STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 10
English only

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Statistical Data Editing
(Rome, Italy, 2-4 June 1999)

Topic (i): Measuring the impact of editing in various phases of statistical survey processing

SELECTED DATA EDITING PROCEDURES IN AN AUTOMATED MULTIPLE CAUSE OF DEATH CODING SYSTEM

Submitted by the National Center for Health Statistics, United States[1]

Contributed paper

I. BACKGROUND

1. As in so many other areas of modern life, computer technology is being rapidly and effectively applied to many aspects of vital statistics, including data collection, data production, and data dissemination. Better data quality, improved timeliness, and reduced resource requirements are among the major benefits of these technological enhancements [1].

2. In 1969, 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 [2] has continued to evolve to meet new requirements and incorporate new technology.

3. ACME, the "Automated Classification of Medical Entities," was the first component of the Mortality Medical Data System. Initially, nosologists, i.e., medical coders, coded all of the medical conditions on the death certificate, using the latest revision of the International Classification of Diseases (ICD). ACME subsequently applied World Health Organization (WHO) rules to the ICD codes and selected an underlying cause of death [3].

4. Since the introduction of ACME, other components of the Mortality Medical Data System have been developed. The second one, TRANSAX, from "TRANSlation of Axis," was developed to facilitate the tabulation and use of multiple cause-of-death data [4]. TRANSAX generates two different sets of multiple cause codes for tabulation. One is referred to as "entity axis," which tabulates each "entity" or "cause" on a record. The other, "record axis," is best for analyzing data on a record by record basis [5].

5. A third component, RETRIEVE, was developed in conjunction with TRANSAX to search the annual files for a particular disease and to tabulate the number of times the disease was reported alone or with other diseases.

---

1 Prepared by Kenneth W. Harris.

II. MICAR

6. MICAR, Mortality Medical Indexing, Classification, and Retrieval, is the fourth component of the Mortality Medical Data System. The major features of MICAR have been described previously [6] but will be repeated here in some detail so as to provide a better lead into the discussion of the final component, SuperMICAR.

7. 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 its development, a wide variety of nosological, statistical, and analytical resources were devoted to the system. 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.

8. 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

9. 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 [7], 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

10. 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 (e.g., respiratory arrest). For purposes of MICAR data entry, it is important to consider entities as being divided into three groups: natural causes, injuries, or adverse reactions caused by some external force (e.g., violent death).

11. 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.

III.1.1 Entity Reference Numbers

12. 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 identifier is referred to as the "Entity Reference Number" (ERN).

13. 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) Arterioscerotic 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)

III.1.2 Conversion of Entity Reference Numbers to International Classification of Disease (ICD) Codes

14. 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 [8]

15. Testing and certification of the MICAR system was based on a benchmark file of 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 codes and coding rules did not adversely affect records not intended to be affected.

16. At the end of the testing period, MICAR was completely processing 85 percent of the file records and eventually reached the 95 percent level. 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.

17. 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 sophisticated logic steps before a final set of ICD codes is assigned.

III.3 Quality of MICAR

18. The quality of MICAR was determined by using multiple cause codes assigned by the NCHS nosologists as the standard. Thus, when a MICAR generated ICD code matched the code assigned by a nosologist, it was considered a correct code. If there was no agreement, another nosologist independently coded the condition and the majority code was considered as the correct code.

19. 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. These rates, which were based on the ICD-9 coding structure, continued to decrease, eventually leveling off at 0.32 percent for multiple cause coding and 0.18 percent for underlying cause coding. For comparative purposes, it can be pointed out that NCHS nosologists have generally performed multiple cause coding 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 the certificate. Thus, it is possible to arrive at the correct underlying cause of death even when one (or more) of the multiple cause codes is incorrect.

20. ICD-10, which was implemented on January 1, 1999, introduced a number of substantive changes from ICD-9. These included: rearrangement of some chapters, regrouping of certain conditions, title changes, and expansion of 4-digit categories from 5,000 to 8,000. The latter change, in particular, allows for a much greater level of specificity in ICD-10.

