Revista de Saúde Pública

J O U R N A L   O F   P U B L I C   H E A L T H

Comparative evaluation of underlying causes of death processed by the Automated Classification of Medical Entities and the Underlying Cause of Death Selection Systems

Avaliação comparativa das causas básicas de morte processadas pelos Sistemas "Automated Classification of Medical Entities" e de Seleção de Causa Básica

Augusto H. Santo, Celso E. Pinheiro e Eliana M. Rodrigues

Departamento de Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo. São Paulo, SP - Brasil (A.H.S.); Departamento de Informática (DATASUS) da Fundação Nacional de Saúde. Rio de Janeiro, RJ - Brasil (C.E.P.); Fundação Sistema Estadual de Análise de Dados, São Paulo, SP - Brasil (E.M.R.)

Comparative evaluation of underlying causes of death processed by the Automated Classification of Medical Entities and the Underlying Cause of Death Selection Systems[*]

Avaliação comparativa das causas básicas de morte processadas pelos Sistemas "Automated Classification of Medical Entities" e de Seleção de Causa Básica

Augusto H. Santo, Celso E. Pinheiro e Eliana M. Rodrigues

Departamento de Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo. São Paulo, SP - Brasil (A.H.S.); Departamento de Informática (DATASUS) da Fundação Nacional de Saúde. Rio de Janeiro, RJ - Brasil (C.E.P.); Fundação Sistema Estadual de Análise de Dados, São Paulo, SP - Brasil (E.M.R.)

Abstract

Introduction   The correct identification of the underlying cause of death and its precise assignment to a code from the International Classification of Diseases are important issues to achieve accurate and universally comparable mortality statistics These factors, among other ones, led to the development of computer software programs in order to automatically identify the underlying cause of death.

Objective   This work was conceived to compare the underlying causes of death processed respectively by the Automated Classification of Medical Entities (ACME) and the "Sistema de Seleção de Causa Básica de Morte" (SCB) programs.

Material and Method   The comparative evaluation of the underlying causes of death processed respectively by ACME and SCB systems was performed using the input data file for the ACME system that included deaths which occurred in the State of S. Paulo from June to December 1993, totalling 129,104 records of the corresponding death certificates. The differences between underlying causes selected by ACME and SCB systems verified in the month of June, when considered as SCB errors, were used to correct and improve SCB processing logic and its decision tables.

Results   The processing of the underlying causes of death by the ACME and SCB systems resulted in 3,278 differences, that were analysed and ascribed to lack of answer to dialogue boxes during processing, to deaths due to human immunodeficiency virus [HIV] disease for which there was no specific provision in any of the systems, to coding and/or keying errors and to actual

problems. The detailed analysis of these latter disclosed that the majority of the underlying causes of death processed by the SCB system were correct and that different interpretations were given to the mortality coding rules by each system, that some particular problems could not be explained with the available documentation and that a smaller proportion of problems were identified as SCB errors.

Conclusion These results, disclosing a very low and insignificant number of actual problems, guarantees the use of the version of the SCB system for the Ninth Revision of the International Classification of Diseases and assures the continuity of the work which is being undertaken for the Tenth Revision version.

Underlying cause of death. Information sistems. Vital statistics.

Resumo

Introdução *A identificação correta da causa básica de morte e a atribuição de código preciso da Classificação Internacional de Doença à mesma são importantes para a obtenção de estatísticas de mortalidade confiáveis e passíveis de comparabilidade universal. Estes fatores, dentre outros, levaram ao desenvolvimento de programas de computador para identificar automaticamente a causa básica de morte.*

Objetivo *Este trabalho teve a finalidade de comparar a causa básica de morte identificada respectivamente pelos programas Automated Classification of Medical Entities (ACME) e pelo Sistema de Seleção de Causa Básica de Morte (SCB).*

Material e Método *O arquivo para a entrada de dados sobre causas de morte (input file) para o Sistema ACME contendo registros de 129.104 declarações de óbito de mortes ocorridas no estado de São Paulo de junho a dezembro de 1993 foi utilizado para o processamento da causa básica pelo SCB. Os problemas identificados pelo processamento dos registros do mês de junho foram considerados para o aprimoramento do sistema SCB.*

