STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 23
English only

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing
(Prague, Czech Republic, 14-17 October 1997)

Item 4 of the provisional agenda

EVALUATION OF AN AUTOMATED MULTIPLE CAUSE OF DEATH CODING SYSTEM

Submitted by the National Center for Health Statistics, U.S.A[1]

[1] Prepared by Kenneth W. Harris.

I. BACKGROUND

1. In 1967 the National Center for Health Statistics (NCHS) began automating the entry, classification, and retrieval of cause-of-death information as reported on death certificates. The Mortality Medical Data System [1] has continued to evolve to meet new requirements and incorporate new technology.

2. ACME, the "Automated Classification of Medical Entities," was the first component of the Mortality Medical Data System. Under ACME, conventional International Classification of Disease (ICD) codes are manually assigned to each entity (e.g., disease, condition, accident, or injury) listed on cause-of-death certificates, preserving the location and order as reported by the certifier. ACME subsequently applies World Health Organization (WHO) rules to the ICD codes and selects an underlying cause of death [2].

II. MICAR

3. MICAR, "Mortality Medical Indexing, Classification, and Retrieval," is the latest development of the Mortality Medical Data System. MICAR was designed to automate the instructions, rules, and code assignments for generating cause-of-death data. The development of the system began in 1983 and was certified for use in 1992. During this period of research and evaluation, a wide variety of nosological, statistical and analytic resources were devoted to the project. In addition, a number of innovative techniques were employed in developing sophisticated computer software packages that contributed to the successful completion of MICAR. The development of the MICAR dictionary, for example, was an iterative process that ultimately resulted in a volume containing more than 100,000 entries. MICAR allows data entry operators to enter full text, abbreviations, or reference numbers for cause-of-death terms. After data from the death certificate are keyed, MICAR matches each cause-of-death term to the MICAR dictionary. All dictionary entries are stored in standardized nomenclature in electronic form. Each matched term is assigned an entity reference number (ERN), its unique identifier in the system. Any record with an unmatched term is rejected for manual review.

4. Records on which **all** terms have an entity reference number are processed by the MICAR rules application software. Code assignments are made based on the presence or absence of cause-of-death terms and their positional relationship to one another. When a record contains a term for which a coding rule has not been automated, or a term involving a sensitive or infrequently occurring ICD category, it is rejected for manual review.

III. DESCRIPTION OF MICAR DATA ENTRY TERMINOLOGY

5. Data entry under MICAR requires that the data entry operator enter the full text in standardized nomenclature, an abbreviation, or a numeric code for each disease, injury, and external cause reported on a death certificate. While MICAR reduces the complexity of manual multiple cause-of-death coding, terms still must be translated into a format acceptable to MICAR. The principles for standardizing medical entities illustrated in the following paragraphs have been incorporated in the Vital Statistics Manual, Part 2g [3], which contains the instructions for reading the medical certification, separating one entity from another, standardizing and entering MICAR nomenclature, and entering other data essential to the

medical classification.

III.1 Description of Medical Entities

6. A medical entity is a word or set of words that describes a cause of death. It may be a disease, a disease process, abnormality, disorder, symptom, complication, injury, poisoning, or a mode of dying (i.e., respiratory arrest). For purposes of MICAR data entry, it is important to consider entities as being divided into three groups: natural diseases, injuries, or adverse reactions
caused by some external force (e.g., violent deaths).

7. Typically both diseases and injuries are reported as either one word (e.g., emphysema, burns) or as multiple words (e.g., cardiac arrest, open wound) that are adjacent to each other. External causes (e.g., accidents, falls, fires) are often reported in a set of words or phrases not adjacent to one another (e.g., collided with another car--ran off road--driver killed). With external causes, the rearrangement is more difficult than with diseases or injuries.

Entity Reference Numbers

8. Each entity listed in the MICAR dictionary has an associated three- to six-digit number indicating the entity's numeric representation in the MICAR system. This unique numerical code is referred to as the "Entity Reference Number" (ERN).

9. Several examples of common variations in reporting medical entities follow. In each group, the data entry operator ignores extraneous information and translates each entity into standardized MICAR nomenclature:

Example 1--Reported as: Disease, hypertensive arteriosclerotic cerebrovascular (OR) Arteriosclerotic cerebrovascular disease, hypertensive

Standardized MICAR Nomenclature:
HYPERTENSIVE ARTERIOSCLEROTIC CEREBROVASCULAR DISEASE (ERN=82792)

Example 2--Reported as:
Fracture of arm (OR) Fracture, arm

Standardized MICAR Nomenclature:
ARM FRACTURE (ERN=94911)

Conversion of Entity Reference Numbers to International Classification of Disease (ICD) Codes

10. The final piece of the MICAR system is the assignment of ICD codes. This is accomplished through a software package that converts the ERNs to ICD codes. Once the data entry operator has taken a death certificate entry and correctly transformed that entry to its corresponding terminology in the MICAR dictionary, the unique ERN is converted to the appropriate ICD code.

