Team: Eliza Loke Email: eloke@usc.edu USC ID: 9840-3757-42

Prescribing Trends of Antidepressants in the U.S. Using the CMS Medicare Part D API

Abstract

This study sought to quantify the geographical variation in the prescribing of five core Selective Serotonin Reuptake Inhibitors (SSRIs) across the five most populated U.S. states using 2023 public data from the Centers for Medicare and Medicaid Services (CMS) Medicare Part D Prescribers API. Over 26 million individual provider drug prescription records were systematically retrieved and aggregated by state and standardized drug groups, using Python's requests and pandas libraries. Analysis of the resulting 25 row summary table revealed significant geographical concentration, validating the research question. The findings establish Florida as the dominant prescribing state, leading in total claims and recording the highest frequency for four out of the five analyzed SSRIs (Sertraline, Escitalopram, Paroxetine, and Citalopram). This suggests a regionalized prescribing intensity that warrants further investigation into state specific public health factors and resource allocation strategies within the Medicare system.

1. Introduction

The study of antidepressant prescribing patterns provides insights into regional mental health needs and healthcare utilization. Significant geographical variations in the uptake of Selective Serotonin Reuptake Inhibitors (SSRIs) is a necessary quantitative assessment to inform public health policy and resource allocation.

This project utilizes the 2023 Centers for Medicare and Medicaid Services (CMS) Medicare Part D Prescribers dataset to analyze prescription claims for five commonly used SSRIs: Sertraline, Fluoxetine, Escitalopram, Paroxetine, and Citalopram. The scope is strategically focused on the five most populated U.S. states to ensure the findings are representative of a substantial portion of national prescribing volume.

The central research question guiding this report is: Which U.S. states exhibit the highest prescribing frequency for each selected antidepressant within the Medicare Part D program in 2023?

2. Methods

2.1 Data Source and Collection

The data for this analysis was systematically retrieved from the Centers for Medicare and Medicaid Services (CMS) Medicare Part D Prescribers - By Provider and Drug API specifically utilizing the 2023 public dataset. The collection process, executed by the get_data.py script, uses Python requests library for HTTP GET requests, targeting the five selected SSRI drug groups, includes querying both Generic and Brand names (e.g., Sertraline and Zoloft) to maximize data capture.

The scope was further refined to focus on five target states, comprehensive query method resulted in a raw dataset consisting of 26,794,878 individual provider prescription records, stored as a csv file (ssri_partd_2023_five_states_raw.csv) in the data/raw directory.

2.2 Data Cleaning and Preparation

Data cleaning and preparation were executed by the clean_data.py script to transform the large raw dataset into a structure suitable for analysis. The initial phase focused on filtering the data to the five target states and ensuring all prescription records corresponding to the five target SSRIs (Sertraline, Fluoxetine, Escitalopram, Paroxetine, and Citalopram) were retained. A critical preprocessing step involved standardizing key categorical columns, including drug names and state abbreviations, by converting all entries to uppercase and stripping whitespace. Following standardization, drug names were grouped by all brand names (e.g., Zoloft, Prozac) to their generic equivalents (e.g., Sertraline, Fluoxetine), confining inconsistencies and accurate aggregation. The final transformation involved aggregating the cleaned dataset. This process was achieved by grouping all records by State Abbreviation (State_Abrvtn) and Antidepressant Group (Antidepressant_Group), summing the Total Claims (Tot_Clms) for each group. This reduced the dataset from over 26 million records to an efficient 25 row-table, which was saved as a processed file (ssri_partd_2023_five_states_aggregated.csv) in the data/processed/ directory for further analysis.

2.3 Deviations from Initial Proposal

The project execution required a strategic deviation from its initial scope in terms of geographical variation. Although initial intent was to analyze prescription trends across all U.S. states, to manage the retrieval and processing of over 26 million records led to a focused reduction in scope. Restricting to the five most populated states allowed for efficient data management, maintaining a representative national prescribing volume essential for a valid comparative analysis.

3. Results and Analysis

The final aggregated dataset was subjected to comparative statistical analysis using the run_analysis.py script. This analysis primarily utilized the pandas library functions, including groupby(), sum(), and pivot() to restructure the data, and rank() to determine the relative prescription frequency of states and drugs. The findings are categorized into three areas: Total SSRI Prescriptions by State, Overall Market Share of 5 SSRIs, and Total SSRI Claims Breakdown by State, which directly addresses the core research question.

3.1 Key Findings

3.1.1 Findings 1. Total SSRI Prescriptions by State (2023)

The states were ranked based on the combined total claims (Tot_Clms) for all five SSRIs. Florida (FL) led the analysis with the highest overall total claims at 2,979,633, slightly surpassing California (CA) at 2,811,206. This is notable, as Florida, only the third most populated state, demonstrates a higher combined prescribing volume than California and the next two largest states, Texas (TX) and New York (NY). Pennsylvania (PA) recorded the lowest total prescription claims among the five selected states at 2,346,093.