Resultados *Foram encontras 3.278 causas básicas de morte identificadas de modo diferente pelos programs ACME e SCB. Essas diferenças foram atribuídas à falta de resposta adequada a janelas de diálogo durante o processamento pelo SCB, a óbitos por doenças devida a vírus da imunodeficiência adquirida para os quais não havia tabelas de decisão específicas, a erros de codificação e/ou digitação e a problemas propriamente ditos. A análise pormenorizada deste últimos mostrou que, em sua maioria, as causas básicas processadas pelo sistema SCB estavam corretas, que diferentes interpretações das regras de mortalidade foram dadas pelos sistemas comparados, que alguns problemas particulares não tiveram explicação adequada por falta de documentação sobre os mesmos e que uma menor proporção de problemas consistia de erros do SCB.*

Conclusões *O número pequeno e praticamente insignificante de problemas encontrados garante o uso da versão do SCB para a Nona Revisão da Classificação Internacional de Doenças e assegura a continuidade dos trabalhos relativos à sua versão para a Décima Revisão.*

Causa básica da morte. Sistemas de informação. Estatísticas vitais.

Comparative evaluation of underlying causes of death
Santo, A. H. et al.

Rev. Saúde Pública, 32 (1), 1998

3

INTRODUCTION

The identification of the underlying cause of death and its subsequent assignment to a code from the International Classification of Diseases (ICD) is an activity undertaken by the coders. This technical professional receives specialised training that is internationally standardised by the World Health Organization; he is taught how to interpret the medical data mentioned on the death certificate and learns how correctly to apply the mortality coding rules and their related guidelines. Nevertheless, this activity is liable to errors that may endanger and impair the quality of the coding, and, in consequence, the mortality statistics also. These mistakes are due to many factors, such as those related to the type of training that the coder receives, to different interpretations given to etiologic relationships among pathological conditions, to lack of consideration of some diagnostic entities registered on the death certificate, to the omission of rules that should be applied and to mistakes in the transcription of codes[10, 12, 13].

The difficulties and problems related to the work done by coders, the limitation of mortality statistics based only on the underlying cause of death and other needs for processing mortality data have led to the development of software designed to identify causes of death. One of these software programs has been developed by the National Center for Health Statistics (NCHS) and received the designation of ACME, acronym for "Automated Classification of Medical Entities". The ACME system guarantees greater precision and quality to the identification, selection and coding of the underlying causes of death, identified by means of decision tables that include the code structure of ICD, etiologic relationships among entries represented by the codes and by processing these according to the mortality coding rules.[1, 2, 9, 10, 11].

The interest of the WHO Collaborating Centre for the Classification of Diseases in Portuguese (S. Paulo Centre) and the high spirit of scientific collaboration prevailing in the NCHS has permitted the introduction of the ACME system to process causes of death in the State of S. Paulo effectively since 1983[11]. As a result of this work, the Ministry of Health and other Brazilian States have also begun to call for access the ACME system. Nevertheless, it has not been possible to expand this use on account of hardware requirements, lack of trained coders and other technical limitations[9]. More recently, the municipalization and decentralization of health services in Brazil has increased the need for timely regional and local mortality data to support planning and health surveillance activities. These factors have led to the launching of the idea of developing a software program suitable for use with microcomputers in order to provide cause of death data[8, 12, 13].

During 1993 this software program was developed in collaboration with the S. Paulo Centre and the "Departmento de Informática (DATASUS) da Fundação Nacional de Saúde". It was called Underlying Cause of Death System for Microcomputers ("Sistema de Seleção da Causa Básica de Morte" - SCB) and its characteristics have been presented elsewhere[8, 12, 13].

In order to develop the SCB, the standard structure of the mortality coding rules and their related guidelines were used and this product was applied to typical death certificates. The adequacy of the process was verified, using however , only a small number of death certificates. There were still doubts about the working of the system when, in everyday practice, routine death certificates would be processed.

One preliminary test with a larger number of death certificates was performed with a mortality data file from the State of Paraná, in which, besides the underlying cause, other associated causes of death were also available. In about 30,000 records, a comparison was made between the underlying cause processed by SCB and the manually selected cause. The results of this comparison were very good, 5,995 differences having been found of which only 350 were due to errors of the system. The other 5,645 differences were due to manual selection or coding errors. The SCB system attained a degree of correctness of 98.8% and it was discovered that the larger part of the errors were due to problems in the decision tables. In spite of these results, a certain insecurity persisted in relation to the real adequacy of the system. The structural format of the records from the mortality file used in this test had not been prepared to input data precisely, i.e., the codes did not retain the identification of the line and its position on lines that included more than one condition on the original medical death certificate. This fact could distort the processing of these causes during the selection and modification of the underlying cause of death[8, 12, 13].

