MEDICARE HEALTHCARE CHARGE DISPARITY ANALYSIS

Prasun Das[1], Gopalakrishna Palem[2], Neha Srivastava[3], Rajat Swaroop[4]

Analytics Division, Symphony-Teleca Corporation, India
Prasun.Das@Symphonyteleca.com, Gopalakrishna.Palem@Yahoo.com,
Neha.Srivastava@Symphonyteleca.com, Rajat.Swaroop@Symphonyteleca.com

ABSTRACT

Transparency in administration and effective corporate governance leads to huge volumes of public data that when processed with analytical procedures yield meaningful insights into the nature of the data and the business operations generating that data. A case in point being the recent public release of United States Medicare charges for healthcare system. In an initiative dubbed Obamacare, United States government released data for hospital charges for top 100 Diagnosis-Related Groups (DRG). The data included in-patient provider charges, number of discharges per each hospital for every DRG and the respective reimbursements. This paper presents our analysis on the intricacies involved in analyzing such public data, along with some interesting results we obtained in the process. We also present the disease classification system that was used to identify the culprit hospitals causing the disparity.

KEYWORDS

Healthcare costs, Medicare, Charge disparities, Obamacare, Health analytics

1.INTRODUCTION

In an initiative dubbed *Obamacare*, United States government released data for hospital charges for top 100 Diagnosis-Related Groups (DRG) [1]. The data included in-patient provider charges, number of discharges per each hospital for every DRG and the respective reimbursements. As part of our research work that is actively being carried with the goal of building an automated expert system capable of analyzing public data and providing actionable insights, we undertook the case-study of analyzing the afore-said *Obamacare* DRG public data, and here we share the process we followed and some of the results achieved.

This study is important for us on two levels:

1. Many of our customers from the healthcare segment are Accountable Care Organizations (ACO) that has to comply with *Obamacare* health reforms in the United States. As part of the reforms, an urgent attention is warranted by the federal government to ensure continued improvements in quality and progress on reducing disparities in ACO service charges [2]. Our analysis helped our customers find insights into the disparity and act upon the outlier ACOs.
2. The insights and results accumulated in the process of this work will act as bootstrap data that in later stages be used as foundations for our *expert system* framework we are working on [3].

DOI: 10.5121/hiij.2013.2302

The data released by Center for Medicare & Medicaid Services (CMS) include hospital-specific charges for the more than 3,000 U.S. hospitals that receive Medicare Inpatient Prospective Payment System (IPPS) payments for the top 100 most frequently billed discharges, paid under Medicare based on a rate per discharge using the Medicare Severity Diagnosis Related Group (MS-DRG) for Fiscal Year (FY) 2011. This is the latest available public data released in the first quarter of 2013, and these DRGs represent almost 7 million discharges or 60 percent of total Medicare IPPS discharges [1].

However, during the analysis of this data it was found that there is a massive disparity in medical costs across service providers. While the *Healthcare Disparities Report* [2] confirms this disparity, it, however, does not identify the individual hospitals that are causing this disparity or by how much the variation is. For our customers this information is important to act upon and conform to Obamacare regulations. Hence we do a more thorough study of the data at the individual hospital level and disease level, while considering the aggregated state-level data as the base norm.

For this analysis, we used a combination of multiple statistical analysis techniques. Outliers in the hospital charges were detected using the *outlier-detection* method suggested by Hodge et al. in [4], while hospital inpatient estimations and aggregations based on number of discharges were done based on the methods suggested in [5]. The disease classification system, presented in section 3, is one of our main contributions in this paper; along with the results achieved (though for brevity we only presented top 5 results for each category). These results are readily reproducible by following the methods explained and can help any ACO gain actionable-insight into the charge disparity and act upon.

The rest of the paper is organized as follows. Section 2 presents an overview of the disparities in the charges, highlighting the top 5 and bottom 5 DRG groups that are being subjected to the disparity and their variation in all states. Section 3 proceeds for a more in-depth analysis towards identifying the individual culprit hospitals that are charging more, and explains the methods we used for that, followed by conclusion and references. For notational flexibility, the terms DRG and disease / disease group are used interchangeably in this paper.

