

A Macro to Estimate Morbidity Indices

Laurel A. Panser, Janice R. Offord, James M. Naessens, Mary E. Campion,
and Sherine E. Gabriel
Mayo Clinic, Rochester, MN

Abstract

In the 1990s, health care systems are under great pressure to reduce costs, to demonstrate the effectiveness of treatments, and to introduce useful reform. Consequently, health systems are being compared. These comparisons need to fairly reflect the variability in disease burden among populations. Populations with greater disease burden not only use more resources (for example, hospitalizations are longer) but have worse health outcomes (for example, mortality rates are higher). Thus, health systems serving patients with a greater disease burden will appear to perform worse than those serving a healthier clientele. These issues also apply when examining the practice patterns of individual physicians. Differences in disease burden must be adjusted (confounding minimized) to make balanced comparisons, but limited software has been available to measure disease burden.

This manuscript describes the first SASTM macro that facilitates the estimation of the Charlson morbidity indices, which are widely used summaries of disease burden. This macro imports data and computes three Charlson morbidity indices, 17 dichotomous disease scores, and creates tables that facilitate comparisons. The many sources of flexibility are described including the ability to use multiple diagnostic coding schemes.

Introduction

The task of health scientists is to draw simple conclusions even when the clinical issues and the data sets analyzed are complex. To make it easier for scientists to quantify an important complicating issue in outcomes research, morbidity load, Charlson and colleagues developed indices that measure disease burden (Charlson 1987). Researchers may use one of these indices as an independent variable, as a factor that modifies other causes, or to achieve greater balance when contrasting groups (less confounding). However, software has been unavailable to facilitate the estimation of the Charlson morbidity indices. A SASTM macro was developed that estimates three Charlson morbidity indices and 17 indicators of the presence or absence of a given disease. This paper describes the macro.

Strategy used and characteristics of the macro

A straightforward strategy was used to create the Charlson macro. We adopted assumptions and definitions of Charlson and colleagues, almost in whole (Charlson 1987). Next, we used a modification of Charlson's work created by Deyo et al. (Deyo 1992). Finally, we embellished upon the work of Charlson and Deyo.

We adopted the assumptions and definitions of Charlson and colleagues. Charlson made the task of estimating the total disease burden tractable by identifying key diseases. Table 1 shows the 17 morbidities from Charlson's original work that were adopted by Deyo. By definition, only diseases shown in Table 1 are considered in the estimation of the indices (Table 2). Charlson derived disease weights for the two original indices (CH_INDEX and CH_CMORT) by using relative risk estimates among a cohort of inpatients. Additionally, she created age-specific weights (Table 3) for the estimation of one index (CH_CMORT).

The three Charlson indices estimated by the macro reflect varying levels of complexity. The indices measure the risk of death (Charlson 1987, D'Hoore 1993). The total number of Charlson diseases is counted in an index we added (CH_COUNT). The morbidity-weighted disease index is captured in the variable CH_INDEX. The third index (CH_CMORT) is a composite of weights from 17 morbidities and age.

Deyo's adaptation of Charlson's work was incorporated into this software. Charlson originally developed her comorbidity indices to predict one-year post-hospitalization mortality by using data abstracted from medical records. Deyo wanted to apply Charlson's indices to patient encounters captured in administrative data in a study of lumbar spine surgery. Administrative data such as billing records, are typically coded according to the ninth edition of the clinical modification of the International Classification of Diseases (ICD-9-CM) (Puckett 1995). Consequently, Deyo operationalized Charlson's definitions by mapping them to ICD-9-CM.

The macro was designed to handle three types of diagnostic codes including ICD-9-CM codes. Deyo's ICD-9-CM definitions were mapped to Mayo Clinic's detailed version of a general coding system, the Hospital Adaptation of the International Classification of Diseases (H-ICDA) (Commission on Professional and Hospital Activities 1973). Mayo Clinic's medical index is coded by using the Mayo modification of H-ICDA. The macro computes 17 dichotomous disease indicators given ICD-9-CM codes, Mayo modified H-ICDA codes, or both. Additionally, the macro processes dichotomous codes for the 17 diseases from any source (such as medical history abstracts or questionnaires) to estimate the three indices.

