Medicaid Drug Spending Report

Author: Bradley D. Presson

May 2024

Introduction:

Medicaid, a crucial government health insurance program, acts as a lifeline for certain low-income individuals and families in the United States. It supports people who meet eligibility criteria established by both federal and state laws, as exemplified in Maryland (HHS, 2021). After Medicaid-eligible patients receive healthcare and other benefits, Medicaid directly pays healthcare providers. This document offers a comprehensive data analysis focused on Medicaid spending, utilizing the drug data dictionary from the Centers for Medicare and Medicaid Services (Publisher Centers for Medicare & Medicaid Services, 2024). This investigation analyzes the total amount reimbursed based on the average spending per dosage unit of drugs by Medicaid and non-Medicaid entities to pharmacies.

The dataset used in this study was obtained from the U.S. Department of Health and Human Services through Data.gov, with the metadata last updated on March 15, 2024. It contains information on Medicaid-covered patient drug prescriptions, including costs associated with physicians and other healthcare providers (CMS, 2023). The dataset primarily examines average spending per dosage unit and tracks changes in this spending from 2018 to 2022 (Publisher Centers for Medicare & Medicaid Services, 2024). It also includes data on drug manufacturer expenditures and details about drug uses and clinical indications. Given these details, concerns were raised regarding factors influencing Medicaid drug spending.

COVID-19, a global health crisis, has significantly affected healthcare utilization among low-income populations in the United States using Medicaid (Roth et al., 2022; McGough et al., 2023). This can manifest in various ways, including health maintenance activities such as vaccines, annual wellness visits, and routine follow-up visits for medication management. By using Medicaid spending on covered outpatient medications as a proxy for healthcare utilization, this study analyzed trends in Medicaid drug spending from the pre-pandemic to the post-

pandemic period, emphasizing the immediate relevance of the research to the current healthcare landscape.

This analysis will use two hypotheses to explain further how the COVID-19 pandemic has impacted Medicaid expenditures and drug manufacturers. The first hypothesis suggests that the number of drug manufacturers for a medication does not influence Medicaid spending; alternatively, it proposes that the presence of more drug manufacturers is linked to increased Medicaid spending. The second hypothesis states that the compound annual growth rate (CAGR) was not significantly related to the average total Medicaid spending per dosage unit from 2018 to 2022; the alternative is that the change in CAGR is significant.

Methodologies:

Before beginning analysis on the dataset (medicaid_spending_by_drug_ data_dictionary), observations with missing values were removed. The data was then sorted in descending order by the number of "overall" manufacturers (Mftr_Name) for each drug, and duplicate entries for generic drug names (Gnrc_Name) were omitted. This process reduced the dataset from 16,459 observations to just 712. The data cleanup focused on macro-level metrics rather than individual micro-level data, enabling comprehensive analysis with various statistical learning models.

The initial analysis used a multiple linear regression model to examine the relationship between total manufacturers (Tot_Mftr) and average spending (Avg_Spending). The data was modeled with Avg_Spending as the dependent variable and Tot_Mftr as the independent variable. Additionally, average spending was identified as the variable that required further exploration by adjusting the total number of manufacturers. Reviewing the regression model summary will provide insights into the connection between the two variables and help predict future average spending based on the total number of manufacturers. It may also account for variances in the dataset that are not apparent from simply examining the raw data.

After conducting the baseline multiple linear regression, the data were rearranged into a data frame and visualized with the ggplot function. A scatter plot showed the relationship between

year and average spending per claim. Next, a linear regression model was fitted to measure this relationship and how it has evolved over the years.

Next, a multiple logistic regression model was used to predict the direction of Medicaid drug spending. Specifically, the model estimated the probability of an upward shift in the CAGR of the average total spending per dosage unit of drugs from 2018 to 2022. Using the prefiltered dataset, a subset selection was performed before fitting the logistic regression, with Mallow's Cp serving as the preferred method to determine the optimal number of predictors. Mallow's Cp was chosen over the Bayesian Information Criterion (BIC) because it indicates a less complex model. The technique revealed that three out of five coefficients (related to average Medicaid drug spending per dosage unit between 2018 and 2022) significantly impacted the model's accuracy. A backward subset selection identified that the average spending per dosage unit from 2019 to 2021 was the most important predictor for the logistic regression's accuracy in forecasting the CAGR direction of Medicaid drug spending. By cleaning the dataset and selecting an appropriate subset of predictors, a robust logistic regression model was developed to improve the accuracy of predicting the direction of CAGR in Medicaid drug expenditure.