21. In view of these changes, it was clear that MICAR would have to be recertified. This time a benchmark file of 137,000 records from the 1996 U.S.Mortality file was selected to test the quality of MICAR coding. The testing procedure was very similar to the one used for the original certification of MICAR.

22. Using 1CD-10 coding rules, MICAR completely processed 95 percent of the records, the same level it had achieved with ICD-9.

23. MICAR coded the ICD-10 benchmark file with a system error rate of 0.68 percent on a multiple cause basis--about the same level (0.74) at which the ICD-9 benchmark file was coded during MICAR's initial certification.

IV. SuperMICAR

24. The latest component of the Mortality Medical Data System is SuperMICAR. Those of you familiar with American football know the word "super" defines the ultimate professional football game played in America, the Super Bowl. Similarly, SuperMICAR is viewed as the ultimate mortality coding system.

25. It usually takes a nosologist about 18 months to become a truly proficient medical coder--5-6 months of classroom training and 12 months of on the job experience.

26. A data entry operator can become a proficient MICAR coder in one month, the amount of time needed to learn and understand a basic level of medical terminology in order to translate death certificate data to standardized MICAR nomenclature.

27. With SuperMICAR, there is virtually no learning period at all. SuperMICAR accepts almost everything in literal text. The data entry operator enters condition descriptions **exactly** as they are shown on

the certificate. In addition, the computer screen the operator uses is in the same format as the standard death certificate. So, the operator takes information from Part 1, line 1 of the death certificate and enters it on a matching computer screen on Part 1, line 1. Even the interval between onset of the disease and death, shown in the "duration field," is entered in full text [5]. The literal text is internally assigned an ERN, which eventually gets converted to an ICD-10 code.

28. Despite its instant viability, SuperMICAR will not completely replace MICAR. The use of one or the other depends on a number of factors. Many states that have the capability to do so will probably continue to use MICAR, rather than switching to SuperMICAR. In 1999, about 64 percent of all mortality records are expected to be processed by MICAR and the remaining 36 percent by SuperMICAR.

29. Perhaps not surprisingly, SuperMICAR is not as efficient as MICAR. About 80-85 percent of records can be processed by SuperMICAR, compared with 95 percent by MICAR. MICAR's system error rate is 0.68 percent, while SuperMICAR's error rate is more than double that level at 1.50 percent.

30. Though not yet as efficient as MICAR, SuperMICAR brings a number of enhancements to the Mortality Medical Data System. In particular, SuperMICAR is facilitated by its use of a processing screen that matches the standard U.S. death certificate. Thus, a data entry operator is able to copy data, word for word, from the death certificate to the same location on a matching SuperMICAR processing screen.

V. AUTOMATED DATA EDITING

31. Automated data editing procedures for MICAR and SuperMICAR are quite extensive [9]. The major ones, in terms of their impact on the standardization and quality of multiple cause coding classification, are listed below. **(Note--Examples of each will be illustrated at the work session.)**

a) Relating and modifying
b) Modification tables
c) Intent of certifier
d) Traumatic conditions
e) Late effects

32. Relating and modifying procedures are implemented when a condition that is usually classified in ICD-10 according to the site affected, e.g., atrophy, congestion, gangrene, rupture, etc., is reported without a site specification. If another condition, with site specified or implied, is also reported on the certificate, then the condition without site is coded to the same site as the related site. For example, the condition "Hemorrhage," reported alone, is coded to "Hemorrhage, NOS" (Not Otherwise Specified). But when reported along with "Salmonella," an intestinal infection, it is coded to "Intestinal Hemorrhage."

