Chapter 11
Compression of Diagnosis and Procedure Codes

INTRODUCTION

Each of the datasets has many different diagnosis and procedure codes to represent a patient’s condition. There are thousands of potential codes, and millions of potential combinations of codes. In order to use patient diagnosis and procedure information in statistical models, there has to be some form of compression as there are far too many to include all of them. Therefore, there has to be some method to compress codes. While such methods are discussed in detail in Cerrito (2009), they will be discussed briefly here.

Information codes are used in billing and administrative data to define patient conditions, and also to define patient treatments. These codes are used to define patient severity indices. Therefore, it is absolutely essential to both understanding the severity indices, and to defining such severity indices to be able to work with these codes. The most difficult data to work with are contained within claims databases where different coding methods are used by different providers; the different codes must be reconciled in some manner.

The standard method to define a patient severity index is to use a number of patient demographics and diagnoses in a linear or logistic regression, comparing patient information to outcomes. Unfortunately, the only way to determine reliability is by using fresh data. There is no good way to validate the model. We will demonstrate several types of patient severity indices to show the current state of the science. Attempts are made to validate by comparing several methods to each other; however, different methods

DOI: 10.4018/978-1-61520-905-7.ch011
Compression of Diagnosis and Procedure Codes

can lead to very different outcomes. In addition to defining a severity index, text analysis can be used to examine how patient conditions interact, and how they relate to treatments.

BACKGROUND

There are a number of different approaches to compressing patient diagnosis and procedure codes. One of the most common is to decide upon specific inclusion/exclusion criteria and then to extract patients who have those specific criteria. (Bateman, Simpson, Bateman, & Simpson, 2006; Glance, et al., 2009; Goff, et al., 2007; Mountford, et al., 2007; Olsen, et al., 2008) Another, as discussed in previous chapters, is to find the most frequently occurring codes in a subpopulation of patients, and to use them in the analysis. While electronic medical records can make text analysis easier to perform to investigate health outcomes, emphasis continues to be on the benefits to the providers, and also on the standardization of language. (Doyle & Doyle, 2006; Jaspers, Knaup, Schmidt, & Jaspers, 2006; Shaw & Shaw, 2006; Yamamoto, Khan, Yamamoto, & Khan, 2006)

Another common method of compression is to define a patient severity score. (Burd, et al., 2008; Chung, Krishnan, & Chakravarty, 2007; Kuykendall, Ashton, Johnson, & Geraci, 1995; Ricciardi, et al., 2007; Rutledge, et al., 1997; West, Rivara, Cummings, Jurkovich, & Maier, 2000) There are several different methods available that define a patient score. Typically, a number of diagnosis codes are used to define the score, and patient information regarding these diagnoses are extracted in the same way that inclusion criteria are used. One common method is called the Charlson Index.

Still other types of compression are proprietary, and often rely upon physician consensus panels to define a patient risk assessment. A novel method discussed in detail in Cerrito (2009) and Cerrito (2007) is to use text analysis to define patient clusters of conditions. (P. B. Cerrito & Cerrito, 2008)

CODED INFORMATION

Information concerning the patient condition is coded into the datasets using a variety of systems. There are a number of coding systems that are used in administrative data. The first is the DRG, or diagnosis related group. These codes are used by Medicare and other insurance providers to provide reimbursements. Generally, a value is negotiated between the insurer and the provider for a particular DRG. Then, DRG grouper software uses “the principal diagnosis, secondary diagnoses (as defined using ICD9 codes), surgical procedures, age, sex and discharge status of the patients treated” to assign inpatient records to a specific DRG. (Anonymous-DRG, 2008b). Examples of DRG codes are listed below: (Anonymous-DRG, 2008a)

- 424 Operating room procedure with principal diagnoses of mental illness
- 425 Acute adjustment reaction & psycho-social dysfunction
- 426 Depressive neuroses
- 427 Neuroses except depressive
- 428 Disorders of personality & impulse control
- 429 Organic disturbances & mental retardation
- 430 Psychoses
- 431 Childhood mental disorders
- 432 Other mental disorder diagnoses
- 433 Alcohol/drug abuse or dependence, left, against medical advice
- 434 Alcohol/drug abuse or dependence, detoxification or other symptom treatment with complications, comorbidities
- 435 Alcohol/drug abuse or dependence, detoxification or other symptom treatment without complications, comorbidities