CAGR, average spending per dosage unit from 2018 to 2022, was categorized as responses "Up" and "Down" depending on whether the average was positive or negative, respectively. After recalculating probabilities using CAGR values, a geometric mean was computed from all 712 observations. The geometric mean of the direction probabilities was approximately 0.54. The prediction model was trained on a random sample of 356 observations without replacement to improve the accuracy of predicting the CAGR direction. The remaining 356 observations were used as test data to evaluate the model's accuracy. The CAGR direction was classified as "Up" if the direction probability exceeded 0.54. After running the prediction model, the next step was to analyze the relationship between total manufacturers (Tot_Mftr) and average total spending per dosage unit (Avg_Spending) from 2018 to 2022.

The final analysis used a hierarchical clustering model. This method examined the relationship between the number of manufacturers and the change in year-over-year growth rate; as a result, it revealed trends that are often not visible when looking at the entire dataset. To perform this

analysis, the first step was to calculate the compound annual growth rate (CAGR) for each year using an iteration function to compute CAGRs for each data line item. Next, a combination function was applied to add the new columns back into the original dataset. To simplify clustering, a filter was then applied to focus only on the top and bottom 10% of the data (1.65 standard deviations from the mean) in terms of total manufacturers. This reduced the observations from 712 to 143. Focusing only on the extremes helps better highlight differences and reduce noise. After selecting this subset, the hclust() function performed hierarchical clustering with complete linkage. Two different plots were generated: one dendrogram displaying total CAGRs by the number of manufacturers, and separate scatter plots to further explore the relationship between individual yearly CAGRs and the total number of manufacturers.

Results

As a result of the data analysis, a multiple regression model examined the relationship between the year and average spending per claim (Figure 3A). The regression analysis showed a stronger correlation. With an adjusted R-squared value of 0.5929, the model explains about 59.29% of the variation in average spending per claim, indicating it is a better predictor of Medicaid spending patterns. The positive coefficient of around $1.27 annually suggests that average spending per claim has steadily increased each year.

Another multiple regression analysis was conducted to examine the effect of the number of manufacturers on average Medicaid spending (Figure 3B). The findings showed a significant association, with each additional manufacturer linked to an increase in average spending of about $1,799,386. However, the model had a low adjusted R-squared value of 0.002683, indicating that the number of manufacturers explains only a small part of the variation in average spending.

The subset selection, logistic regression, and prediction model produced interesting results in the study. Mallow's Cp criterion identified a subset of three out of five predictors as the best for the logistic regression model predicting CAGR in average total spending per dosage unit of drugs between 2018 and 2022 (Figure 2A). Using the coefficient values for this three-predictor subset,

the average total spending per dosage unit (weighted) for 2019 to 2021 was re-evaluated, showing these as the top predictors for a logistic regression model forecasting an upward trend in CAGR for the same measure (Figure 2B). According to the prediction model outlined in the methodology, a confusion matrix indicated that the model was 69% accurate in predicting CAGR direction, based on a training set comprising 50% of the data (356 observations). Importantly, 90% of these predictions correctly identified an upward trend in CAGR for average total spending per dosage unit of drugs, using data from 2018 to 2022 (Figure 2C). Results from the logistic regression and prediction model demonstrated that Medicaid spending from 2019 to 2021 was the strongest predictor of CAGR direction. Despite some noise in the data, the model suggested an upward trend in CAGR for average total spending per dosage unit of drugs approved by Medicaid. Although it was hypothesized that CAGR was not significantly associated with average total Medicaid spending per dosage unit from 2018 to 2022, findings showed that 2019 to 2021 were significantly related to spending per dosage unit, thus supporting the alternative hypothesis.