3. HEALTHCARE CHARGE DISPARITY

The DRG charge data exhibits tremendous unexplained variation in cost of services, not only at national level, but also at street level and county level [2]. The data shows that, even on the same street, hospital charges can vary by upwards of 200% for the same service – a case in point being, two non-profit hospitals that sit on opposite side of the same street in Miami (namely University of Miami Hospital and Jackson Memorial Hospital) costing $166,174 and $89,027 on average for the same *DRG – Heart Attack with 4 stents and major complications.*

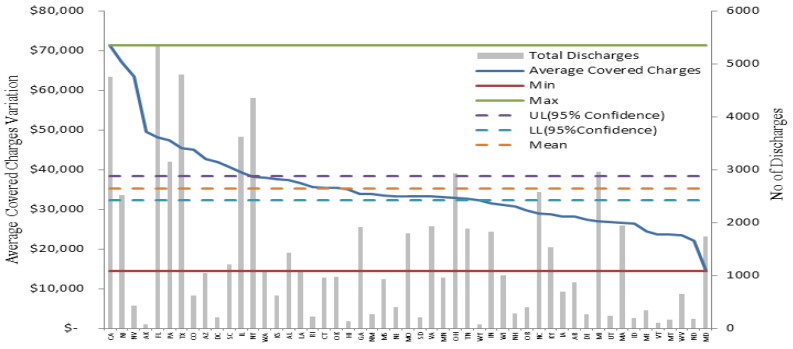

Figure 1. Average charge variation across states in US for all diseases

In another case from downtown New York City, two hospitals 60 blocks apart varied by 320% in the prices they charged to treat complicated cases of asthma or bronchitis. One charged an average of $34,310 while the other billed, on average, $8,159. Figure 1 shows aggregated on state level, the average cover charges across all the hospitals and all DRGs. It can be seen that the states California, Alaska and Nevada emerge as charging high across all the disease groups overall, compared against other states.

To understand the charge disparity on individual DRGs, the top 5 and bottom 5 DRGs with the highest and lowest number of discharges were taken into account and their charge disparity across states were plotted.

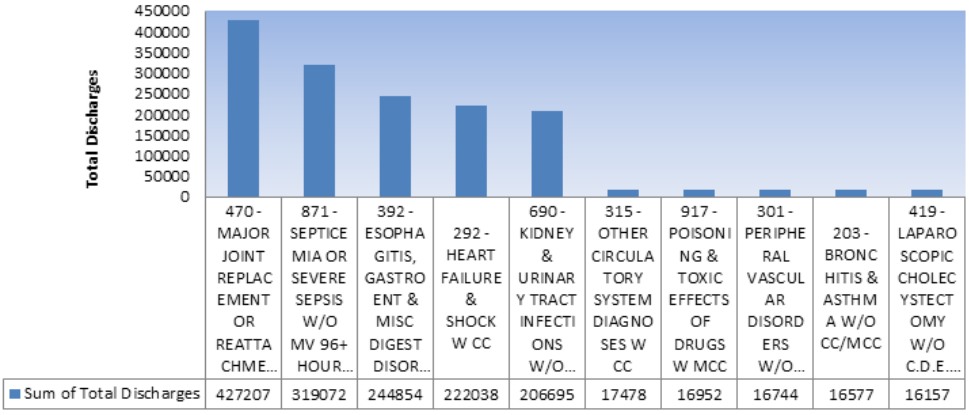

Figure 2. Top 5 and Bottom 5 DRGs with highest and lowest number of discharges

Figure 2 above identifies the DRGs with the top 5 and bottom 5 discharge counts, while Figure 3 below shows the charges for these DRGs across all states. From Figure 2, on national level, DRG 470 (Major Joint Replacement) emerges as the top most disease group with highest number of discharges across states and its charge variation can be seen in Figure 3, with around $88,000 being charged in California, while it costs just about $25,000 in Maryland (CA charging almost 3 times more than MD).

