Chapter 4
Preprocessing Medpar Data

INTRODUCTION

Medicare data provide information that hospitals submit for billing purposes, Medpar, or Medicare Provider Analysis and Review. It is publicly available (for a fee) at http://www.cms.hhs.gov/Limited-DataSets/02_MEDPARLDSHospitalNational.asp. There are multiple forms in Medicare data and we provide the SAS code on how to “unpack” the different forms for use in analysis. We are using the 2005 version of the data. It can be provided for one or several providers. We use Medpar data from a Wound Care Center to investigate the treatment of diabetic foot ulcers. We want to determine how such patients are treated, especially those with infections. The Medicare population is at highest risk for such problems. Using a higher risk population means that there will be more patients with what are, essentially, rare occurrences of a disease. As stated in the Chapter 3, MEPS data is insufficient to examine rare occurrences, so we need to use additional data for such problems.

BACKGROUND

Medpar data have not been used as frequently as other datasets to investigate health outcomes. (Cornelius, Feldman, Marsteller, & Liu, 1994; Ellis & Dushman-Ellis, 2000; Isaac, Jha, Isaac, & Jha, 2008; Preminger, et al., 2008; Sharma, et al., 2009; Vitale, et al., 1999; Wei, Mark, Hartz, & Campbell, 1995; Welke, et al., 2007) It has been used to provide dollar values for comparative effectiveness analysis.

DOI: 10.4018/978-1-61520-905-7.ch004
One of the big questions is how data restricted to the Medicare population can be generalized to the rest of the population. (Ash, et al., 2003; Carey, et al., 2008; Needleman, et al., 2003; Stringham, Young, Stringham, & Young, 2005; Welke, et al., 2007) While it probably cannot be so generalized since the population within Medpar is older than the general population, it can serve to focus on medical problems that can occur in higher proportion with an older population. In particular, it can be used to examine expenditures toward the end of life, especially with diagnoses of terminal diseases such as cancer. (Anderson, et al., 2007; Ash, et al., 2003; Barnato, et al., 2004; Haller, Gessert, Haller, & Gessert, 2007) Therefore, Medpar data are frequently used to examine cost-effectiveness, or comparative-effectiveness. (Birkmeyer, Lucas, & Wennberg, 1999; S. T. Fleming, 1995; McCandless & McCandless, 2002; Preminger, et al., 2008) In particular, one such study examined the speculative cost savings by requiring surgery to be performed in regional centers. (Ellis & Dushman-Ellis, 2000) Another study examined the cost-benefits of “do not resuscitate orders”, suggesting that a lack of such orders significantly increased costs. (Haller, et al., 2007) Other studies have focused on the added cost of complications and comorbidities for patients with specific conditions. (Bond, Raehl, & Raehl, 2006; Reynolds, et al., 2006) One very interesting study examined the results of cost studies when comparing Medpar to a second database, showing inconsistencies. (N. A. Halpern, et al., 2007)

The Medpar datasets have also been used to examine the definition of patient severity indices, and to compare different indices. (Cornelius, et al., 1994; Englesbe, et al., 2009; Needleman, et al., 2003; Radley, et al., 2008) Such indices will be discussed in considerable detail in Chapter 10. They have been used to examine geographic differences and healthcare disparities for the Medicare population. (Barnato, et al., 2004; Culler, et al., 2008; D. B. Smith, 1998; Vitale, et al., 1999) We will demonstrate the effective use of Medpar data.

**PREPROCESSING MEDPAR DATA**

Medpar data generally come in one dataset, containing several different sheets of data. In order to use the data, it must first be unpacked, and each “page” of data stored in a different dataset. Once that is completed, the different datasets can be merged. Each page of data is given a code, and these codes can be used to unpack the data.

The first data page, identified as “20,” contains general patient demographic information, and basic admission information. It also contains patient address information, but that has been removed from the de-identified dataset. The next data page, “50” contains information about accommodations and accommodations revenue. It indicates whether the patient has a private or semi-private room. Data page “60” contains information about ancillary services. This can include patient labs and x-ray information. It is identified using HCPCS and CPT codes rather than ICD9 codes. The data page “61” contains information about outpatient procedures. In this case, HCPCS and CPT codes are used to identify the type of service. There can be multiple “61” pages per patient, if the patient has multiple outpatient procedures in one episode. The data page “70” contains information on an inpatient admission. It uses ICD9 codes to record information about patient conditions and procedures.

At this point, we need to discuss the importance of the codes in defining a patient condition. (Misset, et al., 2008; Pine, et al., 2009; Worth & Mytinger, 1996) ICD9 codes were developed by the World Health Organization and are used in hospitals. CPT codes are used by physicians and were developed by the American Medical Association, remaining proprietary to them. DRG codes were developed by Medicare largely for billing