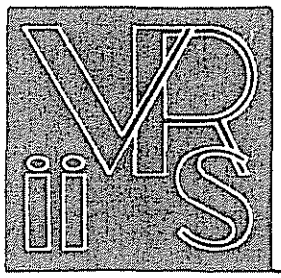

# Technical Papers

Number 50
June 1992

## AUTOMATION OF MORTALITY DATA CODING AND PROCESSING IN THE UNITED STATES OF AMERICA

International Institute for Vital Registration and Statistics
9650 Rockville Pike
Bethesda, Maryland 20814-3998
U S A

TABLE OF CONTENTS



## FOREWORD

One of the most important bodies of data which a well functioning civil registration and vital statistics system can provide to users are mortality data on causes of death. However, the classification of the cause of death based on information provided by the attending physician on the medical certification of causes of death is one of the more difficult tasks facing national vital statistics offices. As the author notes, the International Classification of Diseases (ICD) published by the World Health Organization provides the statistical classification and the rules for classifying causes of death. However, these rules are fairly complex and require well-trained high level coders to achieve desired levels of accuracy, comparability and consistency. The automation of cause of death coding utilized the ACME program to select and code the underlying cause of death, and MICAR to input to the computer the diagnostic information reported in the natural language. By using the ACME and MICAR programs, the underlying cause of death is computer coded without any clerical intervention.

# AUTOMATION OF MORTALITY DATA CODING AND PROCESSING IN THE UNITED STATES OF AMERICA

Robert A. Israel*

## Background

Since the beginning of the 20th century when a national system for collecting and disseminating mortality data based on State registration systems began in the United States of America, the coding of demographic and medical data has been a necessary but difficult task.

One of the most complicated kinds of statistical data classification is the assignment and coding of a single underlying cause of death for each death. This approach involves the determination of "the single disease or injury which initiated the train of morbid events leading directly to death or the circumstances of the accident or violence which produced the fatal injury."[1]

The World Health Organization (WHO) has provided the statistical classification structure and associated rules for determining the underlying cause of death as part of the approximately decennial revisions of the International Classification of Diseases (ICD).[1] While this classification sys- is comprehensive and invaluable from many points of view, it leaves a number of problems for the user. Some of these difficulties are: 1) the long period of training required to develop competent nosologists schooled in the classification and coding of mortality medical data and the related problem of staff retention in an area of work that requires exactness and controlled productivity; 2) the difficulty in achieving desired levels of data comparability and accuracy brought about by complicated rules and classification structures, inexactness in reporting of some medical conditions, and human limitations in the broad range of knowledge required by nosologists; and 3) the selection of only a single underlying cause of death which results in the loss of the remaining information reported on the medical certification portion of the death certificate. Some of the lost information may well be relevant to public health concerns. Losses may have been minimal in the earlier part of this century when acute infectious diseases exacted a heavy toll and often were the only noted cause of death on death certificates. Nevertheless, at no time during this period did the loss of information resulting from the coding of only the underlying cause go unrecognized. However, the expense of coding and tabulating more than one cause of death and the complexities of analyzing and publishing the data discouraged frequent compilations of all of the reported medical data on death certificates during most of the first half of the 1900s.

However, there was a growing realization in a number of countries that the epidemiologic transition which accompanies socioeconomic modernization had resulted in an increasing proportion of deaths from chronic diseases. For example, where less than 20 percent of all deaths in the United States at the beginning of this century were attributed to heart disease, cancer, and stroke, now approximately 70 percent of deaths fall into these same three categories.[2] Deaths due to chronic diseases are often not well characterized by a single cause; rather, they are more likely to result from a number of coexisting conditions among which there may be no direct etiologic chain to facilitate the identification of a single underlying cause.

In order to meet expanding needs for greater utilization of medical information from death certificates and to address problems of coding complexities and loss of additional information under the single underlying cause of death approach, the National Center for Health Statistics (NCHS) undertook in the late sixties and early seventies the development of an automated computer system to produce mortality statistics. This system was designed to provide statistical data on all of the medical information reported on death certificates as well as generate underlying causes of death through computer methodology. The international coding rules are applied by the computer system to each medical term given by certifiers in much the same manner as done in a manual coding operation. The system[3] was given the acronym "ACME", a short name for "Automated Classification of Medical Entities."