The analysis of the hierarchical clustering method revealed several notable trends. Notably, the dendrogram in Figure 1A divided clusters based on total manufacturing volume. The numbers in the chart correspond to the row numbers in the dataset. The higher the number, the more manufacturers are represented by the brand. Ultimately, the chart shows that brands with large numbers of manufacturers (indicated by larger numbers in the chart) share similar characteristics regarding the total CAGR between 2018 and 2022. This trend also appears in brands with fewer manufacturers; however, as you move up the branches, differences emerge between brands with only one manufacturer. Some of these brands have been grouped with brands that have more manufacturers. This suggests that certain brands with a single manufacturer are leveraging their market dominance, while others are following market trends. Figures 1B-1E further illustrate the relationship between yearly CAGRs and total manufacturers by brand. This data highlights a key trend: manufacturers are beginning to converge with year-over-year CAGR as the U.S. works to recover from a recession. Between 2021 and 2022, data shows variation among brands compared to the baseline, with more convergence observed in brands with higher manufacturer totals. In summary, brands with fewer than five manufacturers show greater variance in year-over-year

CAGR, indicating that the number of manufacturers does not significantly impact the average total spent per dose.

Discussion:

The multiple regression analysis for several manufacturers and CAGR on Medicaid drugs consistently showed an increase of $1,799,386 in average spending for each additional manufacturer. When reviewing the adjusted R-square value, the number of manufacturers does influence spending; however, its impact is relatively minor compared to other factors, which may include external market dynamics or regulatory changes. Changes over time are more significant in determining how much Medicaid spends on drugs per claim. A comprehensive approach is essential in developing robust strategies to control costs. This integrated strategy aligns with findings by Wiener et al. (2017), who argue that no single strategy will suffice for Medicaid cost containment, emphasizing the importance of a multifaceted approach.

The number of Medicaid users reporting delays or skipping medical care due to costs or the pandemic increased, mostly in 2021 (McGough, 2023). Another study found that new Medicaid enrollees during the COVID-19 pandemic were less likely to use primary care, experience avoidable hospitalizations, or incur healthcare costs; they also used fewer healthcare services than before the pandemic (Wright et al., 2021). Based on the available literature, the findings from this study contradict the Medicaid spending dataset from the Drug Data Dictionary published by the Centers for Medicare & Medicaid Services.

Mallow's Cp criterion for subset selection indicates that three different predictors offer a more accurate model for the logistic regression. It was shown through backward subset selection that the total average Medicaid drug spending from 2019 to 2021 was the three most important predictors for determining CAGR direction in the logistic regression and predictor models. There was a 90% chance of predicting an upward movement in CAGR for total average Medicaid drug spending based on 356 observations in the test data. These findings suggest that the highest levels of spending occurred during the COVID-19 pandemic years. Although 2022 was not identified as a significant predictor of the upward trend in CAGR, the prediction model also suggests a future increase in CAGR for Medicaid drug spending.

The hierarchical clustering between the total number of manufacturers and CAGR on Medicaid drug spending showed that as the number of manufacturers increased for drugs, the year-over-year growth rates for CAGR on drug spending by Medicaid declined from 2018 to 2022. The analysis indicates that the variation in annual growth rates in drug spending decreased as the number of manufacturers grew.

The large number of Medicaid users who reported delaying or skipping medical care, as mentioned in other studies, did not result in a decline in the CAGR for overall Medicaid drug spending. These findings suggest that an increasing CAGR may be unrelated to the pandemic and instead reflect rising healthcare costs. The trend likely indicates broader changes in the healthcare environment, such as policy shifts or economic factors, which significantly influence spending over time.

The government should seize the opportunity to renegotiate contracts with drug manufacturers through the Inflation Reduction Act of 2022 and leverage brands with multiple manufacturing partners to reduce overall unit prices (Meyers et al., 2023). However, aggressively negotiating drug prices might discourage manufacturers from investing in research and development, so there is a delicate balance between short-term savings and long-term innovation.

Limitations:

Through the analysis of the data, some limitations were found:

First, the dataset is likely derived from claims data from CMS, which typically takes weeks to months to become available for all payors (Strom, 2001). Second, the dataset is based on outpatient prescriptions and does not include any medications administered during inpatient stays or emergency room visits. Third, the dataset was heavily filtered because of the significant variation in qualitative entries of medication names, even when they are the same type of drug. While cleaning the data for analysis, over 15,000 observations were removed due to

redundancies and inconsistencies in data entry by CMS, which impacted the accuracy of the estimates in the analysis.