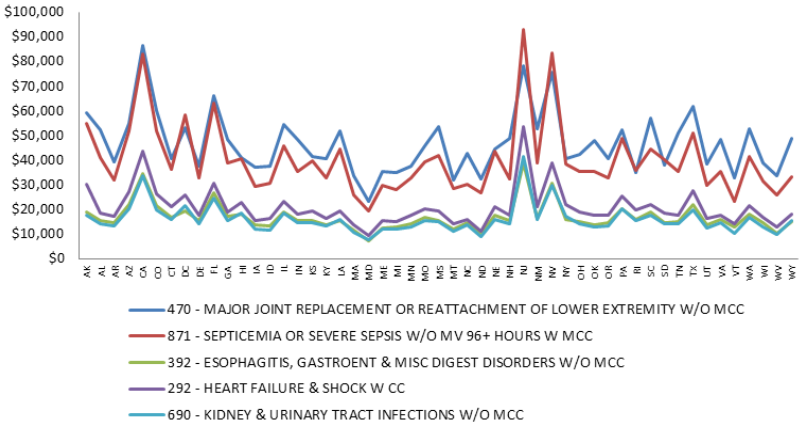

Figure 3. Average charges for DRGs with highest number of discharges showcasing variation

From the above chart it is clear that hospitals in different states have different charges for the same DRG (with variation being more than 2 times or 3 times). However, as Rossiter et al. points out in [8], geographic factors could be argued as the cause for charge variation in hospitals. After all, it is expected that economic policies in one state cannot be the same as those that of another state. However, we rule this possibility out by doing an intra-state comparison of price variation. This establishes that more than geographic factors, the governance and administration factors are the leading cause of charge variation and disparity [6].

For this we drilled down another level, looking into charge distributions for individual diseases across hospitals within a state. We analyze individual hospital's pricing for every disease and classify if they are pricing it higher or lower than the industry average in that state.

We calculated the average charge across all DRGs on a hospital level, and then the outlier hospitals (based on the normal distribution curve) were labeled as *High Priced* hospitals, and *Low Priced* hospitals, based on their charges being higher or lower than the standard deviation [4]. The ratio of *High Priced* to *Low Priced* Hospitals for every state is depicted in the graph of Figure 4.

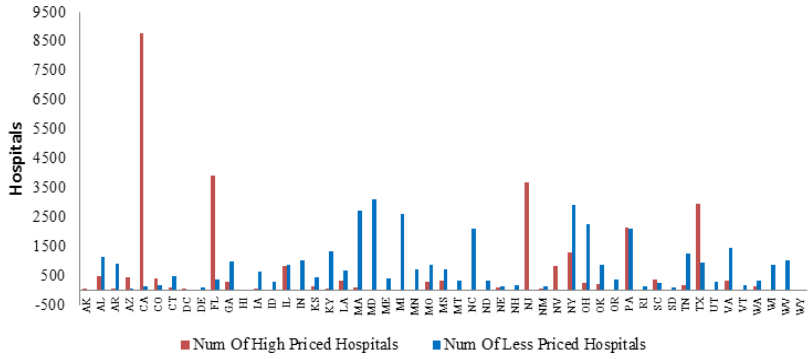

Figure 4. Number of High-Priced and Low-Prices hospitals by State

The intra-state high-priced to low-priced hospital ratio within a state points to the absence of regulated pricing for hospitals. In an ideal state where regulations are enforced, the ratio should be as low as possible, with almost all hospital charging about the same price as that of average. However the graph in Figure 4 presents a different picture, with the high-priced hospital count radically different from that of low-priced ones. States like California (CA), Florida (FL) and New Jersey (NJ) have disturbingly radical variation in the ratio, pointing to the lack of price standardization.

In the following section, we move onto the finding out the exact culprit hospitals that are triggering this variation.

2.1. DETECTING THE CULPRIT HOSPITALS

Statistically it is not possible to identify the culprit hospitals without aggregating the data either on state level or disease level. However, such aggregations will lead to incorrect results causing Simpson's paradox [9, 10] due to the fact that the number of discharges for every DRG varies and the costs associated with them are different. In order to avoid this, we employed a clustering technique.

To detect the culprit hospitals, we first grouped the diseases into classes based on the average cover charges across all hospitals and state. A standard deviation curve is computed and three

classes, namely *High-cost* disease, *Average-cost* disease, and *Low-cost* disease are computed based on the distance from mean.

