# **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-ofpocket 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

#### **Acknowledgements:**

I would like to thank Dr. Guenter Tusch for his work as my advisor on this project, Dr. Timothy Born for his work as my mentor in the PSM program, and Professor Greg Wolffe for his expertise in machine learning techniques that helped clarify my data mining approach.

- 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.

#### References:

- 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.

4	А	В С	D	E	F	G H	1	J	К	L [
1	DRG Definition	Provider Id Provider Name	Provider Street Address P	rovider City	Provider State Provider	Zip Code Hospital Referral Region (HRR	R) Total Discharges Average	Covered Charges	Average Total Payments	Average Medicare Payments
2	039 - EXTRACRANIAL PROCE	10001 SOUTHEAST ALA	1108 ROSS CLARK CIRCLE D	OTHAN	AL	36301 AL - Dothan	91	32963.07692	5777.241758	4763.73
3	039 - EXTRACRANIAL PROCE	10005 MARSHALL MED	01 2505 U S HIGHWAY 431 N B	BOAZ	AL	35957 AL - Birmingham	14	15131.85714	5787.571429	4976.71
4	039 - EXTRACRANIAL PROCE	10006 ELIZA COFFEE N	IE 205 MARENGO STREET FI	LORENCE	AL	35631 AL - Birmingham	24	37560.375	5434.958333	4453.79
5	039 - EXTRACRANIAL PROCE	10011 ST VINCENT'S E	A 50 MEDICAL PARK EAST [ B	BIRMINGHAM	AL	35235 AL - Birmingham	25	13998.28	5417.56	412
6	039 - EXTRACRANIAL PROCE	10016 SHELBY BAPTIST	1000 FIRST STREET NORTI	ALABASTER	AL	35007 AL - Birmingham	18	31633.27778	5658.333333	4851.44
7	039 - EXTRACRANIAL PROCE	10023 BAPTIST MEDIC	A 2105 EAST SOUTH BOULE M	MONTGOMERY	AL	36116 AL - Montgomery	67	16920.79104	6653.80597	5374.14
8	039 - EXTRACRANIAL PROCE	10029 EAST ALABAMA	1 2000 PEPPERELL PARKWA O	PELIKA	AL	36801 AL - Birmingham	51	11977.13725	5834.745098	4761.41
9	039 - EXTRACRANIAL PROCE	10033 UNIVERSITY OF	A 619 SOUTH 19TH STREET B	BIRMINGHAM	AL	35233 AL - Birmingham	32	35841.09375	8031.125	58
10	039 - EXTRACRANIAL PROCE	10039 HUNTSVILLE HO	S 101 SIVLEY RD H	IUNTSVILLE	AL	35801 AL - Huntsville	135	28523.39259	6113.385185	52
11	039 - EXTRACRANIAL PROCE	10040 GADSDEN REGIO	D 1007 GOODYEAR AVENUEG	SADSDEN	AL	35903 AL - Birmingham	34	75233.38235	5541.058824	4386.94
12	039 - EXTRACRANIAL PROCE	10046 RIVERVIEW REG	600 SOUTH THIRD STREET G	SADSDEN	AL	35901 AL - Birmingham	14	67327.92857	5461.571429	4493.57
13	039 - EXTRACRANIAL PROCE	10055 FLOWERS HOSP	I 4370 WEST MAIN STREET D	OTHAN	AL	36305 AL - Dothan	45	39607.28889	5356.288889	44
14	039 - EXTRACRANIAL PROCE	10056 ST VINCENT'S B	IF 810 ST VINCENT'S DRIVE B	BIRMINGHAM	AL	35205 AL - Birmingham	43	22862.23256	5374.651163	4186.02
15	039 - EXTRACRANIAL PROCE	10078 NORTHEAST AL	A 400 EAST 10TH STREET A	NNISTON	AL	36207 AL - Birmingham	21	31110.85714	5366.238095	4376.23
16	039 - EXTRACRANIAL PROCE	10083 SOUTH BALDWI	N 1613 NORTH MCKENZIES F	OLEY	AL	36535 AL - Mobile	15	25411.33333	5282.933333	4383.73