References

Centers for Medicare & Medicaid Services (CMS). *Medicaid Spending by Drug Methodology*. Centers for Medicare & Medicaid Services Data. (2023, February 28). https://data.cms.gov/resces/medicaid-spending-by-drug-methodology

McGough, M., Amin, K., & Twitter, C. C. (2023, January 31). *How has healthcare utilization changed since the pandemic?* Peterson-KFF Health System Tracker. https://www.healthsystemtracker.org/chart-collection/how-has-healthcare-utilization-changed-since-the-pandemic/#Percent%20of%20adults%20(age%2018%20years%20and%20older)%20who%20reported%20delaying%20or%20going%20without%20medical%20care%20due%20to%20COVID-19%20pandemic,%202021

Publisher Centers for Medicare & Medicaid Services. (2024, March 15). *U.S. Department of Health & Human Services - Medicaid spending by drug*. Catalog. https://catalog.data.gov/dataset/medicaid-spending-by-drug-b6f77

Roth, S. E., Govier, D. J., Marsi, K., & Cohen-Cline, H. (2022). Differences in Outpatient Health Care Utilization 12 Months after COVID-19 Infection by Race/Ethnicity and Community Social Vulnerability. *International Jnal of Environmental Research and Public Health*, *19*(6). https://doi.org/10.3390/ijerph19063481

Rudowitz, R., Williams, E., & Burns, A. (2023, April 14). Medicaid financing: The basics. KFF. https://www.kff.org/medicaid/issue-brief/medicaid-financing-the-basics/

Shah, S., Meyers, J., Kirby, M., & Chen, L. (2023, September 13). *Navigating the inflation reduction act's impact on Drug Pricing and innovation*. BCG Global. https://www.bcg.com/publications/2023/navigating-inflation-reduction-act-impact-on-drug-pricing-innovation

Strom, B. L. (2001). Data validity issues in using claims data. *Pharmacoepidemiology and Drug Safety*, *10*(5), 389–392. https://doi.org/10.1002/pds.610

The future of pharmaceuticals - dnesday, June 15, 2022. FTC. (2022, June). https://www.ftc.gov/system/files/ftc_gov/pdf/The-Future-of-Pharmaceuticals-dnesday-June-15-2022.pdf

US HHS. (2021, November 1). *What is the Medicaid program?* HHS.gov. https://www.hhs.gov/ansrs/medicare-and-medicaid/what-is-the-medicaid-program/index.html

Wiener, J. M., Romaire, M., Thach, N. T., Collins, A., Kim, K., Pan, H., Chiri, G., Sommers, A., Haber, S., Musumeci, M., & Paradise, J. (2017). Strategies to reduce Medicaid spending: Findings from a literature review. *Kaiser Family Foundation.* Retrieved from: https://www.kff.org/report-section/strategies-to-reduce-medicaid-spending-findings-from-a-literature-review-key-findings/

Wright, B., Anderson, D., Whitaker, R., Shrader, P., Bettger, J. P., Wong, C., & Shafer, P. (2021). Comparing health care use and costs among new Medicaid enrollees before and during the covid-19 pandemic. *BMC Health Services Research*, *21*(1). https://doi.org/10.1186/s12913-021-07027-6

