Clustering ASP Drug Cost and Utilization Data

**Executive Summary**

This analysis focuses on average sales price (ASP) drugs prescribed and administered by physicians and other providers in Wisconsin, taken from the Medicare Fee-For-Service Provider Utilization and Payment Data.

Using K-means clustering, two main insights were gained. The first is a cost saving measure which is actionable in the form of a marketing or publicity push. Medicare currently covers almost all of the cost of flu/pneumococcal vaccines. However, expenditure is almost twice as much ($48 to $83), on average, when administered by a doctor, as compared to a mass immunizer or pharmacy. However, most Medicare recipients receive these immunizations from a doctor. If Medicare is able to market pharmacies or mass immunizers as more convenient than making an appointment with a physician, then Medicare could reduce costs on the majority of flu vaccinations.

The second insight is one of Medicare’s coverage gaps. With regard to most drugs in most clusters, Medicare covered the vast majority of the bill. However, with three drugs related to macular degeneration, Medicare currently only covers approximately 77% of a bill that averages just over $435. Macular degeneration is chronic, and most Medicare recipients are unemployed. Furthermore, ratio of drugs to patients is very high, so the bill that a patient ends up with is considerable. Since Medicare is not a cost minimizing, profit driven business, but rather a government entity focused on providing quality health insurance, Medicare should improve their coverage of the drugs that treat macular degeneration to ease the financial burden of elderly citizens with a chronic illness.

**Introduction**

Within the Medicare Fee-For-Service Provider Utilization and Payment Data[[1]](#footnote-1), there is a feature that flags observations corresponding to drugs whose manufacturers must send their average sales prices to CMS quarterly. Even without regard to ASP drugs, CMS has the unique ability to negotiate drug prices.[[2]](#footnote-2) The data is relatively large, so I subset ASP expenditures specifically in my home state of Wisconsin. I used clustering to examine patterns in Medicare expenditure on ASP drugs with the goal of recommending areas in which Medicare could expand its coverage on drugs where patients feel a significant financial burden or reduce costs by eliminating anomalous expenditures. I chose to subset based on medication because CMS has the ability to both negotiate drug prices and have manufacturers publish average sales prices. CMS has more ability to decrease drug related expenditures than procedure-based expenditures (i.e. surgery) since CMS has more leverage over the manufacturers of the ASP products than it does over physicians (re: negotiation).

**Methodology**

Using sqlite3, a SQL database was created and then queried. Some exploratory queries were performed. Then the data was filtered to only include observations related to ASP drugs within the state of Wisconsin. 9754 observations were in the final dataset, with 3184 unique healthcare providers.

Categorical variables were dummy coded with a column for each unique level. I reduced the variable *nppes\_credentials* to a two category variable that distinguished between doctors and all other providers. My hypothesis was that a doctor administering drugs may end up as a higher cost encounter for Medicare. The variable *nppes\_provider\_type* was handled in a similar manner. Categories that comprised less than 1% of the data were collapsed into an “other” category. Provider type encoded important information about the department in which a FFS provider works.

Numeric variables were highly skewed. Log transformations were used on line service count and beneficiary service count columns as well as the payment and charge columns. After transformation, all features were scaled using StandardScaler. K-means clustering was then applied to the data.

To select the optimal value of k, two methods were utilized. First was to look for the “elbow” in a scree plot of SSE vs number of clusters. The elbow is where returns on more clusters diminish.

A screenshot of a cell phone

Description automatically generated

Figure SSE and Number of Clusters

The second test was an inspection of the silhouette plot. The silhouette plot confirmed the scree plot in selecting an optimal number of clusters. Both suggested 16 clusters.

A screenshot of a cell phone

Description automatically generated

Figure Silhouette and Number of Clusters

HCPCS provides descriptions of drugs and procedures as a column. So after the clusters were adjoined to the original dataset, the HCPCS descriptions were as well. This allowed for deeper insight into the types of drugs contained within each cluster.

**Results**

With 16 clusters, each one took a considerable amount of time to analyze. I will focus on the clusters that provide the most actionable insights to CMS.

Cluster 0 was characterized by low costs and almost full coverage of the bill by Medicare. These samples were all non-facilities (not hospitals), and were described as centralized flu facilities. Inspection of the HCPCS descriptions associated with this cluster confirm. All the drugs in this cluster are flu or pneumococcal vaccine related.

Cluster 1 also received almost full coverage by Medicare and was relatively low cost. This cluster was mostly family practice doctors giving flu shots. However, while Medicare did cover almost all of the bill, in comparison to flu shots administered by non-doctors of Cluster 0, Medicare’s expenditures per beneficiary were doubled.

Cluster 2 is comprised of flu shots provided by a roster biller. Or a large organization that provides mass immunizations. As such, costs are relatively low, particularly compared to the doctors of Cluster 1.

I recommend that Medicare encourage its recipients to go to mass immunizers or pharmacies for their flu immunizations. Medicare could decrease its expenditures on flu vaccinations considerably, as currently the majority of individuals receive their flu vaccine from a physician.

Cluster 15 is unique in that the average amount that Medicare covers is drastically lower than usual. Based on the difference between average number of services and the number of unique encounters, it seems that high cost drugs are administered multiple times a day for a given patient. The administered drugs (Aflibercept, Bevacizumab, and Ranibizumab primarily) are used to treat macular degeneration, which causes blindness. Since Medicare recipients are largely retirement age, massive out-of-pocket spending on healthcare is not sustainable. In this case, Medicare could improve their service level. CMS has two ways to address this shortcoming. Either negotiate down the price of drugs treating macular degeneration, or increase the payment amount to minimize burden on elderly patients with a chronic disease.

1. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Physician-and-Other-Supplier2017> [↑](#footnote-ref-1)
2. CMS Takes Action to Lower Prescription Drug Prices and Increase Transparency, CMS, 2019, <https://www.cms.gov/newsroom/press-releases/cms-takes-action-lower-prescription-drug-prices-and-increase-transparency> [↑](#footnote-ref-2)