Accessing and using the macro

Users may invoke the macro repeatedly to do simple tasks or to do more complex procedures. The call statement for the macro is described first. The execution of simple tasks are explained next. Later, less simple procedures are described.

The macro is submitted by using the following call statement with some combination of the parameters below described below and in the appendix:

```
%MACRO CHARLSON (PROCESS= DXDATA= ,PTDATA= ,DXTYPE= ,  
DXARRAY= ,HITS= ,OUT_TIME= ,TEXT= ,PRINT= ,ID= ,  
DATE= ,AGE= ,DXDATE= ,ICD9= ,HICDA= ,MOD= ,SYN= ,  
BY= );  
RUN;
```

Simple uses of the macro

A primary input data set containing patient or study subject information (PTDATA) must be made available to the macro. It must contain one record per subject. PTDATA must include the subject identification variable (ID) used to link this data set to the other data sets (DXDATA, HITS and OUT_TIME) handled by the macro. If the age-weighted index (CH_CMORT) will be added to the primary data set later, then the subject's age (AGE) and reference date (DATE), such as a date of a surgery, must be present.

In addition to providing primary patient data, the user may enter a diagnostic data set (DXDATA) from billing records or another source coded in ICD-9-CM, Mayo-modified H-ICDA, or both. The diagnostic data set contains one record per diagnosis for each subject on a given date. It is preferable for the diagnostic data to include only subjects in the primary data set. The diagnostic code type (DXTYPE) must be specified. As an example, if both complex coding schemes are used, then DXTYPE= 3. Under this

scenario, the variable containing ICD-9-CM data (ICD9) and the three fields containing H-ICDA data (HICDA, MOD and SYN) must be specified. The variable HICDA contains the unmodified H-ICDA code created by the Commission on Professional and Hospital Activities. The variable MOD contains the Mayo modifier used to differentiate subclassifications of diagnoses within H-ICDA codes. References to synonymous terms are found in the variable SYN, another Mayo Clinic embellishment. Additionally, whenever a diagnostic data set is used, the dates associated with diagnoses (DXDATE) must be defined.

All of the uses of the macro may be accomplished by setting PROCESS equal to 6. Data processing occurs in the following order. First ICD-9-CM and Mayo-modified H-ICDA codes are condensed to those resulting in a 'yes' or 'hit' on one of the 17 diseases. The results are stored in an output data set (HITS) with an additional variable (_DX_) indicating which of the 17 diseases were flagged. Additionally, investigators may output a data set (OUT_TIME) to review the dates of the earliest (FRST_DT) and latest (LAST_DT) diagnosis dates in the DXDATA for each person in PTDATA. FRST_DT and LAST_DT will be blank if a subject has no diagnostic information. After the HITS dataset has been created, it is used as input to add the array (DXARRAY) to the master input data set (PTDATA). The array contains the variables DX1 through DX17 (Table 1) that specify the 17 disease flags necessary for the computation of the three indices. If a disease is not coded as present (1=yes), the disease is assumed to be absent (0=no). Missing diagnostic data are not distinguished from diagnoses known to be absent. The final step in the process is to use the diagnostic array to compute the three indices (CH_COUNT, CH_INDEX, CH_CMORT) and add them to the PTDATA data set.

Users may also print results by using the PRINT option. To facilitate the review of individual ICD-9-CM and Mayo-modified H-ICDA codes, the user may use the parameter TEXT= Y. The macro will then add detailed text descriptions for the ICD-9-CM and Mayo-modified H-ICDA codes to the diagnostic data set. The user may also request frequency distributions of the individual ICD-9-CM and Mayo-modified HICDA codes within the 17 morbidity categories, the frequency of the 17 morbidities, and a univariate distribution of the three indices. The user may specify a BY group when printing these frequency and univariate distributions.

