Using SAS to Map ICD-8, ICD-9 and ICD-10 Codes to Common Codes

Angela M. Thompson, MSPH

Timothy S. Church, MD, MPH, PhD

The Cooper Institute, Dallas, TX, 75230

Address for correspondence and reprints:
Angela M. Thompson
The Cooper Institute
12330 Preston Road
Dallas TX 75230
Telephone: 972-716-7091
Telefax: 972-341-3221
Email: athompson@cooperinst.org

Abstract

The ICD-10 was implemented in 1999 with substantial changes to nomenclature and increased level of detail. Due to these changes, direct mapping from ICD-8 or ICD-9 to ICD-10 codes is not possible. Many studies face the challenge of using ICD-8 or ICD-9 and ICD-10 codes in analyzing mortality data produced by a longitudinal cohort. To overcome this problem while analyzing cancer deaths among a large cohort of men with diabetes, a translation 'database' was created to assign a common code for neoplasms as a cause of death to this cohort whose information dated from 1974 through 2003 and employed both ICD-8, ICD-9 and ICD-10 mortality codes. Once written, the translation code can be used with death certificate information from any cohort with minor modifications.

Introduction

An analysis of mortality data spanning nearly thirty years (1974-2003) faces the problem of using mortality data coded using the International Classification of Diseases versions eight (1968-78), nine (1979-1998), and 10 (1999-present). Our objective was to analyze all cancer deaths among diabetic men who had participated in the Aerobics Center Longitudinal Study (ACLS) in the years 1974-2003. Due to significant changes which were made to the ICD-10, it precluded using one version for all deaths in this cohort. In order to compare cancer deaths across ICD versions, it was necessary to establish a unidirectional map from each ICD version to a set of common codes for neoplasms due to significant changes which were made preventing using one version for all deaths in this cohort. The common codes were then used to classify cancer deaths by site. This paper details one method of mapping the ICD codes; it may be particularly useful to students and junior researchers who are new to SAS coding.

Objectives

The goal of data mapping in this case was to improve the information available for analysis by standardizing the cause of death across ICD versions without losing information. For malignant neoplasms, the version changes had relatively little effect on the overall cancer mortality statistics. The age-specific comparability ratio for malignant neoplasms in the United States ranged from 0.9773 (Relative Standard Error (RSE): 0.3574) in the 15-24 year old age group to 1.0219 (RSE:0.5587) among people aged 25-34 years indicating that relatively the same number of deaths were assigned to malignant neoplasms under ICD-9 and ICD-10. (Anderson et al. 1-32)A separate study using deaths

in England and Wales showed similar results with a comparability ratio of 1.009 for malignant neoplasms in people less than 75 years of age, 1.025 for people 75-84 years, 1.050 for people 85 and over and 1.022 among all ages. (Brock, Griffiths, and Rooney 7-17)

ICD Revisions

The ICD-8 coding used three and four digit numeric codes. The ICD-9 provided a fifth numeric digit. The ICD-10 provided the greatest set of modifications to the system of classification substantially increasing the level of detail, with nearly 3000 more categories, and employing an entirely new code using one alphabetic character and two to four numeric characters. (Anderson et al. 1-32;McBride et al. 44-48) Additionally, the method of selecting the primary site of malignant neoplasm was changed in the 10th revision whereby the order of entry no longer determines if the cancer is defined as primary or secondary. (Anderson et al. 1-32)

Methods

Our primary method of mortality surveillance was the National Death Index (NDI). The underlying cause of death was determined from the NDI report or by a nosologist's review of official death certificates obtained from the department of vital records in the decedent's state of residence. Cancer mortality was defined as *International Classification of Diseases* (10th revision) (World Health Organization) codes C00 to D48 for deaths from 1998-2003. Deaths prior to 1998 were classified using ICD 8th revision (1974-78) (World Health Organization) and 9th revision (1979-1997) (World Health

Organization) codes 140 to 239. A cancer death was classified as such if the primary cause of death or any of the five underlying causes of death were cancer.