Appendix A: Hierarchical Clustering: CAGR v.s. Total Manufacturer

Figure 1A: 2018 to 2022 CAGR v.s. Manufacturer #'s

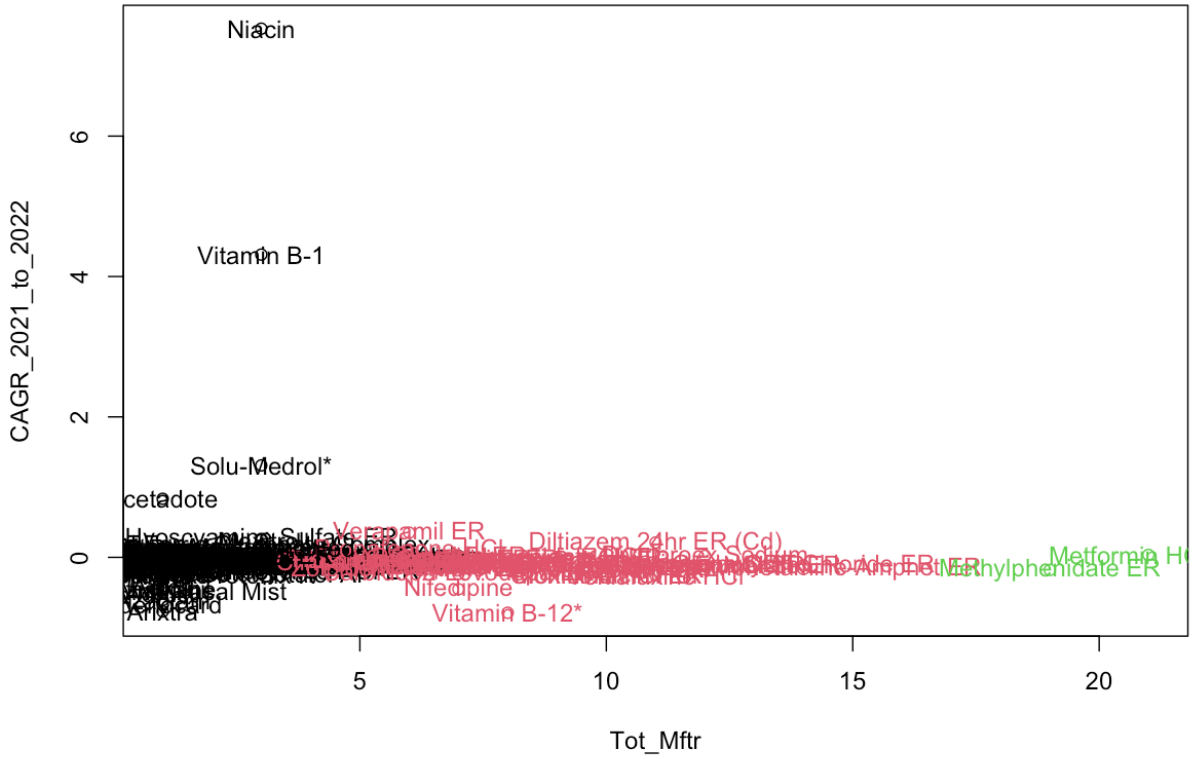

Figure 1B: CAGR v.s. Manufacturer #'s from 2020 to 2021

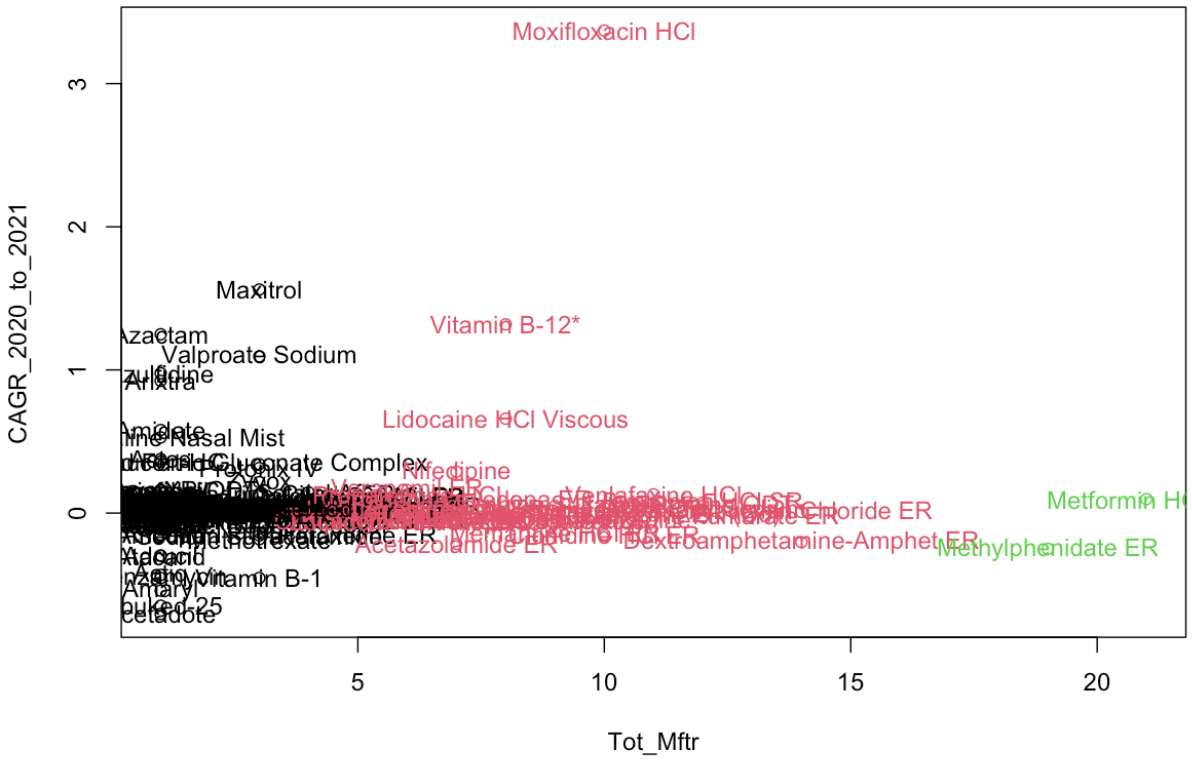

Figure 1C: CAGR v.s. Manufacturer #'s from 2021 to 2022

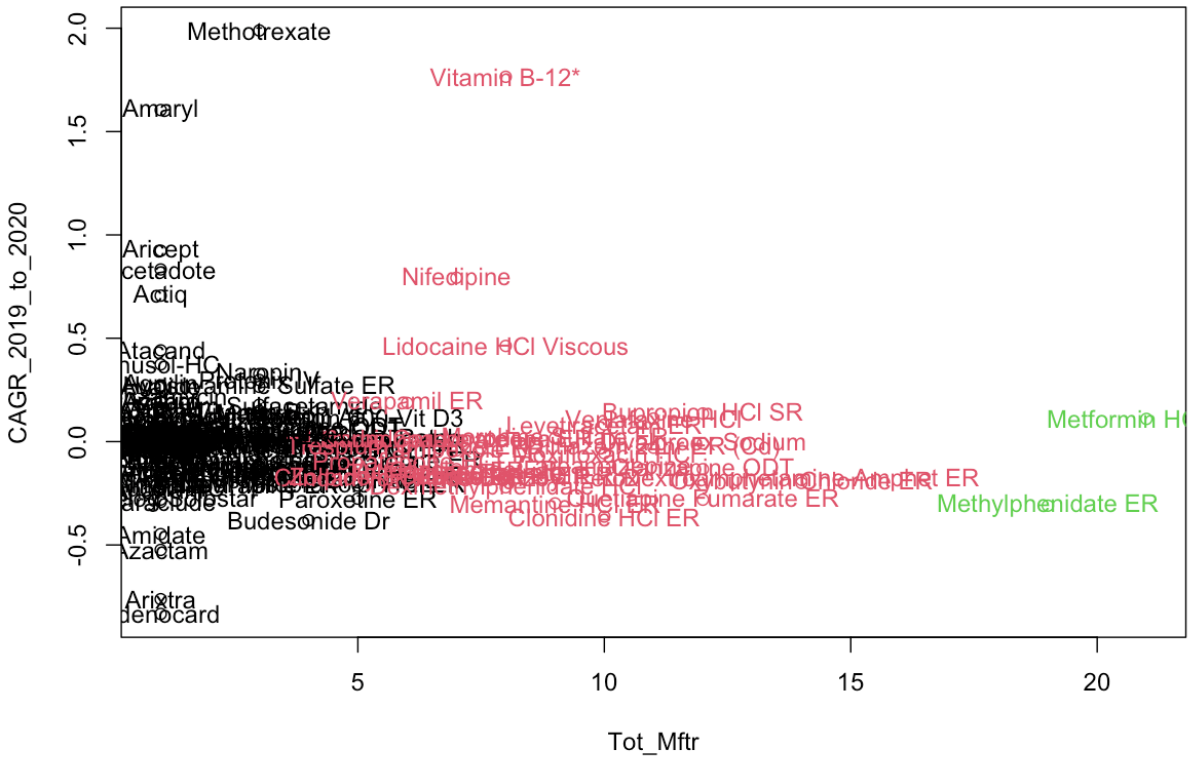