Additional advantages and less simple procedures

In addition to the advantages already enumerated, the macro may be used to cull inappropriate subjects from a given study and to combine morbidity data from multiple sources to estimate the 17 diseases and the three indices. Users may print tables to identify subjects who should be removed from a study when an entire Charlson morbidity defines an exposure in a cohort study, an outcome in a case-control study, or another exclusion criteria. Additionally, the PRINT option with TEXT= Y may be used to identify and exclude subjects with disorders under the rubric of one of the 17 morbidities (such as men with prostate cancer among subjects with cancer histories). Apart from exclusions, stepwise data processing may be used to prioritize the retention of diagnoses from one data source among other data sources accessed. For example, recent administrative data may provide information on recent events while a search of the same data over a longer time period would ensure better representation of chronic conditions. In this way, users may prioritize the retention of data on a disease-by-disease basis to more accurately portray the acuity versus chronicity of the conditions. Prioritization may also be used when morbidity data from one source is more valid than other sources; as an illustration, investigators may choose to retain a diagnosis of metastatic cancer from a tumor registry and dismiss questionnaire data from a subject acknowledging a history of malignancy without mention of metastasis.

Table 1. The Seventeen Morbidities Considered in the Estimation of the Three Charlson Indices and the Disease-Specific Weights Applied to Two of Those Indices*

Morbidity	Weights*	Variable Name†
Myocardial infarction	1	DX1
Congestive heart failure	1	DX2
Peripheral vascular disease	1	DX3
Cerebrovascular disease	1	DX4
Dementia	1	DX5
Chronic pulmonary disease	1	DX6
Peptic ulcer	1	DX7
Liver disease, mild	1	DX8
Diabetes mellitus without end organ damage	1	DX9
Diabetes mellitus with end organ damage	2	DX10
Hemiplegia	2	DX11
Renal disease, moderate or severe	2	DX12
Liver disease, moderate or severe	2	DX13
Cancer, metastatic solid tumors	6	DX14
Acquired immune deficiency syndrome	6	DX15
Rheumatologic disease	1	DX16
Cancer, non-metastatic	1	DX17

*The weights were derived from relative risk estimates. The weights are applied to an index that only takes disease into account (CH_INDEX) and another index that summarizes both disease and age (CH_CMORT).

†The name of variables in the Charlson macro are shown. Dichotomous (present/absent) variables are computed for each of these 17 morbidities.

Table 2. The Three Charlson Comorbidity Indices Estimated By the Charlson Macro

Index*	Description of Index
CH_COUNT	A simple count of the number of Charlson morbidities. Assumes all diseases are equally burdensome. Age is ignored. Range: 0 - 17.
CH_INDEX	Charlson morbidities are weighted to account for varying degrees of burden. Age is ignored. Range: 0 - 33.
CH_CMORT	Charlson morbidities are weighted to account for varying degrees of burden and the index is incremented for advancing age. For each decade of age after age 49, 1 point is added to the index. Range: 0 - 39.

*Name of the Charlson index within the Charlson macro.

Table 3. Age Weights Added to One Charlson Comorbidity Index Labeled CH_MORT* in the Charlson Macro

Age†	Age-Specific Weights Added to the Score
< 50	0
50-59	1
60-69	2
70-79	3
80-89	4
90-99	5
>=100	6

A comorbidity index that is weighted according to the burden of disease and age.

†Age in years

Discussion

There are things to consider when using this macro. It may be difficult to know if a disease being adjusted for (a confounder) occurred before the outcome of interest in some administrative data sets. This is a concern when using Mayo Clinic's medical index data before 1994; diagnoses were captured within windows of approximately one year (registration brackets) and not linked to dates of diagnoses. Other considerations pertain to the indices. Not all diseases affecting the quality of life are represented in the indices. Also, the indices capture multiple levels of severity for only two morbidities (liver disease and diabetes mellitus). Additionally, the disease weights reflect the one-year risk of death and may be misleading for some outcomes.