III.2 Certification of MICAR [4]

11. Testing and certification of the MICAR system was based on a benchmark file of approximately 92,000 records systematically selected from the 1988 U.S. Mortality file of two million records. Changes in the MICAR system were made on a continuous basis throughout the testing period. The file was reprocessed after each system update/change to assure that new coding rules and codes did not adversely affect records not intended to be affected.

12. At the end of the testing period, MICAR was completely processing 85 percent of the file records. Today, the level is at 90 percent. A record is completely processed when **every** disease entity on the record (average of three per record) is matched with one of the 100,000+ terms in the MICAR dictionary. Of equal importance, MICAR correctly flags the 10 percent of the file that it can not completely process.

13. At its most detailed level, (usually 4 digits) the ICD-9 contains approximately 5,000 codes. The next broadest level of detail is the "282 cause list." To assure representativeness of the benchmark file, its percentage distribution was compared with that of the 1988 U.S. file at the 282 cause list level. The 15 leading disease entities (i.e., diseases listed somewhere on the death certificate in the multiple cause of death sequence) accounted for 62 percent of the U.S. file and 65 percent of the benchmark file. For the leading 30 entities, the percentages were 80 and 82, respectively.

14. Procedurally, when a term on the death certificate is determined to be valid (i.e., matches a MICAR dictionary term) it is assigned an Entity Reference Number (ERN). For each completely processed record, the automated coding rules examine the relationship between the ERNs on that record and apply a series of rather sophisticated logic steps before a final set of ICD codes is assigned.

III.3 Quality of MICAR

15. As mentioned earlier, MICAR can now completely code about 90 percent of all records. In terms of individual disease entities, it can code in the order of 96 percent.

16. The quality of MICAR coding was determined by using multiple cause codes normally assigned by the NCHS nosologists as the standard. In other words, when a MICAR-generated ICD code matched the code assigned by a nosologist for that disease entity, it was considered a correct code.

17. Overall, at the time of certification, MICAR coded the benchmark file with a system error rate of 0.74 percent on a multiple cause basis and 0.42 percent at the underlying cause level. Many causes, of course, occurred very infrequently in the benchmark file. As an additional measure after certification, every cause that had, during the testing period, 50 or fewer occurrences and/or an error rate greater than five percent at the time of certification, was automatically flagged for manual review when it was selected. This manual review process continued (on a cause by cause basis) until sufficient knowledge was gained about the validity and reliability of the MICAR system with regard to the particular cause in question. For comparative purposes, it can be pointed out that NCHS nosologists have generally performed multiple cause and underlying cause coding at error rates of about 3 1/2 percent and 2 percent, respectively. It should be recalled that the underlying cause of death is derived by the ACME system from the set of multiple cause codes on each certificate. Many times an incorrect, but "close," multiple cause code will still lead to the same

underlying cause code; that is why the underlying cause error rate is lower (sometimes by as much as 50 percent) than the multiple cause error rate.

III.4 Benefits of MICAR

18. NCHS made a major resource investment in developing MICAR because of its anticipated tremendous payoff potential. The successful implementation of MICAR has:

> **R**educed **P**ersonnel **R**equirements - Automated coding rules allow a data entry person to become productive in one month, compared with the four month class-room training and twelve month on the job training required to become a proficient nosologist.
>
> **I**mproved **A**ccuracy and **C**onsistency - Automated coding rules eliminate errors in recognizing terms, applying coding rules, using the ICD index, and keying.
>
> Simplified the ICD Revision Process - MICAR assigns each cause-of-death term an entity reference number that is independent of ICD codes. Once causal relationships are established among the terms, the relationships do not have to be reestablished with each revision of the ICD. The use of entity reference numbers also simplifies developing comparability ratios between ICD revisions. Because data entry personnel do not manually assign codes, retraining with each ICD revision will not be necessary.
>
> Allowed for the Retrieval and Reconstruction of Exact Terminology - Under conventional ICD coding, considerable terminological detail is lost when diagnostic descriptions are collapsed into summary categories. With MICAR, it is possible to resurrect the exact diagnostic terms used by the physician, which are captured in the ERNs.

IV. QUALITY CONTROL PROCEDURES FOR MICAR DATA ENTRY CODERS

19. In addition to the ongoing procedures for assessing the quality of the automated MICAR system, a systematic sample of the work of each data entry operator is measured using a two way independent verification procedure. It should be remembered that the quality of the MICAR system relies on the correct input (i.e., proper use of MICAR nomenclature) from the data entry operator. The two way system is a modification of the three way independent verification system which has long been the preferred method by many organizations that use difficult or extensive coding schemes in their data reduction operations.

Three Way Independent Sample Verification

20. Under this procedure, a batch of records is coded by three coders. A production coder codes all of the records and two sample coders independently code the same sample of records from the batch. The two sets of sample records and the corresponding records from the production coder are computer matched and the majority code (i.e., coded by at least <u>two</u> of the coders) is selected as the correct code. Error rates for each coder are computed on this

basis. If the production coder's error rate exceeds the acceptable level, the entire production batch (100 percent) is recoded and re-verified. The three way system assumes that a condition or description leads to only one valid code; thus, when two or three out of three coders with comparable coding skills independently arrive at the same code, there is a high probability that the selected code is correct.