Figure 1D: CAGR v.s. Manufacturer #'s from 2019 to 2020

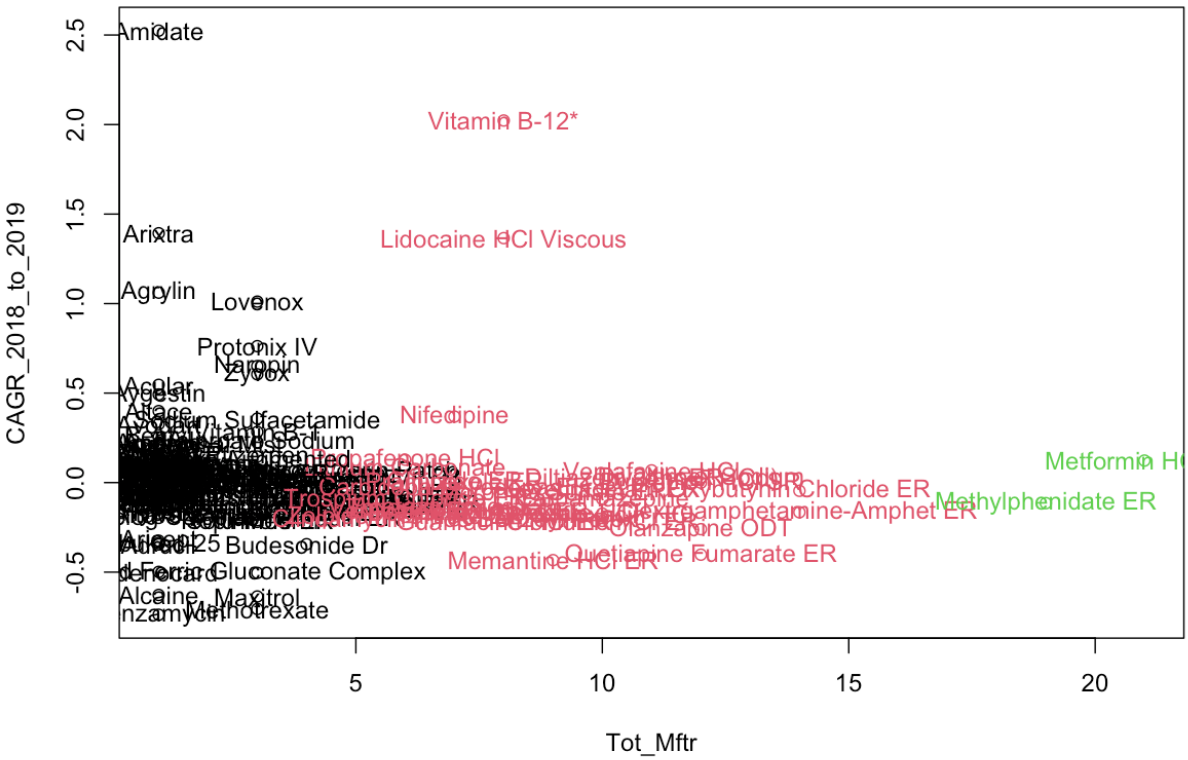

Figure 1E: CAGR v.s. Manufacturers #'s from 2018 to 2019

Appendix B: Subset Selection, Logistic Regression, and Prediction Model

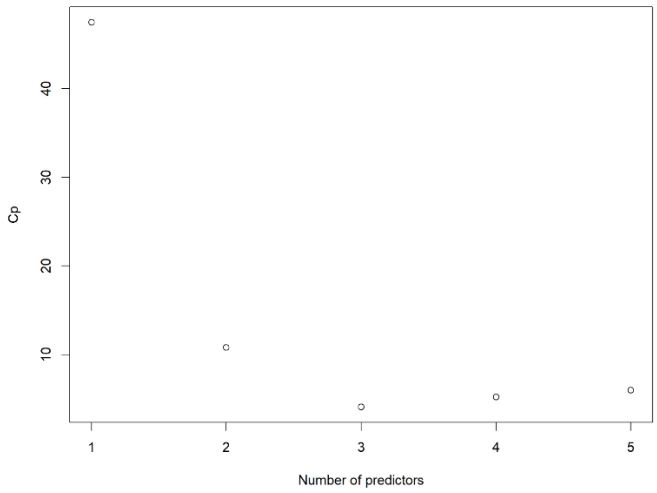

Figure 2A: Mallow's Cp Criterion

```
