Chapter 9
Working from Claims Data

INTRODUCTION

Claims data are more difficult to work with to extract the necessary information about patient conditions in relationship to costs. There can be multiple claims for the same patient episode from different sources. A physician visit after an inpatient claim can be followed up for the inpatient stay rather than to consider the inpatient stay as the start of a new patient episode or a new patient problem. Therefore, in addition to analyzing patient conditions as represented by ICD9 codes, we must also attempt to define an episode and to distinguish between new problems and follow up of old problems.

The second major difference in using claims data as opposed to the data we have been using up to this point is that there exists a one-to-many relationship in the data that must be taken care of. For example, one inpatient stay may have several follow up visits, and one or two of those visits may be classified as outpatient stays while others are classified as physician visits. We need to find a way to convert the one-to-many relationship to a one-to-one so that we can investigate outcomes. Another difference is that not all medical services are necessarily allowable, so that the medical record may be incomplete. For example, until recently, Medicare disallowed almost all medications. Therefore, such information

DOI: 10.4018/978-1-60566-752-2.ch009
would not be recoverable. However, inpatient medications are usually not identified separately in the billing record, so such information would still remain incomplete.

Moreover, although hospital providers tend to use ICD9 codes, physicians are more likely to use the CPT coding system. Therefore, in addition to defining episodes, it is necessary to be able to use multiple coding systems. In this chapter, we will discuss the means of working with such data to determine patient severity. Because of the nature of text mining, it is very well equipped to use multiple coding systems since it mostly relies upon the linkage between codes instead of relying on a specific list of codes. It is the linkage between codes that is used in text mining to define patient clusters.

BACKGROUND

Claims data are used to identify general trends concerning patient treatment. (Smith- Bindman, Quale, Chu, Rosenberg, & Kerlikowske, 2006; Wilkinson, Askew, & Dixon, 2006) Claims data are also used to make cross-references to different treatments. In Ceratti, Roger France, and Beguin, claims information was compared to the hospital clinical database to determine if treatment and diagnosis were related, or whether treatment was absent given a diagnosis. (Ceratti, France, & Beguin, 2008) Such observational studies relating billing data to clinical outcomes are fairly common. (Kaushal, Bates, Franz, Soukup, & Rothschild, 2007) Billing charges for prophylactic medication can also be compared to average costs of disease treatment. (Collinet- Adler et al., 2007)

Another study cross-referenced billing data from general practice to that of hospital emergency departments to see if patient visits for infection in the general practice could predict near term increases in emergency room utilization for similar infections. (Sloane et al., 2006) Another use of claims data has been to determine whether appropriate testing is conducted for patients with chronic illnesses. For example, claims are used to determine if patients with diabetes are administered regular A1C tests or to determine if patients with heart conditions are prescribed an ace inhibitor. (Philipneri et al., 2008)

Since multiple providers treat the same patients, the development of a patient severity index will not be as useful to rank the quality of providers since it will become difficult to separate the contribution of each provider. Instead, it can be used to find patterns of treatment and to find those treatment patterns that lead to the best patient outcomes. It can be used to find the relationship between different treatments and different outcomes for the same patient conditions.

First, the data have to be examined to define separate episodes, and to investigate episodes in relationship to patient outcomes. Solutions under the general category of episode grouper have been developed specifically to fuse claims data. The methodology is difficult to find since it is mostly proprietary, and little exists in the research literature. (P. Claus, P. Carpenter, C. Chute, D. Mohr, & P. Gibbons, 1997; Forthman, Dove, & Wooster, 2000; Rosen & Mayer-Oakes, 1999) A brief summary is given in Forthman, Dove and Wooster. (Forthman et al., 2000) The main purpose of these groupers is to identify homogeneous groups of patients so that cost comparisons and summaries can be made. These “episode groupers” are used in analysis with little understanding as to how episodes are defined or how patients are grouped. (Bassin, 1999; Currie et al., 2005; Kerr, McGlynn, Vorst, & Wickstrom, 2000; Thomas, 2005; Wan, Crown, Berndt, Finkelstein, & Ling, 2002) However, it is known that the groupers do not take into consideration the severity of an individual patient’s condition. (Thomas, 2005)

One method of grouping is to examine medications of a similar nature, and to define the end of an episode if there is at least one day between prescriptions. (Bonetto, Nose, & Barbui, 2006) The Medicare
Related Content

PSSRC: A Web Service Registration Cloud Based on Structured P2P and Semantics
Qian He, Baokang Zhao, Liang Chang, Jinshu Su and Ilsun You (2016). *International Journal of Data Warehousing and Mining* (pp. 21-38).
[www.igi-global.com/article/pssrc/146851?camid=4v1a](www.igi-global.com/article/pssrc/146851?camid=4v1a)

Identify Cross-Selling Opportunities via Hybrid Classifier
[www.igi-global.com/article/identify-cross-selling-opportunities-via/1807?camid=4v1a](www.igi-global.com/article/identify-cross-selling-opportunities-via/1807?camid=4v1a)

Data Mining Using Fuzzy Decision Trees: An Exposition from a Study of Public Services Strategy in the USA
[www.igi-global.com/chapter/data-mining-using-fuzzy-decision/44282?camid=4v1a](www.igi-global.com/chapter/data-mining-using-fuzzy-decision/44282?camid=4v1a)

A Survey of Extract–Transform–Load Technology
[www.igi-global.com/article/survey-extract-transform-load-technology/3894?camid=4v1a](www.igi-global.com/article/survey-extract-transform-load-technology/3894?camid=4v1a)