Data Mapping

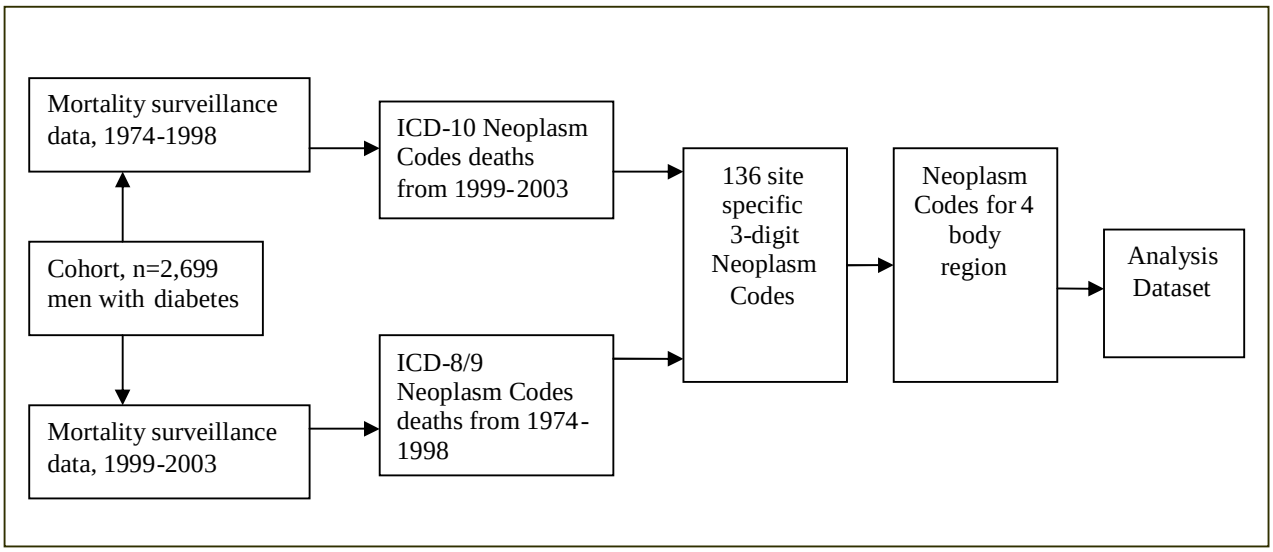
Since the ICD-10 is the most current revision, we determined that it would be most useful to map the neoplasm codes from the older revisions to the ICD-10 categories. Due to the increased level of detail in the ICD-10, it is not possible to directly map the neoplasm codes for the ICD-8/9 to the ICD-10 codes therefore we created a unidirectional map that initially grouped the smallest subgroups into the larger categories using the 3 digit neoplasm site codes employed by the ICD-10 (1 alpha character and 2 numeric characters). Due to the small number of site specific cancer deaths in our cohort, we then regrouped these into four body systems that provided a more useful unit on which to report for our cohort (Table 1).

Table 1. Examples of recoding Neoplasm Codes.

ICD-8/9 Code	ICD-10 Code	Common Code	Examples of Site Specific Cancers
140.0-140.9	C00.0-C00.9	C00	lip
153.0-153.9	C18.0-C18.9	C18	Colon
157.0-157.9	C25.0-C25.9	C25	Pancreas
			Body Regions:
140.0-157.9, 159.0-159.9,	C00.1 –C26.9	C00thruC26	Gastro-Intestinal
160.0-163.1, 164.0-165.9	C30.0-C39.9	C30thruC39	Pulmonary
187.1-189.9	C60.0-C68.9	C60thruC68	Genital-Urinal
200.0-202.9	C81.0-C96.9	C81thruC96	Heamatpoetic

Separate files were created to map ICD8/9 codes and ICD-10 codes to the common codes. A third file contained the code used to subset the male patients with diabetes and to define various descriptive variables that would be used in the analysis and report.

Figure 1.



The following sequence was used to establish the data maps and the final analysis dataset (figure 1):

1. Cohort file: Create file subsetting our cohort of interest from ACLS database containing approximately 190,000 patient records. For patients who had more than one record in the database, the first visit where the patient identified as diabetic was used. Diabetes cases were defined as men who reported taking insulin, had a physician-diagnosed history of diabetes, or had a fasting plasma glucose concentration ≥7.0 mmol/L (≥126 mg/dL). (The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus 1183-97) To increase study

generalizability, individuals with both type 1 and type 2 diabetes were included in the analyses.

2. Mortality Surveillance file: Information from our mortality surveillance was pulled for men in the cohort who had died. It was merged with the cohort file by ID.

3. ICD-9 file: Create file mapping ICD-8/9 neoplasm codes to Common Terms. Since the ICD-9 codes are numeric, the primary tasks associated with mapping these terms were initializing the 136 common terms for site specific cancers and the 4 terms for the body region specific cancers, then regrouping the cancers under the new common terms. The following is an example of the code:

```
if (156.1 <=ICD9CodeNumber1 < 157.0) then C24=1; else
if (157.0 <=ICD9CodeNumber1 < 158.0) then C25=1; else
if (159.0 <=ICD9CodeNumber1 < 160.0) then C26=1; else

if (140.0 <=ICD9CodeNumber1 < 160.0) then C00thruC26=1;
```