In view of the fact that the input data file for the ACME system presents a structure and syntax format that permits the identification of the line and the position on the line where the causes are mentioned on the original medical death certificate, it was considered convenient to use this type of file to test the SCB system[7, 9]. A particular death certificate, in order to be included as a record in this file, must re-

ceive codes for all entries registered in Parts I and II. The entries in Part I are separated by a slash (/) which corresponds to the phrase "due to or as a consequence of" and are distinguished from the entries in Part II by an asterisk (*). More than one entry on the same line in Part I and entries in Part II are registered with a blank space between their corresponding codes. The decimal point between the third and fourth digits of four digit categories is not entered when coding. An example of the format of this type of record, considering the above described structure and syntax, may be illustrated as follows:

```
        Part I        Part II
     a   /   b   / c
 7800/436 431/4019*2500 4409
```

This format of input file for the ACME system can also be used as an input file for the SCB with some adaptations.

MATERIAL AND METHOD

The comparative evaluation of the underlying causes of death processed respectively by ACME and SCB systems was performed using the input data file for the ACME system that included deaths which occurred in the State of S. Paulo from June to December 1993, totalling 129,104 records of the corresponding death certificates. Table 1 presents the distribution of these records according to the month of occurrence of the death and some facts and variables that were observed.

Some adaptations were made in order to process the files in batch mode since the SCB system processes records individually, i.e., one record at a time. On account of these adaptations, the variables sex and age were not used during the confrontation and it was not possible to check their compatibility with causes of death. Also, for batch processing, all answers to dialogue boxes were fixed as negatives.

The initial differences between underlying causes selected by ACME and SCB systems verified in the month of June, when considered as SCB errors, were used to correct and improve SCB processing logic and its decision tables. Therefore, the months from July to December were processed after the changes deemed necessary had been made.

From the original monthly files, 1,389 records containing invalid ICD and ACME codes were eliminated. These records are identified as invalid in Table 1. Also, 317 records in which only the underlying cause of death was registered were considered as null records, since it would not be possible to perform the confrontation between ACME and SCB, as the fields corresponding to the medical form were blank, without any entry.

Conventionally, the number of death records used as the basic number for the calculation of proportional values was obtained by subtracting invalid and null records from the number of the original files, resulting in 127,398 records for comparison.

RESULTS

The differences observed between underlying causes selected by ACME and SCB systems were 3,278. It can be noticed that the number of differences found in June is larger than the corresponding numbers from July to December. This fact is to be explained by the improvement of the SCB system after the experience of the first month and correction of the errors observed in June; for this reason the total percentages of columns are not calculated (NC).

The factors contributing to these differences were classified as due to "answers to dialogue boxes", processing of death certificates with mention of "human immunodeficiency virus [HIV] disease (AIDS)", problems of "codification and/or of keying" and, generally speaking, "problems", which will be presented and discussed later.

Differences due to answers to dialogue boxes have occurred on account of the necessity to fix all answers as negative, to allow SCB batch processing. If the records were processed one by one, the adequate

Table 1 - Distribution of death records by month of occurrence, invalid records(Inval), null records (Null), basic number for calculation (Base), differences between ACME and SCB (Diff), dialogue box differences(Dial), AIDS deaths, codification or keying issues (C/K) and final problems (Probl), State of S. Paulo, 1993.

Month	Deaths	Inval	Null	Base	Diff	%	Dial	%	AIDS	%	C/K	%	Probl	%
Jun	20,552	174	0	20,378	603	2.96	322	1.58	55	0.27	64	0.31	162	0.79
Jul	19,013	203	46	18,764	435	2.32	279	1.49	23	0.12	60	0.32	73	0.39
Aug	20,102	248	252	19,602	483	2.46	318	1.62	17	0.09	65	0.33	83	0.42
Sep	17,943	199	3	17,741	455	2.56	283	1.60	23	0.13	72	0.41	77	0.43
Oct	16,364	197	1	16,166	421	2.60	266	1.65	27	0.17	67	0.41	61	0.38
Nov	17,709	167	5	17,537	440	2.51	300	1.71	25	0.14	59	0.34	56	0.32
Dec	17,421	201	10	17,210	441	2.56	276	1.60	17	0.10	74	0.43	74	0.43
Total	129,104	1,389	317	127,398	3,278	NC	2,044	NC	187	NC	461	NC	586	NC

Comparative evaluation of underlying causes of death
Santo, A. H. et al.