However, the Charlson indices have been shown to be indicators of outcomes other than death (Katz 1996) including discharge to nursing homes and other rehabilitation facilities, complication rates, length of stay, functional status, surgical results and the use of treatments. Moreover, a questionnaire was developed that translated the spirit of Charlson's medical history abstraction process into lay language (Katz 1996). The questionnaire yielded valid results, as measured against medical record review data (Spearman correlation coefficient 0.63). Twenty-four hour, test-re-test reliability was also high for the questionnaire (intraclass correlation coefficient 0.91). Thus, this macro may be used to estimate the Charlson indices from medical history abstracts, administrative and questionnaire data.

Summary

This macro is an easy to implement, highly functional advancement that may be used to minimize patient disease burden in comparisons among groups.

Acknowledgments

The development of this macro was made possible through funding, in part, by a grant from the Public Health Services, National Institutes of Health (AR30582).

References

- Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chron Dis* 1987;40:373-383.
- Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *J Clin Epidemiol* 1992;45(6):613-619.
- D'Hoore W, Sicotte C, Tilquin C. Risk adjustment in outcome assessment: the Charlson comorbidity index. *Meth Inform Med* 1993;32:382-387.
- H-ICDA: Hospital Adaptation of ICDA. 2nd Edition. Vol. 1. Ann Arbor, Michigan: Commission on Professional and Hospital Activities, 1973.
- Craig D. Puckett. *The Educational Annotation of ICD-9-CM*. 4th Edition Volumes 1-3 Reno, Nevada: Channel Publishing, Ltd., 1995.
- Katz NJ, Change LC, Sangria O, Fossel AH, Bates DB. Can comorbidity be measured by questionnaire rather than medical record review? *Medical Care* 1996;34(1):73-84.

Appendix

This appendix contains documentation for the main macro parameters. In addition, variables are present in the input data sets, such as ID.

The main parameters needed to run the macro are:

1. PROCESS = The main processing options: (required)
 - 1 = creates a dataset of the records from DXDATA resulting in yes for one of the 17 morbidities and adds the variable _DX_ to each record.
 - 2 = creates an array of 17 morbidity flags (DXARRAY) from the HITS data set and adds this array to PTDATA.
 - 3 = computes the 3 Charlson indices for each subject from data in DXARRAY and adds them to PTDATA.
 - 4 = 1 and 2 above.
 - 5 = 2 and 3 above.
 - 6 = 1,2,3 above.
2. PTDATA = The name of the patient dataset.
3. DXDATA = The name of the diagnostic dataset.
4. DXTYPE = The type of diagnostic coding used on the diagnostic dataset. The options are:
 - 1= IDC-9-CM
 - 2= Mayo-modified HICDA
 - 3= Both of the above
5. HITS = The name of the output dataset to be created containing the records from DXDATA resulting in one of the 17 morbidities.
6. OUT_TIME = The name of the output dataset to be created containing the time span of the Diagnostic Dataset for each patient. It will not be created unless this parameter is specified and both PTDATA and DXDATA are present.
7. PRINT = The printing options:
 - 1 = print frequencies of diagnostic codes found in the HITS data set. Detailed text descriptions of the diagnoses are printed with the code, if TEXT=Y has been specified.
 - 2 = print frequency of the 17 morbidity flags.
 - 3 = print univariate distributions of the 3 indices
 - 4 = 1 and 2
 - 5 = 2 and 3
 - 6 = 1, 2, 3
8. TEXT = Flag to indicate if text descriptions for the diagnostic codes are to be added to the DXDATA dataset. Y=add, otherwise do not add text.

Additional documentation will be included in a technical report on the Charlson indices from the Section of Biostatistics, Department of Health Sciences Research, Mayo Clinic and Mayo Foundation, Rochester, MN 55905.

For further information, please contact Laurel A. Panser or Janice R. Offord, Mayo Clinic, 200 First Street Southwest, Rochester, Minnesota 55901.

Laurel A. Panser
e-mail: panser.laurel@mayo.edu
Telephone: 507-284-1438
Fax: 507-284-1731

Janice R. Offord
e-mail: offordj@mayo.edu
Telephone: 507-284-5630
fax: 507-284-9542

SAS is a registered trademark or trademark of SAS Institute, Inc. in the USA and other countries. SASTM indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.