The classification was done using the percentile distribution of the average covered charges across all hospitals treating that disease. Hospitals treating a disease are calculated by looking at the number of discharges for that corresponding DRG for that hospital. If the number of discharges is greater than zero then the hospital is implied as providing the treatment for that disease.

Table 1 Disease class grouping based on percentile charges

Percentile	**Disease cost group**
90-100	High cost
50-89	Average cost
0-49	Low cost

The overall charge distribution for every individual disease cost group was divided into inter-quartile range, and charges for a disease in a hospital were mapped within that inter-quartile range of charges of all hospitals, thus labeling the hospital into a charge group for that disease. An example of the outliers in the charges within the charge distribution can be seen from the below diagram (the small circles indicate the outliers, while the boxes indicate the normal range).

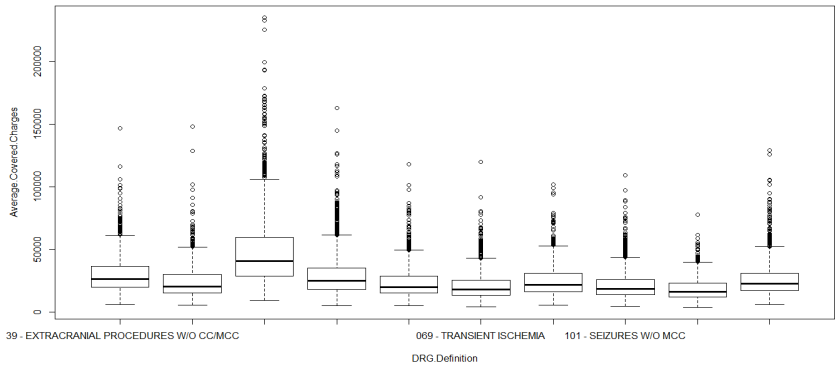

Figure 5. Disease charge variations across hospitals

The procedure outline is as below:

- Extremities in hospital charges has been diagnosed based on count of DRGs for each hospital
- Within each DRG, hospitals charging more number of DRGs at much higher cost than their group's average charge has been marked as culprit hospitals

This cross tabulation of hospital-disease-charge combination shows how many Low and Avg. Cost diseases have been charged more than their average value within their corresponding group. The count of such charge disparities is then calculated for every hospital. The more the number of charge disparities count for the hospital, the higher it will raise in its rank as the culprit hospital.

This study revealed us that many hospitals are treating as much as 48 low-cost diseases for much higher price than their average cost. Further, below details are observed:

- 1089 hospitals are overcharging on more DRGs than similar peer hospitals in their group
- Among the 1089, most hospitals, as many as 462, are overcharging in two categories (high pricing a low cost disease as well as average cost disease)
- Overall 189 hospitals across US are found to overcharge patients in all charge groups

The top 5 of these overcharging hospitals can be seen in Table 2 below.

Table 2. Top 5 Hospitals overcharging most number of Low Cost and Avg. Cost diseases

Hospital	State	Count of Heavily Priced Low-cost diseases	Count of Heavily Priced Average-cost diseases
Cedars-Sinai Medical Center	CA	48	40
Crozer Chester Medical center	PA	43	36
Somerset Medical center	NJ	48	31
Eisenhower Medical center	CA	42	36
Robert wood Johnson Univ. Hosp.	NJ	46	31

If a hospital has been overcharging more number of DRGs than similar discharge hospitals then it will be classified as violation. For instance, in California 31 hospitals meet such criteria of over-charging more number of diseases from one DRG cost group compared to other hospitals handling similar volume of discharges. Similarly in Florida 25 hospitals has over charged more number of diseases in all cost groups.

However, while the cost variations across hospitals are clear, the reasons behind them are not. Some commonly attributed factors for cost variations are quality of care provided by each hospital, complexity of disease for each patient, demand for the service in the locality and so on. However, it is not clear if those factors accurately justify the variation. A quick study by us (not covered in this paper) mapping the hospital infrastructure (the number of beds, number of physicians, specialists and so on) to the hospital charges for diseases revealed no significant correlation to justify the charges variation.