Rev. Saúde Pública, 32 (1), 1998

5

answer would be offered and the underlying cause of death would also coincide with the one selected by the ACME system. It must be noticed that in many of these cases, that include ambivalent decision tables entries, the underlying cause should also be confirmed by a qualified coder after the rejection by the ACME system. These differences totalled 2,044 and a significant variation between months was not observed, with proportional values between 1.49% and 1.71%.

The differences due to human immunodeficiency virus [HIV] disease must be stressed because the version of the ACME system used in the State of S. Paulo comes from the early 1980's and had not included provisions in its decision tables for the processing of death certificates with mentions of AIDS. The underlying, cause was assigned manually in the 187 records from the ACME input file[4-7].

The differences observed, classified as due to codification and/or keying, were 461. Examples of this type of difference in the original file are records of deceased male persons for whom codes on the medical form corresponded to conditions applying to females only but, after the errors being noticed, corrections were made by the coder only in the field of the underlying cause. The SCB system processed the medical form and selected the wrong sex cause, thus presenting the difference mentioned. Analogous facts occurred when valid codes for restricted age groups were used incorrectly and the mistake was corrected only in the field of the underlying cause of death. It must be remembered that the compatibility of sex and age was not checked.

The newly created codes for the ACME system are included among these differences due to codification and/or keying. Some categories in the ICD, generally called residual categories, include terms which are dissimilar in their causal relation to or combination with other categories and must be designated as having an ambivalent (maybe) relationship in the ACME decision tables, compelling a manual review of the underlying cause assignment by the coder. This fact reduces the capability of automated assignment. NCHS has chosen to eliminate ambivalent rejects for 16 such high frequency ICD categories by removing certain inclusion terms and creating an artificial category not currently used for those terms for which precise specification can be foreseen in the decision tables. If any of these terms are registered in the medical form, the coder must remember to use the created ACME code and not the regular ICD code; if this is not done an improper underlying cause will be identified. The SCB system also adopted the idea

of creating codes for processing these ambivalent categories and automatically transforms the regular ICD code into a created one, depending on the context of the causes mentioned on the medical form or, when this is not possible, asks the coder about the related conditions through a dialogue box in order to perform the transformation or not. Differences occurred when the ACME coder used a regular code and the SCB transformed it into a created code, resulting in a proper cause[4-8, 12, 13].

Adapted three digit categories, mainly included in Chapter XVII and related to the nature of injuries, ampersand (&) and the signs to denote the first digits "8" and "9" for nature of injuries categories, respectively "(" and ")", also lead to differences classified here[3-7].

The issues related to codification and/or keying that impair the quality of the underlying cause of death directly interest the SEADE Foundation and therefore will be discussed in a separate paper, complementing a previous evaluation made on another occasion[9].

The differences named "problems", verified in the month of June, to a percentage of 0.79%, were significantly greater than those observed in the subsequent months from July to December, which presented a variation of from 0.32% to 0.43%. These values reflect the improvement in the SCB processing.

Under the designation "problems", several kinds of differences are classified and presented in Table 2. The most frequent, identified in the table as SCB OK and totalling 303 records, refer most frequently to differences that occur on the ground of adaptations introduced in the SCB system in order to cope with and interpret more properly causes of death in Brazil. As an example, for Chagas' disease and for Schistosomiasis (mansoni), conditions not usually prevailing in the United States of America, specific decision tables have been prepared and included in the SCB. These differences should not, therefore be, deemed errors of either system[5, 6].

Nevertheless, those death certificates with regard to which, in the opinion of the authors, the ACME system gave a less adequate interpretation and application of the mortality coding rules, were considered as being correctly processed by the SCB system and were included among these 303 records.

Another type of difference also classified as problematic, relates to certain specific causes of death, each of them in very small numbers, for which no plausible explanation in the documentation of either system could be found. It is supposed that codification or keying mistakes must lie behind some of these

Table 2 - Distribution of death records by month of occurrence, basic number for calculation(Base), differences (Diff), total problems (Probl), correct solutions by SCB, certain specific problems (Spec) and errors (Err) of SCB, State of S. Paulo, 1993.