                    (Intercept) Avg_Spnd_Per_Dsg_Unt_Wghtd_2019
                   0.0579175339                   -0.0008038362
Avg_Spnd_Per_Dsg_Unt_Wghtd_2020 Avg_Spnd_Per_Dsg_Unt_Wghtd_2021
                   0.0003798873                    0.0003949070
```

Figure 2B: Subset selection for regression model in CAGR direction.

```
          CAGR_Direction.test
glm.pred Down  Up
    Down   24   4
    Up    105 223
```

Figure 2C: Confusion matrix for predicting CAGR spending direction

Appendix C:

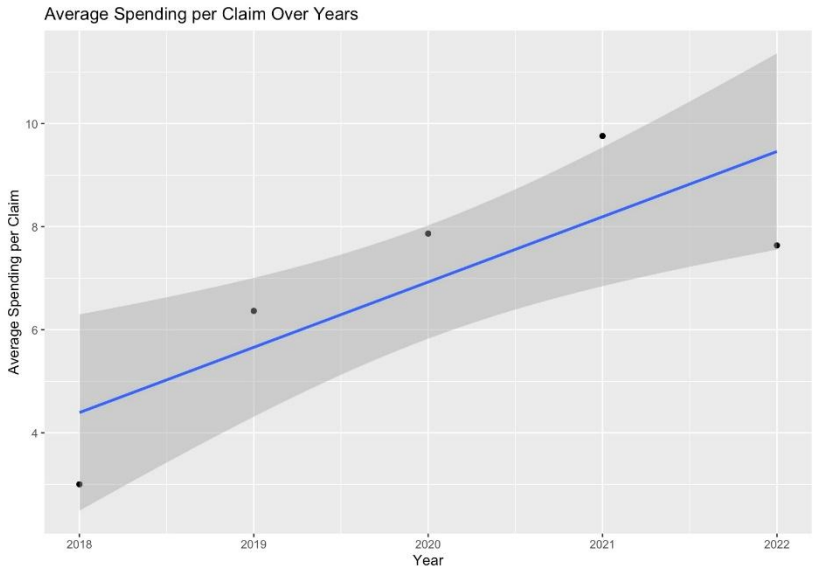

Figure 3A: Claims spending v.s. years (2018-2022)

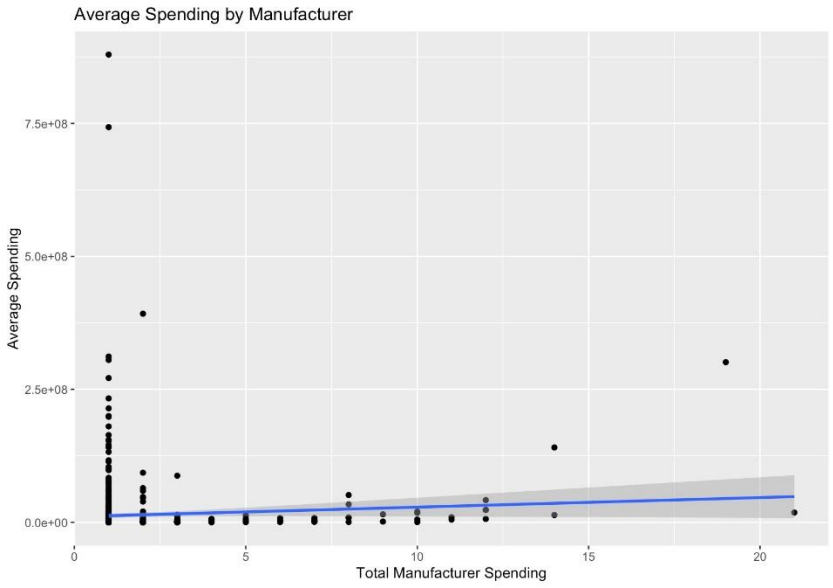

Figure 3B: Drug spending v.s. Manufacturer numbers