A detailed socio-economic factors study relating the charge groups with in-patient inflow, average income level and other social factors might reveal addition insights. One of the major hurdles for such studies becoming wide spread, though, is the absence of publicly available hospital infrastructure data in relation to its socio-economic factors of geographies. However, with the health reforms heavily being focused on improving the quality and transparency, we hope to see more public data available, making such detailed studies possible.

3. CONCLUSIONS

The United States federal healthcare costs data shows massive disparity in medical costs among service providers. While there is no common agreement on how much patients, insurance providers and government actually end up paying, the disparity of charges among hospitals for same DRG, nonetheless, points to the absence of standards and ineffective governance. The *Obamacare* initiative attempts to address these issues by bringing transparency and quality control into the system.

Although government has collected hospital charge information for years, it was till now housed in closed databases that one has to pay for access. With the public release of this data, it now becomes possible to assess it much granular level and identify the reasons for disparity.

The data displays charge disparity of about 150 to 200% across states. While it can be argued that geographic and local economies are the cause of charge variation in hospitals, we, however, ruled this possibility out by doing an intra-state comparison of price variation. It is established in this paper that more than geographic and location specific factors, the governance and regulatory factors are the leading cause. The list of top 5 and bottom 5 DRGs with highest and lowest number of discharges was presented, along with the analysis of their charge variation across all states. With our disease classification system that segregated the diseases into quintiles, we were able to find out the culprit hospitals and the list of top 5 of them was presented.

REFERENCES

[1] Medicare Provider Charge Data, Center of Medicare and Medicaid Services, Available at: http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and- Reports/Medicare-Provider-Charge-Data/Inpatient.html

[2] U.S. Department of Health and Human Services, (2012) "National Healthcare Disparities Report", AHRQ Publication No. 12-0006.

[3] Gopalakrishna Palem, (2013) The Practice of Predictive Analytics in Healthcare, Symphony Teleca Corp

[4] Hodge, V.J. and Austin, J. (2004) "A survey of outlier detection methodologies", Artificial Intelligence Review, 22 (2). pp. 85-126.

[5] Levit KR, Friedman B, Wong HS. (2013), "Estimating Inpatient Hospital Prices from State Administrative Data and Hospital Financial Reports", Health Services Research, doi: 10.1111/1475-6773.12065

[6] Giles K, (2011) "6 strategies for building bundled payments", Healthcare financial management, 65(11): 104-110.

[7] Steven Brill, (2013) "An End to Medical-Billing Secrecy", Time Magazine

[8] Rossiter LF, Adamache KW, (1990), "Payment to health maintenance organizations and the geographic factor", Healthcare financing review, 12(1): 19-30

[9] Clifford H. Wagner (1982). "Simpson's Paradox in Real Life". The American Statistician 36 (1): 46–48

[10] Marios G. Pavlides, Michael D. Perlman (2009), "How Likely is Simpson's Paradox?", The American Statistician 63 (3): 226–233

Authors

Gopalakrishna Palem is a Corporate Technology Strategist specialized in Distributed Computing technologies and Cloud operations. During his 12+ year tenure at Microsoft and Oracle, he helped many customers build their high volume transactional systems, distributed render pipelines, advanced visualization & modelling tools, real-time dataflow dependency-graph architectures, and Single-sign-on implementations for M2M telematics. When he is not busy working, he is actively engaged in driving open-source efforts and guiding researchers on Algorithmic Information Theory, Systems Control and Automata, Poincare recurrences for finite-state machines, Knowledge modelling in data-dependent systems and Natural Language Processing (NLP).

Prasun is a consultant at Symphony Teleca Corporation working in the field of Data analytics and Statistics. He has completed his B. Tech from West Bengal University of Technology.

Neha Srivastava is an MBA in Marketing from IIPM, Brussels. She is presently working as a consultant at Symphony Teleca Corporation in the domains of Healthcare and Retail.

Rajat Swaroop is a senior consultant at Symphony Teleca Corporation, working for the past 3.5 years in field of Data Science. He has completed his B. Tech from IIT Bombay and Masters from Technical University of Delft and University of Paris-11.