Month	Base	Diff	%	Probl	%	SCB OK	%	Spec	%	Err	%
Jun	20,378	603	2.96	162	0.79	50	0.2454	20	0.0981	92	0.4515
Jul	18,764	435	2.32	73	0.39	43	0.2292	12	0.0640	18	0.0959
Aug	19,602	483	2.46	83	0.42	50	0.2551	18	0.0918	15	0.0765
Sep	17,741	455	2.56	77	0.43	45	0.2536	18	0.1015	14	0.0789
Oct	16,166	421	2.60	61	0.38	37	0.2289	17	0.1052	7	0.0433
Nov	17,537	440	2.51	56	0.32	34	0.1939	17	0.0969	5	0.0285
Dec	17,210	441	2.56	74	0.44	44	0.2557	28	0.1627	2	0.0116
Total	127,398	3,278	NC	586	NC	303	NC	130	NC	153	NC

differences; to confirm this supposition the original Death Certificate should be made available to evaluate the coding and the syntax used to reproduce the medical form. Also considered as specific problems are death certificates with mention of surgical procedures or of adverse reactions of drugs, medicaments or biological substances, whose codes are included in the Supplementary Classification of External Causes and Poisoning of ICD-9. An ampersand (&) is used to identify the underlying cause by the ACME system but, on the other hand, the SCB system depends mostly on decision tables in which such cases are not foreseen[4-7, 9].

Finally, 153 records described and now acknowledged as SCB problems, were identified, the benefits of the changes made after the critical analysis of the June processing results having become evident. The error rate moved from about 1/220 death certificates, observed in June, to 1/8,600 death certificates in December. These results, disclosing a very low and insignificant number of actual problems, gives greater security for the use of the version of the SCB system for the Ninth Revision of the International Classification of Diseases and assures the continuity of its use with the Tenth Revision, the version of which is being undertaken.

REFERENCES

1. CHAMBLEE, R.F. & EVANS, M.C. New dimensions in cause of death statistics. *Am. J. Public Health,* **72**:1265-70, 1982.

2. ISRAEL, R.A.; ROSENBERG, H.M.; CURTIN, L.R. Analytical potential for multiple cause-of-death data. *Am. J. Epidemiol.,* **124**:161-79, 1986.

3. MANUAL de classificação estatística internacional de doenças, lesões e causa de óbito, baseada nas recomendações da Nona Conferência de Revisão. São Paulo, Centro da OMS para a Classificação de Doenças em Português, 1980. v. 1.

4. NATIONAL CENTER FOR HEALTH STATISTICS, VITAL STATISTICS DATA PREPARATION. *Instruction manual. Part 2b: Intructions for classifying multiple causes of death - 1984.* Hyattsville, Md., 1983.

5. NATIONAL CENTER FOR HEALTH STATISTICS, VITAL STATISTICS DATA PREPARATION. *Instruction manual. Part 2c: ICD-9 ACME decision tables for classifying underlying causes of death - 1984.* Hyattsville, Md., 1983.

6. NATIONAL CENTER FOR HEALTH STATISTICS, VITAL STATISTICS DATA PREPARATION. *Instruction manual. Part 2c: ICD-9 ACME decision tables for classifying underlying causes of death - 1992.* Hyattsville, Md. 1992.

7. NATIONAL CENTER FOR HEALTH STATISTICS, VITAL STATISTICS DATA PREPARATION. *Instruction manual. Part 2d: NCHS procedures for mortality medical data system*

file preparation and maintenance. Effective - 1979. Hyattsville, Md., 1980.

8. PINHEIRO, C.E. & SANTO, A.H. Um sistema especialista para a seleção da causa básica da morte. In: Congresso Brasileiro de Informática em Saúde, 4º. Porto Alegre, RS, 1994. *Anais.* Sociedade Brasileira de Informática em Saúde, 1994. v. 2, p. 161-2.

9. SANTO, A.H. *Avaliação da codificação e do processamento das causas de morte pelo Sistema ACME no Estado de São Paulo, 1992.* São Paulo, 1994. [Tese de Livre-Docência - Faculdade de Saúde Pública da USP].

10. SANTO, A. H. *Causas múltiplas de morte: formas de apresentação e métodos de análise.* São Paulo, 1988. [Tese de Doutorado - Faculdade de Saúde Pública da USP].

11. SANTO, A.H. & LAURENTI, R. Estatísticas de mortalidade por causas múltiplas: novas perspectivas com o Sistema ACME. *Rev. Saúde Pública,* **20**:397-400, 1986.

12. SANTO, A.H. & PINHEIRO, C.E. Selection of the underlying cause of death by microcomputer. In: Meeting of Heads of WHO Collaborating Centres for the Classification of Diseases, Caracas, Venezuela, 1994. Washington, D.C., World Health Organization, 1994. p. 1-7. (ESS/ICD/C/ 94.30).

13. SANTO, A.H. & PINHEIRO, C.E. Uso do microcomputador na seleção da causa básica de morte. *Bol. Oficina Sanit. Panamer.,* **119**:319-27, 1995.