17	039 - EXTRACRANIAL PROCE	10085 DECATUR GENE	R 1201 7TH STREET SE D	DECATUR	AL	35609 AL - Huntsville	27	9234.518519	5676.555556	4509.11
18	039 - EXTRACRANIAL PROCE	10090 PROVIDENCE H	0 6801 AIRPORT BOULEVAF N	NOBILE	AL	36608 AL - Mobile	27	15895.85185	5930.111111	3972.85
19	039 - EXTRACRANIAL PROCE	10092 D C H REGIONA	809 UNIVERSITY BOULEVAT	USCALOOSA	AL	35401 AL - Tuscaloosa	31	19721.16129	6192.548387	5179.38
20	039 - EXTRACRANIAL PROCE	10100 THOMAS HOSPI	T 750 MORPHY AVENUE FA	AIRHOPE	AL	36532 AL - Mobile	18	10710.88889	4968	3898.88
21	039 - EXTRACRANIAL PROCE	10103 BAPTIST MEDIC	A 701 PRINCETON AVENUE B	BIRMINGHAM	AL	35211 AL - Birmingham	33	51343.75758	5996	4962.45
22	039 - EXTRACRANIAL PROCE	10104 TRINITY MEDICA	800 MONTCLAIR RD B	BIRMINGHAM	AL	35213 AL - Birmingham	29	55219.31034	5710.310345	4471.68
23	039 - EXTRACRANIAL PROCE	10113 MOBILE INFIRM	45 MOBILE INFIRMARY CIR N	MOBILE	AL	36652 AL - Mobile	66	14948.15152	5550.909091	4219.90
24	039 - EXTRACRANIAL PROCE	10139 BROOKWOOD	2010 BROOKWOOD MED B	BIRMINGHAM	AL	35209 AL - Birmingham	19	73846.21053	4987.263158	3944.42
25	039 - EXTRACRANIAL PROCE	20017 ALASKA REGION	2801 DEBARR ROAD A	NCHORAGE	AK	99508 AK - Anchorage	23	34805.13043	8401.956522	6413.78
26	039 - EXTRACRANIAL PROCE	30002 BANNER GOOD	S 1111 EAST MCDOWELL R(P	HOENIX	AZ	85006 AZ - Phoenix	11	34803.81818	7768.909091	6951.45
27	039 - EXTRACRANIAL PROCE	30006 TUCSON MEDIC	A 5301 EAST GRANT ROAD TO	UCSON	AZ	85712 AZ - Tucson	40	24474.75	6799.85	5764
28	039 - EXTRACRANIAL PROCE	30007 VERDE VALLEY	M 269 SOUTH CANDY LANE C	COTTONWOOD	AZ	86326 AZ - Phoenix	18	28571.61111	9133	8008.11
29	039 - EXTRACRANIAL PROCE	30010 CARONDELET ST	1601 WEST ST MARY'S RC T	UCSON	AZ	85745 AZ - Tucson	12	35968.5	6506.5	5379.83
30	039 - EXTRACRANIAL PROCE	30011 CARONDELET S	350 NORTH WILMOT ROAT	UCSON	AZ	85711 AZ - Tucson	42	26294.52381	6083.428571	4903.33
31	039 - EXTRACRANIAL PROCE	30012 YAVAPAI REGIO	1003 WILLOW CREEK ROAP	RESCOTT	AZ	86301 AZ - Phoenix	28	26771.78571	7140.857143	6133.57
32	039 - EXTRACRANIAL PROCE	30013 YUMA REGIONA	L 2400 SOUTH AVENUE A	'UMA	AZ	85364 AZ - Phoenix	20	29967.8	6978.75	596
33	039 - EXTRACRANIAL PROCE	30023 FLAGSTAFF MED	1200 NORTH BEAVER STR F	LAGSTAFF	AZ	86001 AZ - Phoenix	15	27349.4	11026.33333	9056.06
34	039 - EXTRACRANIAL PROCE	30024 ST JOSEPH'S HO	S 350 WEST THOMAS ROAL P	HOENIX	AZ	85013 AZ - Phoenix	18	59443.83333	8487.444444	7422.66
35	039 - EXTRACRANIAL PROCE	30030 PHOENIX BAPTI	S 2000 WEST BETHANY HO! P	HOENIX	AZ	85015 AZ - Phoenix	42	24714.66667	7673.214286	6541.92
36	039 - EXTRACRANIAL PROCE	30038 SCOTTSDALE HE	7400 EAST OSBORN ROALS	COTTSDALE	AZ	85251 AZ - Phoenix	13	37706	6590.230769	
37	039 - EXTRACRANIAL PROCE	30061 BANNER BOSW	E 10401 WEST THUNDERBIFS	UN CITY	AZ	85351 AZ - Sun City	62	38149.85484	6048.612903	4951.90
20	MO EVEDA PANIAI DECCE		2200 600/1/10/1/14/2 8/6	HOWLOW	۸7	95901 A7 Dhooniy	17	15021 520/1	10000 92252	7221 64
	inpatient2011	<u></u>				: [	<u> </u>			<b>)</b>

