Chapter 5
The Charlson Comorbidity Index

INTRODUCTION

In this chapter, we consider the Charlson Comorbidity Index (CCI). This index is published, and the weights used to define risk adjustment in the logistic regression model are clearly identified as well. (Sundararajan et al., 2004) Therefore, we are able to examine this index in detail, and to see if the index is meaningful in terms of adjusting risk based upon patient condition. We will suggest an alternative to the Charlson Index in Chapter 8 that gives improvements in terms of its relationship to outcomes.

In Chapters 6 and 7, we will discuss other commonly used severity index measures. We will compare their results to those computed using the Charlson Index. We will also continue to discuss the examples given in Chapter 4 of patients having cardiovascular bypass surgery and COPD (chronic obstructive pulmonary disease) in addition to other examples introduced in this chapter.

Since the Charlson Index is published, it is not used by payers to rank the quality of providers. If it were used, the providers could focus on this minimal set of patient conditions. There is, of course, a similar danger if the developers of proprietary indices released, or sold, their model to healthcare providers who could use the values to focus on specific diagnosis codes. Therefore, in order to reduce the problem of upcoding, providers should not know the details of the model used.

DOI: 10.4018/978-1-60566-752-2.ch005
BACKGROUND

The Charlson Index was developed to provide risk adjustment and to predict patient mortality or other outcomes in the hospital. (Charlson, Pompei, Ales, & MacKenzie, 1987) It was tested and validated on a general population. (Romano, Roos, & Jollis, 1993a, 1993b) Other data have modified the Charlson Index for specific patient populations. (Ghali, Hall, Rosen, Ash, & Moskowitz, 1996) It is representative of many different indices. (Iezzoni, 2003)

The problems with the Charlson Index show some of the problems generally with risk adjustment measures. (Melfi, Holleman, Arthur, & Katz, 1995) In any index that limits the choice of the diseases used to define the index, there will be omitted diseases that can have as much importance in relationship to patient outcomes as the ones in the model. As a result, the indices tend to be poor predictors of actual outcomes.

One of the problems with using the Charlson Index is that information is not uniformly entered by healthcare providers. In one study, hospitals consistently under-coded for diabetes and for diabetes with complications compared to physicians. In contrast, physicians tended to under-code on myocardial infarction and cerebral vascular disease. (Klabunde, Potosky, Legler, & Warren, 2000) Similarly, there is under-coding in the billing data, usually discovered through a review of patient charts. (Doorn et al., 2001; Kieszak, Flanders, Kosinski, Shipp, & Karp, 1999) Under-coded patients will show a reduced risk compared to what they should normally be assigned, and their predicted risk will be reduced as well.

Since providers are measured in terms of quality based upon the difference between predicted outcomes and actual outcomes, providers that have a high proportion of patients with the diseases listed in the index will have a greater difference between predicted and actual values compared to providers that have patients with severe problems that are not in the index list. A different list of diseases can completely change the risk adjustment and the rank order of providers. We will demonstrate how such a different list compares to the Charlson Index.

STATISTICAL EXAMINATION OF THE CHARLSON INDEX

Table 1 gives the patient condition, the related ICD9 codes, and the associated weights for the Charlson Index. (Sundararajan et al., 2004)

The score is the sum of the weights of the nonzero indicators of categories 1-17 in Table 1. Other references give a weight of 3 to metastatic cancer as well in spite of the high risk of mortality. (Gettman et al., 2003) For our purposes, the actual weights are not important; the methodology remains the same regardless of the assigned weights.

However, there is a clear problem of multicollinearity given the relationship between diabetes and diabetes with complications, and so on. Consider, for example, diabetes with complications. This can include diabetes if both 2501 and 2504 are listed as secondary patient conditions. Approximately 9% of those with complications of diabetes are also identified as having diabetes. A higher 14% of those with diabetes complications also have renal disease. To avoid the issue of collinearity, only the score is used in a logistic regression to examine mortality, or in a linear regression to examine costs and utilization.

First, we define the Charlson Index on the 2005 National Inpatient Sample using the SAS code listed below. We use the SUBSTR function to reduce the ICD9 codes to 3-digits. Next, we concatenate the 15 columns of ICD9 codes into one column. Then, the rxmatch function is used to find the relevant codes for
Related Content

A Query Beehive Algorithm for Data Warehouse Buffer Management and Query Scheduling

ASCCN: Arbitrary Shaped Clustering Method with Compatible Nucleoids
www.irma-international.org/article/asccn-arbitrary-shaped-clustering-method/46940/

Mining Customer Knowledge for a Recommendation System in Convenience Stores
Shu-Hsien Liao, Chih-Hao Wen, Pei-Yuan Hsian, Chien-Wen Li and Che-Wei Hsu (2014). *International Journal of Data Warehousing and Mining* (pp. 55-86).
www.irma-international.org/article/mining-customer-knowledge-for-a-recommendation-system-in-convenience-stores/110386/

A Comparative Study on Medical Diagnosis Using Predictive Data Mining: A Case Study
Seyed Jalaleddin Mousavirad and Hossein Ebrahimpour-Komleh (2014). *Data Mining and Analysis in the Engineering Field* (pp. 327-360).
www.irma-international.org/chapter/a-comparative-study-on-medical-diagnosis-using-predictive-data-mining/109989/

Estimating Semi-Parametric Missing Values with Iterative Imputation
Shichao Zhang (2010). *International Journal of Data Warehousing and Mining* (pp. 1-10).
www.irma-international.org/article/estimating-semi-parametric-missing-values/44955/