vinyl chloride, coke oven emissions, benzene, asbestos, and water and air pollutants. Available evidence related to standards both for the general public and for workers will be discussed.

Members of the academic community, government, and industry are invited to attend. Registration fee is \$40 and the deadline for pre-registration is March 15, 1981. Attendance will be limited. For further information please contact:

Center for Environmental Epidemiology
 Graduate School of Public Health
 University of Pittsburgh
 Pittsburgh, PA 15261
 Phone: 412/624-1559 or 624-1560