Figure 1: Inpatient Data Example

o ⇔   Æ   ▼ Filte	er	<b>\( \( \)</b>		
Provider.State †	Average.Covered.Charges		DRG.Definition	Average.Covered.Charges
CA	898573118	92	853 - INFECTIOUS & PARASITIC DISEASES W O.R. PRO	222549381
FL	536347383	50	329 - MAJOR SMALL & LARGE BOWEL PROCEDURES W	195616349
TX	505994996	93	870 - SEPTICEMIA OR SEVERE SEPSIS W MV 96+ HOURS	186867638
NJ	319216240	21	207 - RESPIRATORY SYSTEM DIAGNOSIS W VENTILATO	167927572
PA	314169575	67	470 - MAJOR JOINT REPLACEMENT OR REATTACHMENT	153936583
NY	291843226	64	460 - SPINAL FUSION EXCEPT CERVICAL W/O MCC	139599372
IL	290520385	94	871 - SEPTICEMIA OR SEVERE SEPSIS W/O MV 96+ HO	136622261
ОН	204392541	22	208 - RESPIRATORY SYSTEM DIAGNOSIS W VENTILATO	132246317
GA	156832133	51	330 - MAJOR SMALL & LARGE BOWEL PROCEDURES W CC	116466021
VA	133685618	70	481 - HIP & FEMUR PROCEDURES EXCEPT MAJOR JOINT	113093068
МІ	132385584	27	247 - PERC CARDIOVASC PROC W DRUG-ELUTING STE	110378585
NC	129459104	26	246 - PERC CARDIOVASC PROC W DRUG-ELUTING STE	108096043
МО	127428347	37	291 - HEART FAILURE & SHOCK W MCC	102577294
TN	126057017	17	193 - SIMPLE PNEUMONIA & PLEURISY W MCC	101635334
IN	124868333	29	252 - OTHER VASCULAR PROCEDURES W MCC	101424759
AZ	117526851	11	177 - RESPIRATORY INFECTIONS & INFLAMMATIONS W	99228578
AL	116099808	23	238 - MAJOR CARDIOVASC PROCEDURES W/O MCC	96738772
sc	107300945	30	253 - OTHER VASCULAR PROCEDURES W CC	91186877
LA	101548953	32	280 - ACUTE MYOCARDIAL INFARCTION, DISCHARGED	90710957

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

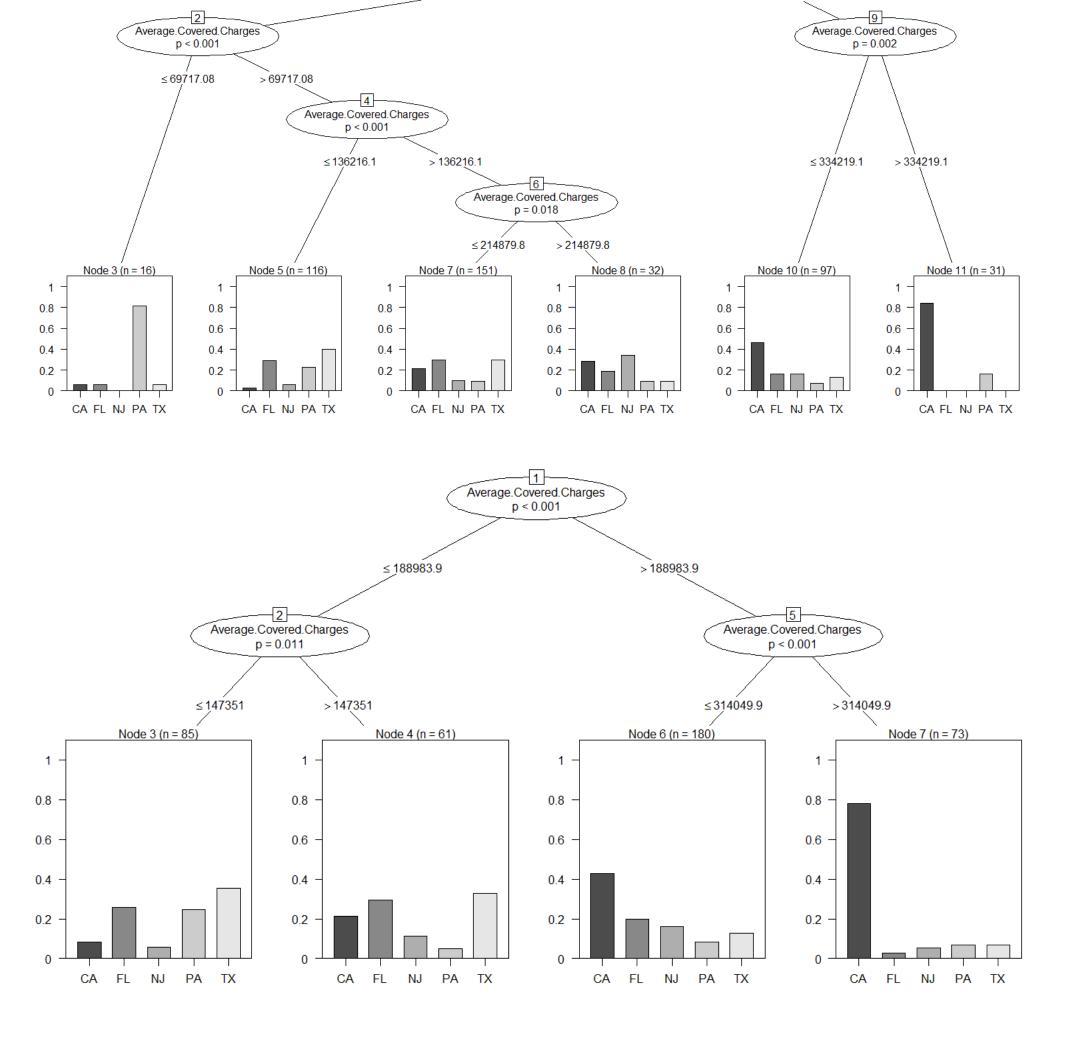


Figure 3: Decision trees for bowl procedures (2011) and infectious diseases (2012)