Title:

Analyzing ‘Medicare Part D Drugs’ Spending Trends and Implications in relation to Diabetes

Introduction:

This project delves into Medicare Part D spending, specifically focusing on diabetic medications. The purpose of this analysis is to unveil spending patterns, identify outliers, and provide insights that can inform policy decisions, optimize resource allocation, and enhance care within the Medicare system.

Diabetes is a big problem for medicare as it is a major prevalent chronic condition. According to an article by CDC, 61% of diabetes costs are for adults aged 65 or older, which is mainly paid by Medicare(CDC, 2020). Hence, understanding its financial aspects of diabetes related medications is crucial for the stakeholders at Medicare. By examining the costs associated with diabetic medications, we aim to offer valuable information to Medicare stakeholders, contributing to informed decision-making and program effectiveness.

Objectives:

* Objective 1: To identify how many diabetic medications are within the dataset and to assess the proportion of diabetic medication to the remainder of the medication and how it affects the spending of Medicare Part D Program.
* Objective 2: To identify diabetic medications with outlier spending patterns within the Medicare Part D program, which may have a significant impact on the program's budget.
* Objective 3: To assess the relationship between the number of drug manufacturers for diabetic medications and drug spending within this specific category.
* Objective 4: To examine the annual trends in Medicare Part D drug spending for diabetic medications from 2017 to 2021, including changes in average spending
* Objective 5: To provide actionable insights and recommendations for policymakers to optimize drug spending on diabetic medications within the Medicare Part D program.

Problem Statement:

We are focusing on understanding the challenges posed by rising costs, outlier spending patterns, manufacturer influence, and long-term trends in diabetic medication spending within the Medicare Part D program. With an emphasis on diabetic drugs, the analysis of Medicare Part D spending data from 2017 to 2021 aims to offer actionable insights and suggestions that will improve the program's effectiveness, increase its financial sustainability, and raise beneficiary care standards. This project will give decision-makers the information they need to make wise choices and promote improvements to the Medicare program.

Scope:

* This project focuses specifically on diabetic medications within the Medicare Part D program.
* The list of drugs to filter the dataset for Diabetes specific drug will be taken from the official government booklet, Medicare Coverage of Diabetes Supplies, Services, & Prevention Program. (*Large Print Edition*, n.d)
* It includes the analysis of spending patterns, beneficiary data, and drug categories related to diabetes treatment from 2017-2021.
* The project will take 70 days to complete, with two data consultants working for approximately 3 hours per week.

Exclusions:

* The project will exclude the drugs that are not mentioned in the resource we will use to reference Diabetes specific drugs.
* It will also exclude null entries since we will not have enough data for taking them into account.
* Prescription drug claims identified as "over-the-counter" are excluded by the Centers for Medicare & Medicaid Services (CMS). Hence, will be excluded from our project.

Limitations:

* The dataset represents only a subset of Medicare beneficiaries enrolled in Part D.
* All of the spending data for a drug that has fewer than 11 claims is redacted.
* When the drug information is based on more than 11 claims but less than 11 unique beneficiaries, Total Beneficiaries are redacted since that measure uses the unique beneficiary count as its denominator.
* The analysis is limited to the data available from 2017 to 2021; does not account for more recent data.
* Does not consider the broader context of healthcare policy changes beyond the scope of Medicare Part D.
* We are still learning visualization tools, so the final results might not be perfect in their representation.
* The Part D spending metrics do not reflect any manufacturers’ rebates or other price concessions as CMS is prohibited from publicly disclosing such information.

Data and Methodology

Data Source:

* Primary data source is Medicare Part D spending data for diabetic medications taken from Centers for Medicare & Medicaid Services (CMS). It is taken from Data.CMS.gov
* The data is in a “.csv” format, i.e Comma Separated Values.
* The Medicare Part D Spending by Drug dataset focuses on average spending per dosage unit and change in average spending per dosage unit over time.
* [Link to data](https://data.cms.gov/summary-statistics-on-use-and-payments/medicare-medicaid-spending-by-drug/medicare-part-d-spending-by-drug)
* Data is sourced from publicly available datasets.

Data Collections(*Medicare Part D Spending by Drug Methodology*, 2022):

* Medicare Part D PDE records were summarized by drug by linking National Drug Codes (NDCs) available in the PDE data to a commercially available database and aggregated across all strengths, dosage forms, and routes of administration to the drug brand name and generic name.
* Drug spending metrics for Medicare Part D drugs are based on the gross drug cost, which includes ingredient cost, dispensing fees, sales tax, and applicable vaccine administration fees. Medicare Part D drug spending represents total spending for the prescription claim, including amounts paid by the Medicare Part D plan and beneficiary payments.
* For each NDC in the PDE-level data, lower and upper bounds for average cost per dosage unit are defined as: i. Lower Bound: 25th percentile – 1.5\*Interquartile Range ii. Upper Bound: 75th percentile + 1.5\*Interquartile Range
* Data cleaning will be done using python, pandas library and visualizations will be made using pandas, Seaborn and Matplot lib in python.
* Data includes information about drug spending, dosage units, claims, beneficiaries, and drug categories.