* Associate Director for International Statistics, National Center for Health Statistics, USA

The utilization of other medical data on death certificates, in addition to the underlying cause, has been undertaken in various ways in several countries. For example, in the United States the coding of more than just a single underlying cause was carried out using deaths occurring in 1917, 1925, 1936, 1940, and 1955, prior to the development of the ACME system. Renewed interest in multiple causes of death at the international level resulted in two meetings--one in 1967 in London and one in 1969 in Geneva--bringing together a number of interested nations for the purpose of exploring uses of multiple cause analysis and methodology, with emphasis on minimum standards for international comparability.[4] At both of these meetings experiences were exchanged and various approaches to coding multiple causes were discussed. In the end, it was agreed that those countries interested in multiple cause of death analysis should experiment with methodology and keep WHO and each other informed of their activities. With that understanding NCHS proceeded to incorporate its own set of multiple cause coding rules into the evolving ACME system.

## Development of an Automated System

The system requires the entry into the computer of all diseases, conditions, accidents, and injuries given in the medical certifier's statement. The information is recorded onto tape utilizing codes based on the current version of the ICD. Special instructions are used in applying the ICD codes to the reported diseases, conditions, and injuries. Once the medical information is placed on magnetic tape in coded form and in the same sequence as it appeared on the death certificate, the entries are matched against a series of decision tables which analyze the ICD codes and their relationships to each other in terms of each of the international coding rules. Defining the specific content of each decision table presented a difficult systems development problem, requiring a deliberate decision on the relationship of each detailed code to every other ICD code relative to the purpose of each of the international rules. Traditionally, such decisions have been made by individual coders who have had only some guidelines on causal relationships of diseases and on priorities of some conditions over others. In order to automate the process a much more thorough and systematic approach needed to be taken.

The computer program is written in PL1 Computer Programming Language. It provides for storage of the decision tables in core memory, minimizing reference time requirements, which, in turn, increases output speed.

The contents of the decision tables are updated by introducing individual changes when approved by the staff. The tables are now routinely maintained on disc storage and a special program provides for the introduction of changes, coupled with a printout of the updated tables with a special notation showing the changes for visual verification.

In addition, a quality control process is maintained through manual coding of a sample of records. The disagreements between the manual and automated coding are fully reviewed and, when appropriate, adjustments to the decision tables are incorporated into the system. Furthermore, if a particular decision is changed and is considered to be sufficiently significant to warrant retroactive correction of records previously processed, the file can simply be reprocessed through the computer system using the corrected decision tables.

## System Overview

Thus, simply stated, the ACME system utilizes information based not only on the ICD codes for each reported condition on the death certificate but their actual location on the certificate as well (e.g., first condition on line Ia; second condition on line Ic; first condition in Part II; etc.). These individual codes and their respective location codes are manually assigned by a coder who requires less training than a fully experienced underlying cause of death coder. To these codes the computer program then applies, in sequence, each international selection rule, resulting in a code for a "temporary underlying cause." This "temporary underlying cause" code is then subjected to each international modification rule in sequence, finally arriving at an assignment of a single underlying cause of death code. In this process, each selection and modification rule is represented in the computer system by a decision table which expresses the rule in terms of acceptable or unacceptable relationships among specified ICD Codes. Thus, the process is carried out in a manner quite similar to the way a trained human coder would apply the rules to arrive at an underlying cause of death code.

However, manual application of the international coding rules is subject to varying degrees of inconsistency. These inconsistencies include not only differences in interpretation and application of rules and guidelines from coder to coder, but also changes that occur over time. The computer

system provides an avenue of absolute consistency and a means of isolating troublesome certifications for further study. An important by-product is the full documentation of the assumptions and decisions going into cause of death classification. Heretofore, documentation has been given in terms of general guidelines with limited minimum reference as to how the guidelines are to be applied in specific situations. Automating this element of the classification process has given users detailed insight into the data and permits more intelligent analyses of the end product.

The same data that serve as input to the ACME system for the assignment of the underlying cause of death, i.e., the ICD codes for each diagnostic term appearing on the medical certification of death, also serve as the data input to multiple cause of death analysis which takes cognizance of all conditions reported at death on the certificate. The availability of multiple cause of death data is a significant by-product of the computer selection of underlying cause of death by the ACME system.

Although the underlying cause of death is conceptually easy to understand and is a well-accepted measure of mortality, often it does not convey the complexity of the reported medical conditions at the time of death. Because many deaths are the result of more than one disease entity, a single underlying cause excludes useful information on intervening contributory, and even concurrent, conditions. This is a particular problem when a death is reported as being caused by two or more unrelated diseases. In addition, for persons who die of injuries, only the external cause is traditionally tabulated as an underlying cause of death. For example, if a person dies from a broken neck as a result of a fall from a ladder, the underlying cause is represented as the fall rather than as the nature of injury (damage to spinal cord).