4. ICD-10 file: Create file mapping ICD-10 neoplasm codes to Common Terms. Since the ICD-10 codes are alpha-numeric, slightly more manipulation was required to map the cancer codes to the common code. First, the alpha and numeric portions of the ICD codes were divided.

```
DATA work.MedArcMort99to03_a;
  SET work.MedArcMort99to03;
      ICD10CodeLetter1=substr(ICD10code1,1,1);
      ICD10CodeNumb1=substr(ICD10Code1,2,4);
      ICD10CodeLetter2=substr(ICD10code2,1,1);
      ICD10CodeNumb2=substr(ICD10Code2,2,4);
      ICD10CodeLetter3=substr(ICD10code3,1,1);
      ICD10CodeNumb3=substr(ICD10Code3,2,4);
      ICD10CodeLetter4=substr(ICD10code4,1,1);
      ICD10CodeNumb4=substr(ICD10Code4,2,4);
      ICD10CodeLetter5=substr(ICD10code5,1,1);
      ICD10CodeNumb5=substr(ICD10Code5,2,4);
      ICD10CodeLetter6=substr(ICD10code6,1,1);
      ICD10CodeNumb6=substr(ICD10Code6,2,4);
RUN;
```

Next, the 136 common terms for site specific cancers and the 4 terms for the body region specific cancers were initialized. The cancer codes were then regrouped under the common terms. The following is an example of the code:

```
if    ICD10CodeLetter1="C" then do;

if (24 <=ICD10CodeNumber1 <25) then C24=1; else
if (25 <=ICD10CodeNumber1 <26) then C25=1; else
if (26 <=ICD10CodeNumber1 <27) then C26=1; else

if (00 <=ICD9CodeNumber1 < 27) then C00thruC26=1;
```

5. Mortality 1998: Deaths from 1974-1998 were then subset from the Cohort file containing the mortality data. These deaths were then recoded and mapped to the common terms using the ICD-9 file.
6. Mortality 1999 to 2003: Deaths from 1999-2003 were then subset from the Cohort file containing the mortality data. These deaths were then recoded and mapped to the common terms using the ICD-10 file.
7. Concantenate datasets: The Mortality 1998 and 1999to 2003 datasets were combined to create the recoded dataset ready for analysis.

Discussion

Publicly available SAS code to map previous versions of the ICD to the ICD-10 are difficult to find. The CDC has made available SAS code for assigning ICD-9 to ICD-10 for 130 select causes of infant death and 113 selected causes of death. (Minino et al. 2-58;Minino et al. 5-51) The mapping process is tedious and prone to errors of omission or miscoding due to the shear volume of ICD codes. As our cohort spans 30 years and the number of deaths continues to increases, the amount of information we will be able to

analyze will also increase. Creating code to map the mortality data to the most recent revision of the ICD so that it can be run with various subsets of the cohort saves time and ensures consistency as your mortality data is run through the same matrix and classified the same way each time.

Conclusion

This method of mapping meets the need that many studies encounter when working with mortality data that spans multiple versions of the International Classification of Diseases.

Acknowledgments

This study was supported in part by U.S. Public Health Service research grant AG06945 from the National Institute on Aging, Bethesda, Md.

We thank our many participants, Kenneth H. Cooper, MD, for establishing the Aerobics Center Longitudinal Study, the Cooper Clinic physicians and technicians for collecting the baseline data.

Reference List

Anderson, Robert N., Minino, Arialdi M., Hoyert, Donna L., and Rosenberg, Harry M. Comparability of Cause of Death Between ICD-9 and ICD-10:Preliminary Estimates. National Vital Statistics Reprts 49[2], 1-32. 5-18-2001.

Ref Type: Generic

Brock, Anita, Griffiths, Clare, and Rooney, Cleo. The effect of the introduction of ICD-10 on cancer mortality trends in England and Wales. Health Statistics Quarterly 23, 7-17. 2004.

Ref Type: Generic

McBride, Susan, Gilder, Richard, Davis, Robert, and Fenton, Susan. Data Mapping. Journal of AHIMA 77[1], 44-48. 2006.

Ref Type: Generic

Minino, Arialdi M., Anderson, Robert N., Glenn, Donna E., Parr, Gail A., Parsons, Van L., and Kochanek, Kenneth D. A Guide to State Implementation of ICD-10 for Mortality, Part II: Applying Comparability Ratios. Centers for Disease Control and Prevention. 5-51. 12-4-2000.

Ref Type: Generic

Minino, Arialdi M., Anderson, Robert N., Glenn, Donna E., Parr, Gail A., Parsons, Van L., and Kochanek, Kenneth D. Documentation for the Public Use Multiple Cause of Death File on Comparability between ICD-9 and ICD-10: A double-coded file

based on teh 1996 data year multiple cause of death file. Centers for Disease Control and Prevention. 2-58. 12-4-2000.

Ref Type: Generic

The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. "Report of the Expert Committee on the diagnosis and classification of diabetes mellitus." Diabetes Care 20.7 (1997): 1183-97.

World Health Organization. International Classification of Diseases and Related Health Problems, Eighth Revision. Geneva, World Health Organization.

Ref Type: Generic

World Health Organization. International Statistical Classification of Diseases and Related Health Problems, Nineth Revision. Geneva, World Health Organization.

Ref Type: Generic

World Health Organization. International Statistical Classification of Diseases and Related Health Problems, Tenth Revision. 1992. Geneva, World Health Organization.

Ref Type: Generic