Exploring Medicare Costs using Machine Learning

Alec Ashburn

Medical and Bioinformatics Graduate Program, School of Computing and Information Systems,
Grand Valley State University, Allendale, MI, USA

Summary

As various forms of technology become more ubiquitous in the field of health care, an enormous amount of data is being collected in hope of making new scientific discoveries and reforming the way we understand health care as a society. Specifically, data mining has opened up a portal to discovery and comprehension of otherwise meaningless information. Preprocessing and cleaning techniques, advanced machine learning algorithms, and data visualization tools can be of extraordinary use when trying to make sense of the vast amount of health information at our fingertips. One area of health care that is always undergoing reform and debate is Medicare. I decided to analyze inpatient Medicare coverage data in R for the years 2011 through 2013 to get a better idea of how Medicare dollars are being spent in recent years, how they compare to past spending rates, and what future rates may look like. I was able to determine that among the medical procedures that receive the most Medicare coverage, the top causes of death among the elderly were not included. Many other factors contribute to Medicare costs and were not explored during my research, but the data that I was able to analyze using data mining techniques provides a great deal of insight into an area of much discussion and controversy.

Background

Created in 1966, Medicare provides health insurance to Americans over the age of 65 who have worked and paid into the system. Additionally, Medicare provides insurance to those with disabilities, end-stage renal disease, and amyotrophic lateral sclerosis. As of 2010, Medicare provided insurance to over 48 million Americans and is the primary payer for over 15 million inpatient stays, which accounts for 47.2% of total inpatient costs in the United States. On average, Medicare covers about half of all health care charges for those enrolled. This means that the remainder of the bill must be paid out-of-pocket or with a supplemental insurance. Medicare can currently be divided into four distinct parts, two of which vary among states and two of which are federally ensured. Part A and Part B are federally provided by the United States. Part A is insurance used for hospital visits and Part B is utilized during medical procedures. Part C and Part D vary by state. Part C is primarily concerned with private insurance companies and Part D (added in 2006) refers to a prescription drug plan among those qualified. In order to gain a better understanding about how costs contrast throughout the United States, I looked at inpatient Medicare cost data from 2011 – 2013 provided by Centers for Medicare & Medicaid Services.

Methods

The methods I used for my research primarily consisted of data mining within the statistical software R. The original datasets were in CSV Excel sheet formats, consisted of over 450,000 entries, and needed much cleaning prior to information extraction. Once the datasets were standardized throughout all 3 years, I aggregated the data into sub tables that compared average covered charges, provider state, and the medical surgery covered. The final step was creating decision trees for the top 5 Medicare-covered states to see which state offers the best coverage pertaining to each diagnosis.

Results and Discussion

For 2011 – 2013, I determined that these states had the highest average amount covered for Medicare charges:

1) California - $881,932,621
2) Florida - $513,311,133
3) Texas - $492,121,064
4) New Jersey - $319,122,582
5) Pennsylvania - $309,303,455

I also determined these medical procedures to have the highest Medicare coverage:

1) Major Small and Large Bowl Procedures
2) Infectious and Parasitic Diseases
3) Respiratory System Diagnosis
4) Septicemia or Severe Sepsis
5) Major Joint Replacement

Although each decision tree determined that states varied in the amount of coverage for the top 5 medical procedures mentioned earlier, I also found that these top 5 states were also the biggest Medicare spenders in the last 20 years. However, many states, such as Nevada and Alaska, have increased their spending by over 10% in the past 20 years.

After looking at the results from my decision trees, I was curious as to what the distribution of Medicare looked like for each state based on certain demographics. For the top 3 spenders in the United States, it was found that the majority of its Medicare recipients were Caucasian, female, and above the poverty line. I also discovered that the top 5 causes of death among the elderly were not included entirely in the top 5 medical conditions covered by Medicare throughout 2011, 2012, and 2013. This is an extremely important finding for future research as both health care systems and politicians craft the future of Medicare in the United States. It is also important to point out that since Medicare covers half the cost, on average, these costs only represent about half of what the medical procedures actually cost throughout 2011, 2012, and 2013.

Acknowledgements:

References:

1. Centers for Medicare & Medicaid Services. *Medicare Provider Utilization and Payment Data: Inpatient.* 2011 – 2013.

2. Altman, Drew; Frist, William H. (July 28, 2015). *Medicare and Medicaid at 50 Years: Perspectives of Beneficiaries, Health Care Professionals and Institutions, and Policy Makers.* JAMA 314(4): 384-395.

