Chapter 6
Creating a One-to-One Relationship in the Data from a Many-to-Many

INTRODUCTION

One of the biggest problems in dealing with healthcare data is that there can be multiple patient events for the same individual. Without finding some way to combine these many observations into one row in the dataset, it is very difficult to analyze the data by patient. When dealing with cost and reimbursement information, this can be relatively simple since a total sum of costs or reimbursements can be used to examine any one patient. However, if we are looking at patient conditions, or physician decisions in sequence, this consolidation of patient information can be especially difficult. When using the technique of survival data mining, as we will discuss in later chapters, such consolidation of information is absolutely essential. In this chapter, then, we will discuss some of the requirements for combining patient information into one data row.

We will first look at converting a one-to-many relationship into one-to-one and then examine a many-to-many. Suppose a patient has a chronic condition and visits a physician once a month. Then there are a total of 12 physician visits. If we want to investigate the total costs for a chronic condition, we need to find some way to combine these twelve visits into one patient record. In addition, suppose that patient had one inpatient stay under the care of a second physician, with several subsequent follow up visits. Then these visits, too, need to be combined with the regular physician visits into one patient record. The patient may have numerous prescriptions that need to be considered along with the rest of the healthcare contacts. Some
of these prescriptions may be related to the patient’s chronic condition; others may be needed to treat acute infections. As patients will not always respond to one medication, sometimes there are several medications for the same chronic conditions; medications can be prescribed together, or in sequence as the physician attempts to find one that satisfies the patient’s needs. These, then are the one-to-many, with the many defined as the different patient contacts with healthcare; the one is the individual patient.

**BACKGROUND**

A search of Medline using the keywords “preprocessing data” and “health outcomes” returns exactly two results. There are a few textbooks on the market as well that are focused on health outcomes. (Arslanian, 2001; Fink, 2004; Kane, 2005) However, these books tend to be more concerned with devising measurement indices than they are in examining the data analysis and preprocessing requirements. In particularly, there is very little basic information on converting a many-to-one or a many-to-many relationship into a relationship that is one-to-one so that it can be used for the analysis of outcomes. (Austin & Austin, 2008a; Dilba, Bretz, Guiard, & Hothorn, 2004; Hothorn & Hothorn, 2006; Nakagawa & Nakagawa, 2005; Schaarschmidt, et al., 2009)

**Preprocessing Data to Examine Costs**

One of the most important reasons to convert a one-to-many relationship is to investigate the issue of costs. For example, suppose we want to know the cost of congestive heart failure or diabetes. We first go into the dataset and filter using ICD9, CPT, or DRG codes to find those patients, as discussed in previous chapters. Next, we find all claims related to these patients. Next, we create summaries by patient, defining a one-to-one relationship from a one-to-many.

One of the most important uses is in the area of cost effectiveness analysis. If we look at some of these studies, we see that many of the costs are estimated rather than computed from the data. For example, suppose we wanted to examine the cost-effectiveness of a drug-eluting stent for blocked arteries. In order to do this, we first have to find the actual cost of the treatment compared to alternate treatments. We have to determine whether we are going to investigate follow up cost as well as the cost of failure that results in a second procedure. We need to examine the data carefully to make certain that we do not confuse a follow up visit with a repeat procedure because they are coded similarly.

Relatively recently, a drug eluting stent was introduced in the performance of angioplasty procedures. It contains a timed-release medication that reduces the likelihood of a clot developing around the stent, reducing the risk of a second procedure. With the use of this stent, it is possible that fewer bypass surgeries are performed. (De Labriolle, et al., 2009; Kornowski & Kornowski, 2009; Thompson, et al., 2009) In addition, we need to see if the use of the drug eluting stent results in changes to actual treatment. If we just compare the cost of using the drug eluting stent to the cost of the traditional bare stent in treatment, it is clear that the drug eluting stent is more expensive. In addition, we would also need to examine the need for repeat treatments because the stent failed. Is failure more likely with the drug eluting stent, or the bare stent? However, if some patients who would ordinarily receive coronary artery bypass surgery (CABG) now receive percutaneous coronary angioplasty (PTCA), then there is an obvious and considerable cost savings. In order to investigate the true cost, it must be in relationship to an examination of the data to determine if this new device results in a shift in treatment, and the corresponding change in costs. (Kashan, 2010) Long term outcomes should also be investigated; these can be different even
Related Content

The Effectiveness of Health Informatics
www.igi-global.com/chapter/effectiveness-health-informatics/53575?camid=4v1a

Trust and Clinical Information Systems
www.igi-global.com/chapter/trust-clinical-information-systems/53689?camid=4v1a

Prototype Based Classification in Bioinformatics
www.igi-global.com/chapter/prototype-based-classification-bioinformatics/53603?camid=4v1a

Evaluation of Medical Image Compression
www.igi-global.com/chapter/evaluation-medical-image-compression/19611?camid=4v1a