Data Overview and summary statistics:

* It has 13,751 rows and 46 columns.
* MetaData :
  + Brnd\_Name: Brand name, Trademark name that is filed
  + Gnrc\_Name: Generic Name of drug, chemical ingredient of a drug
  + Tot\_Mftr: Total Manufacturers of the drug
  + Mftr\_Name: Name of manufacturers.
  + Tot\_Spndng\_Year: Aggregate drug spending for the Medicare Part D program during the benefit year.
  + Tot\_Dsg\_Unts\_Year: The sum of the dosage units of medication dispensed across the calendar year
  + Tot\_Clms\_Year: Number of prescription fills for each drug.
  + Tot\_Benes\_Year: Number of Part D beneficiaries utilizing the drug during the benefit year.
  + Avg\_Spnd\_Per\_Dsg\_Unt\_Wghtd\_Year: Medicare Part D drug spending divided by the number of dosage units, which is weighted by the proportion of total claims.
  + Avg\_Spnd\_Per\_Clm\_Year: Part D drug spending divided by the number of prescription fills.
  + Avg\_Spnd\_Per\_Bene\_Year: Total Part D drug spending divided by the number of unique beneficiaries utilizing the drug during the benefit year.
  + Outlier\_Flag\_Year: Yearly outlier flag variable, which is set to “1” when a drug’s Average Spending per Dosage Unit is substantially impacted by outlier records in a given year.

Data Types

* Brnd\_Name, Mftr\_Name, Gnrc\_Name: object
* Tot\_Mftr, Outlier\_Flag\_2021: int64
* Total Spending, Total Dosage Units, Total Claims, Total Beneficiaries, Outlier Flags, Avg Spending/Dosage Unit, Average spending per claim, Average spending per beneficiary for the years 2017-2021 are Float Data Types.
* There are a total of 95534 null values in the entirety of the dataset, but since it is a very large dataset, mentioning the values here is not sensible. However, we will be showing them in the code we write for the Data Overview section.
* There are 13,751 rows in the dataset and 46 columns in the original dataset.

Data Preprocessing:

* Involves cleaning data, handling missing values and null values, and ensuring data consistency.
* May use split, apply and combine to the dataset to get specific value sets from the data source.
* The data will be split into two dataframes, which has the overall data on each generic drug and another with data for all the diabetes drugs from all the manufacturers.
* Also use methods taught in class over the duration of the semester.

Expected Outcomes and Future Implementation

Expected Outcomes:

* The policymakers and the government health agencies will use this analyses to make informed policy decisions and improve the Medicare Part D program. This will help enhance the affordability and accessibility of diabetes medications for beneficiaries.
* It will provide insights into how spending per beneficiary differs across various diabetes treatments, which will help providers and insurers in recommending cost-effective treatments to beneficiaries.
* Because of its reproducibility, the analyses can be used for further projects and research on the spending and medication trends.

Deliverables:

* Presentation for stakeholders to summarize the keyfindings and insights in a non-technical manner for stakeholders.
* Detailed report with collection methods, methodology and potential visualizations showing trends in Diabetes medication in Part D Medicare.
* Visualizations will be made using Python libraries like Pandas, Seaborn and Matplot.

Future Implementations:

* Can be used to inform Medicare policy decisions related to diabetic medications by the government agencies. For eg. Use it for analysis of medicaid drug datasets.
* The analyzes also be used by the pharmaceuticals for their future projects and finding potential areas of research for drugs with limited manufacturers and higher spending.
* Hypothetically can involve adjusting Medicare Part D coverage for diabetic medications based on the analysis and use it to negotiate prices with the pharmaceutical companies.
* Can use the analysis tools with recent data sets to help improve real time decisions made regarding coverage thereby improving healthcare access and cost savings.

Ethical Considerations and Risk Assessment

Ethical considerations:

* Clearly documenting data sources and methodologies to maintain transparency and trust
* Ensuring patient data is anonymized and protected.
* Avoiding any bias in the analysis that could impact policy decisions unfairly.

Ethical Issues out of hand:

* Potential biases in the original data collected by the Medicare Part D Program. This could be based on gender, region, race, etc.

Mitigation Strategies:

* Implementing data validation and cleaning processes.
* Clearly documenting any assumptions or limitations in the analysis.

Risks:

* Data quality and availability
* Changes in healthcare policies that may affect the data.
* Technical challenges in data processing and analysis.

Risk Mitigation:

* Monitor data source for updated data.
* Staying informed about policy changes and adapting into our analysis.
* Using alternative methods for our data analysis to obtain needed outcomes.

—--------------------------------------------------------------------------------------------------------------

References:

1. *Medicare Part D Spending by Drug Methodology*. (2022). <https://data.cms.gov/sites/default/files/2022-01/DSD_PTD_R21_20220118_Methodology_WDDSE.pdf>
2. *Large Print Edition*. (n.d.). <https://www.medicare.gov/Pubs/pdf/11022-LE-Medicare-Coverage-of-Diabetes.pdf>
3. CDC. (2020, September 29). *Cost-Effectiveness of Diabetes Interventions | Power of Prevention*. Www.cdc.gov. https://www.cdc.gov/chronicdisease/programs-impact/pop/diabetes.htm

Project Members:

1. Preksha Shah
2. Mohammad Awad