The known limitations of underlying cause data have led some to suggest complete abandonment of the underlying cause for statistical tabulations. Yet, underlying cause data are still important and appropriate for many mortality analyses including trend analysis and international comparisons. Because of this, multiple cause data should be viewed not as a replacement for, but as an adjunct to, basic underlying cause statistics.

Routine production of multiple cause of death data in the United States has been instituted; a set of computer tapes that contain multiple cause of death data for each death in the United States for each of the years 1968-1987 are available and these data are now being produced on an annual basis and made widely available. This routine availability is the result of the development and use of the ACME system to produce the "official" traditional underlying cause of death statistics.

It is important to note that with the present version of medical certificate of cause of death, the design and structure of the ICD itself, and the quality and quantity of the basic diagnostic information reported on certificates, there are a number of limitations to multiple cause data. Despite these limitations, however, multiple cause of death data allow researchers to maximize the use of available diagnostic information on the death certificate. In particular, multiple cause data can give a more complete description of trends for chronic diseases such as diabetes, chronic respiratory disease, hypertension, and kidney-related diseases, because the underlying cause trends may be masked by the changing priority given to these causes as underlying rather than as nonunderlying causes of death.

The availability of national multiple cause data should generate an increasing number of applications for those data. Once the usefulness of the data in epidemiologic and health research has been established, it will be easier to implement programs aimed at educating medical certifiers on the importance of accurate and complete reporting of all conditions, underlying and nonunderlying, that were believed to contribute to each death. If multiple cause data are to realize their full potential, then significant interaction between producers and consumers of the data must occur. In the long run, continuing assessment, evaluation, and research should help to improve the system further.

## System Enhancements

NCHS is currently developing a major addition to the ACME software. This addition is known as the Mortality Medical Indexing, Classification, and Retrieval System (MICAR). MICAR is being developed as a "preprocessor" or "front end" to the ACME system and will automate the assignment of ICD codes to each reported condition on the death certificate by accepting natural language (in this case, the diagnostic terms in English) typed into the computer system. MICAR will utilize less skilled data entry operators instead of more highly trained nosologists and will reduce training

requirements. MICAR will also store and retain for retrieval both the natural language text and the conventional ICD codes that are required for underlying cause processing.

In addition to speeding the coding of death certificates, reducing the training required of staff to use the system, and improving the consistency and accuracy of ICD coding, there are several other anticipated by-products of MICAR:

- MICAR will provide, in standardized form, computer retrievable access to every term reported on death certificates. These terms will be an invaluable source of data for researchers who wish to monitor a disease which is an inclusion term within one ICD category without its own unique ICD number.

- MICAR will simplify conversion from one revision of ICD to the next. Normally ICD conversion requires considerable retraining of coders. Under MICAR, the majority of staff will not have to be trained in code number and rule changes since the data entry will be essentially independent of the structure and content of ICD. MICAR will initially assign to each term a fixed reference number. Of course, with each revision of ICD, the MICAR reference numbers will have to be linked via software to the appropriate ICD codes.

- MICAR plus ACME will ease comparability or bridge studies between ICD revisions by allowing the entry of records once via natural language and then processing them to obtain the underlying causes of death using both the old and new revisions of the ICD.

## Summary

The development in the United States of an automated system for coding mortality data was undertaken with two major objectives in mind: 1) to introduce consistant and rapid assignment of underlying cause of death coding with reduced needs for manpower training; and 2) to allow better utilization of medical information on death certificates for multiple cause of death analyses. The ACME system meets both of these objectives; NCHS produces all of its underlying caose of death statistics for the United States based on this system, and multiple cause of death data are routinely available for additional epidemiologic study beyond the traditional methods of vital statistics analyses.

Enhancements of the automated system, primarily through the software known as MICAR, reduce even further the levels of training necessary for persons doing the basic data entry. MICAR additionally will ease transitions between ICD revisions by reducing the need for coder reorientation and by permitting rapid calculation of comparability ratios when new revisions are introduced.

## References

1. World Health Organization. International classification of diseases. Manual of the international statistical classification of diseases, injuries, and causes of death. Vol 1. 9th revision. Geneva: World Health Organization, 1977.

2. Israel, R.A., Rosenberg, H.M., Curtin, L.R. Analytic potential for multiple cause of death data. American journal of epidemiology, 124(2): 161-179 (1986).

3. Israel, R.A., Templeton, M.C., Evans, M.C. New approaches to coding and analyzing mortality data. Proceedings of the Social Statistics Section. Washington, D.C.: American Statistical Association, pp 20-24 (1972).

4. Israel, R.A., Armstrong, R. An alternative procedure for classifying and analyzing mortality data. Liege, Belgium: International Union for the Scientific Study of Population, pp 231-241 (1973).