Two Way Independent Verification

21. Although three-way independent verification is considered a highly efficient system, resource constraints forced the Center to investigate less expensive, but equally valid, ways of measuring quality - particularly in view of the recognized less precise knowledge regarding measurement errors in data collection processes. The successful introduction of two-way independent verification in the Center's three major medical coding operations occurred in 1982 [5] and, after testing, was incorporated into the MICAR data entry operation in 1994. The appeal of the two-way independent system is that it also uses the majority rule concept, which is the key feature of the three way system. When at least two coders in a three-way system agree on a code (AAB or AAA), then A, the majority code, is considered the correct code. In a two-way system, AAA cases become AA cases and AAB cases become AA (one-third) and AB (two-thirds). Three-way differences (ABC) are rare, but they, too, become AB cases (i.e., two non-matching codes) in the two-way system.

22. In an earlier study of the quality of mortality medical coding [6], Harris and French found that AAA cases comprised 90.3 percent of all cases and AAB cases accounted for just over nine percent of all cases. At these levels, the use of a two way system would provide the same measurement precision (majority rule) for more than 93 percent of the codes [90.3+ 1/3(9)] while reducing the number of coders used from three to two. The third coder would be needed as an adjudicator only for the AB cases (6.7 percent).

23. These findings indicated that substantial resource savings could be realized, at no diminution of quality measuring capabilities, when

(a) The quality of coding within a coding unit is, in general, homogeneous.
(b) Correct coding solutions are unique.

These conditions enhance the concept of the majority rule as a valid identifier of the correct code and a reliable measure for establishing coding quality. As one might expect, however, the validity and reliability are much greater when the majority code is arrived at unanimously.

24. In the mortality study cited above, it was found that Code A was correct 98.5 percent of the time that each of three coders independently selected Code A (AAA cases) but was correct slightly less than 80 percent of the time when only two of the three coders selected it ($\overline{AAB}$ cases). Overall, the selection of Code A was correct about 97 percent of the time. In addition, that study found a medical description led to a unique coding solution 98 percent of the time.

25. These additional findings further supported the plan to convert from three-way independent verification to two way independent verification. Since AAA cases comprised about 91 percent of the majority rule cases (AAA + $\overline{AAB}$), the AA cases in the two-way system are much more likely to represent what would have been AAA cases in a three-way

system (only a third of the $\overline{AAB}$ cases will become AA cases), thus, the correctness of code A is strongly indicated. If the adjudication of the remaining AB cases yields the correct code 80 percent of the time, then the two-way system will also produce the correct code about 97 percent of the time, thereby providing a measure of coding quality very similar to that of the three-way system. Several follow-up studies (the most recent in 1995 as part of the continuing quality control process for MICAR) conducted since the introduction of two-way independent verification have yielded similar results.

V. SUMMARY

26. After extensive research and testing, it was determined that the automated MICAR system could completely process about 85 percent (now at 90 percent) of the annual mortality file of 2 million records with greater accuracy than through manual multiple cause coding. In addition, the system consistently "recognizes" what it cannot code and flags those records for nosological review. Since training for data entry of natural language text is considerably less time consuming and less costly than nosological training, these results are of considerable import. Furthermore, the use of a two-way independent verification system provides a continuing measurement tool for assessing the quality of the data entry operators, whose work serves as input to the MICAR automated coding system. Thus, the overall assessment led to a recommendation to the Center Director that MICAR be certified for use in selected states beginning with 1991 mortality data and that it be implemented in additional states each year until full geographic coverage is achieved. At the beginning of 1997, more than 40 states were using MICAR.

27. The development and certification of MICAR should be regarded as a major contribution by NCHS, one that will surely be widely adopted by other countries, and that will likely be adopted in modified form for other applications, such as morbidity coding.

References

1. "The Mortality Medical Data System," Division of Vital Statistics and Division of Data Processing, NCHS, Research Triangle Park, NC, 1990.

2. National Center for Health Statistics. Instruction Manual, Part 2c, ICD-9 ACME Decision Tables for Classifying Underlying Causes of Death, 1995. Hyattsville, Maryland: Public Health Service, 1995.

3. National Center for Health Statistics. Instruction Manual, Part 2g, Data Entry Instructions for the Mortality Medical Indexing, Classification, and Retrieval System (MICAR), 1996-97. Hyattsville, Maryland: Public Health Service, 1996.

4. "Certification of MICAR." Memorandum: Harry Rosenberg and Kenneth W. Harris to Dr. Manning Feinleib, Hyattsville, Maryland, October 1992.

5. Kenneth W. Harris, "Quality Control Procedures for Medical Coding: National Center for Health Statistics," Proceedings of the American Statistical Association, Social Statistics Section, Washington, DC, 1983, PP. 480-484.

6. Kenneth W. Harris and Dwight K. French, “A Methodological Study of Quality Control Procedures for Mortality Medical Coding,” PHS Pub. No.80-1355, Series 2, No. 81, Public Health Service,Hyattsville, Maryland, March 1980.