33. Modification tables are applied when the reported condition is (or may be) affected by the age or sex of the decedent, or by the duration of the condition between onset and death. With regard to age, "newborn" means less than 28 days of age at time of death; "infant" means less than one year of age at time of death; "child" means less than 18 years of age at time of death. In terms of duration, "acute" means up to four weeks; "chronic" means more than four weeks. A condition is "congenital" if it existed from birth; otherwise it is "acquired." All of these distinctions are considered in determining the appropriate cause of death. For example, the reported condition "Brain Damage Syndrome" is normally coded to "Disorder of Brain," but for under age one year, it is coded to "Infantile Cerebral Palsy, Unspecified."

34. In order to arrive at the most appropriate code for a given diagnostic entity, it is sometimes necessary to take other reported information and the order in which that information is reported into account in order to

assure that the certifier's intent is conveyed. Here the objective is to code each diagnostic entity to reflect the certifier's intent without combining separate codes into one. For example, the condition "Interventricular Septal Defect," when reported as due to any of several entities, such as "Renal Failure" or "Urinary Tract Infection," becomes "Acquired Heart Septal Defect."

35. In ICD-10, some conditions have both a traumatic and non-traumatic code. These conditions are to be coded as traumatic when reported as "due to" or with an injury, a trauma, or an external cause. For example, "Emphysema," when reported alone, is coded to "Emphysema, Unspecified." When it is reported as due to an injury, such as "Rib Fracture," it is coded to "Traumatic Emphysema."

36. When death is attributed to a condition that occurred a year or more before death, the "late effects" procedures are used to modify the ICD-10 code representing the cause of death. For example, the ICD-10 code for "Acute Poliomyelitis" is A80.9, "Acute Poliomyelitis, Unspecified." However, if death occurred a year or more after onset, the code becomes B91.0, "Sequelae of Poliomyelitis."

37. Automatic data editing procedures also apply to specific conditions, such as "Rheumatic Heart Disease," and to specific families of conditions, such as "Neoplasms."

38. The advances made by NCHS in the development of MICAR and SuperMICAR, and the automation of a wide variety of complex data editing procedures designed to assure their high quality, continue to provide major contributions in the field of mortality medical coding. Their growing widespread use is ample evidence of that fact. These systems, in whole or in part, have been adopted by England, Canada, Spain, Japan, Sweden, Brazil, Australia, Scotland, Israel, and Italy.

References

1. Harry M. Rosenberg, "International Collaborative Effort on Automating Mortality Statistics: Background and Issues." Proceedings of the International Collaborative Effort on Automating Mortality Statistics, Washington, D.C. November 1996.

2. "The Mortality Medical Data System," Division of Vital Statistics and Division of Data Processing, NCHS, Research Triangle Park, N.C. 1990.

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

4. Ronald F.Chamblee and Marshall C. Evans, "TRANSAX, the NCHS System for Producing Multiple Cause of Death Statistics, 1968-78." PHS Pub. No.86-1322, Series 1, No. 20, Public Health Service, Washington, D.C. June 1986.

5. Donna E. Glenn, "Description of the National Center for Health Statistics Software Systems and Demonstrations." Proceedings of the International Collaborative Effort on Automating Mortality Statistics, Washington, D.C. November 1996.

6. Kenneth W. Harris, Harry M. Rosenberg, Kenneth D. Kochanek, Ronald F. Chamblee, and Donna E. Glenn, "Evaluation of an Automated Multiple Cause of Death Coding System." Proceedings of the American Statistical Association, Social Statistics Section. Washington, D.C. August 1993.

7. National Center for Health Statistics. Instruction Manual, Part 2g, Data Entry Instructions for the Mortality Medical Indexing, Classification, and Retrieval (MICAR), 1998-99. Hyattsville, MD. Public Health Service, 1998.

8. “Certification of MICAR.” Memorandum: Harry M. Rosenberg and Kenneth W. Harris to Dr. Manning Feinleib, Hyattsville, MD. October 1992.

9. National Center for Health Statistics. Instruction Manual, Part 2b, Instructions for Classifying Multiple Causes of Death, 1998-99. Hyattsville, MD. Public Health Service, 1998.