3. Torio CM, Andrews RM. *National Inpatient Hospital Costs: The Most Expensive Conditions by Payer, 2011.* HCUP Statistical Brief #160. Agency for Healthcare Research and Quality, Rockville, MD. August 2013.

4. Kaiser Slides, The Henry J. Kaiser Family Foundation. Facts.kff.org. Retrieved on April 3, 2016.

5. Agency for Healthcare Research and Quality, *Statistical Brief #327: Top Five Most Costly Conditions among the Elderly, Age 65 and Older. 2008.*

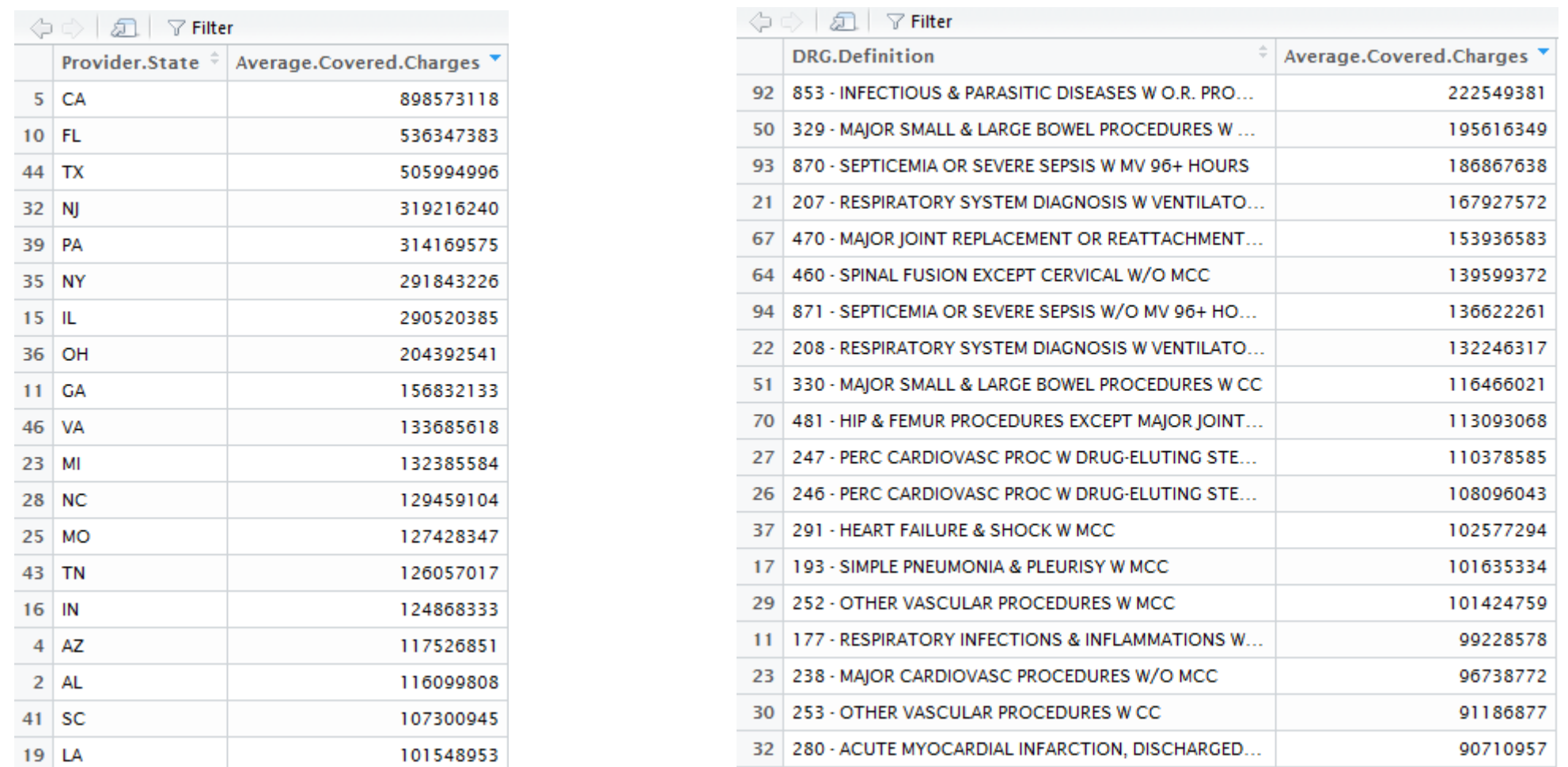*Figure 1: Inpatient Data Example*

Figure 2: Data tables created in R to find top 5 states and procedures by amount covered
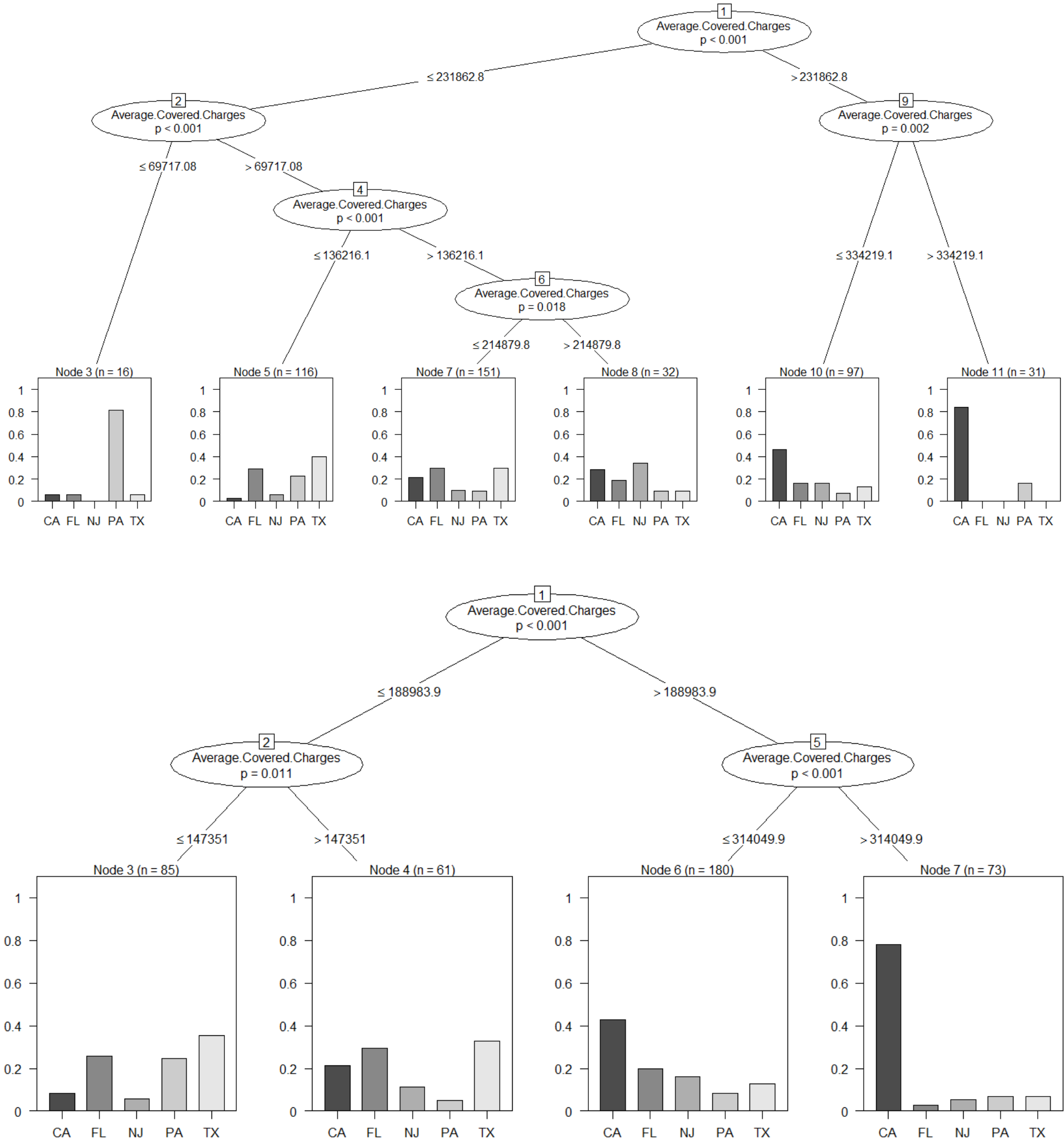

Figure 3: Decision trees for bowl procedures (2011) and infectious diseases (2012)