Findings 2. Total SSRI Claims Breakdown by State (2023)

The core analysis identified the specific state that exhibited the highest prescription frequency for each individual SSRI, revealing a highly specialized prescribing trend. The results confirm significant geographical variation, but the pattern is driven by the consistent and overwhelming

dominance of Florida (FL), which records the highest prescribing frequency for four out of the five analyzed SSRIs. California (CA) leads the prescription volume for only one drug, Fluoxetine (Prozac). This suggests that the high total claim volume observed in Florida (Figure 1.) is not merely a statistical anomaly but a pattern of consistently higher per-capita prescribing across nearly the entire class of analyzed antidepressants.

3.1.2 Findings 3. Overall Market Share of 5 SSRIs (Total Claims)

When aggregating claims across all five states, Sertraline (Zoloft) was definitively the most prescribed SSRI, accounting for the largest total market share with 4,575,184 claims. Escitalopram (Lexapro) followed as the second most prescribed drug at 3,893,570 claims. Conversely, Paroxetine (Paxil) was the least prescribed SSRI among the five groups analyzed, recording only 867,143 claims, indicating a lower general preference in the Medicare Part D population across the states.

3.2 Data Visualization

Data visualization was executed by the visualize_results.py script, which utilized the matplotlib library to generate graphical representations of the analytical findings. All figures were saved to the results/visuals/ directory.

3.2.1 Figure 1. Total SSRI Prescriptions by State (2023)

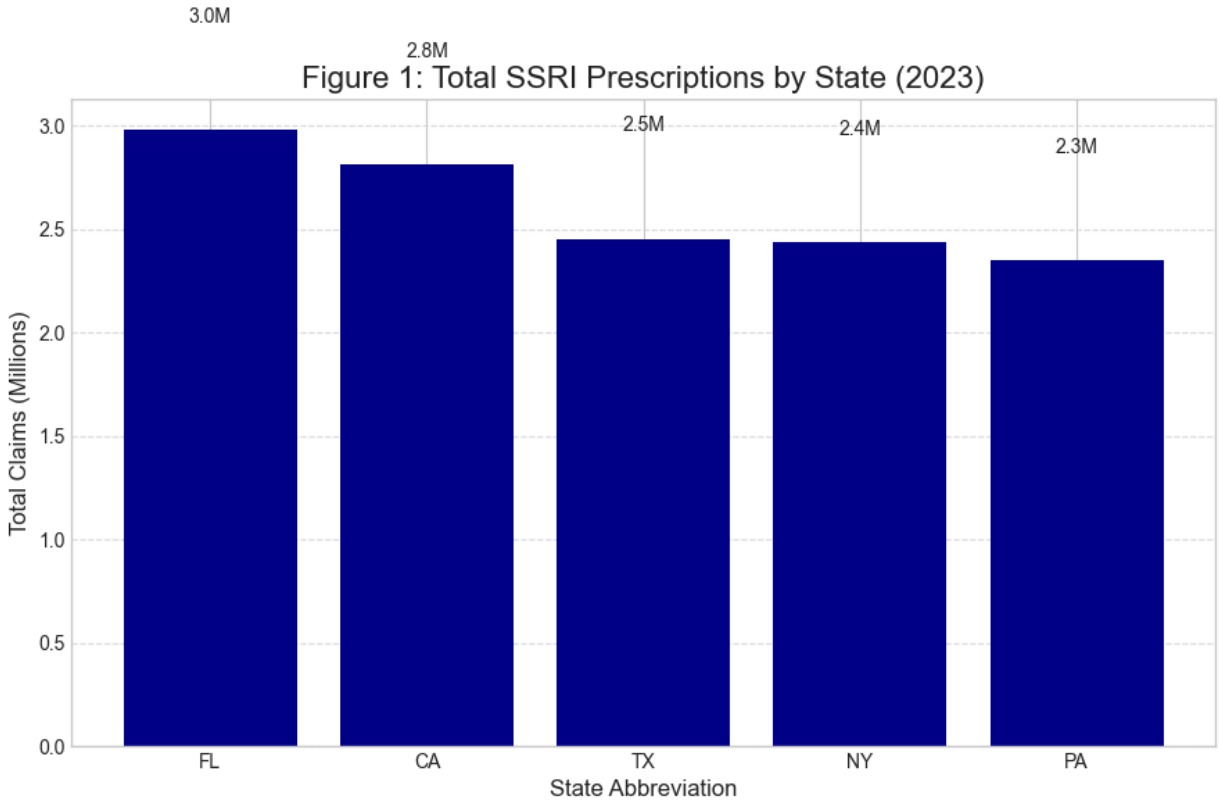


Figure 1 is a vertical bar chart that provides context for overall prescribing volume across the five target states. This visualization clearly illustrates that Florida (FL) and California (CA) are the two highest prescribers.

3.2.2 Figure 2. Total SSRI Claims Breakdown by State (2023)

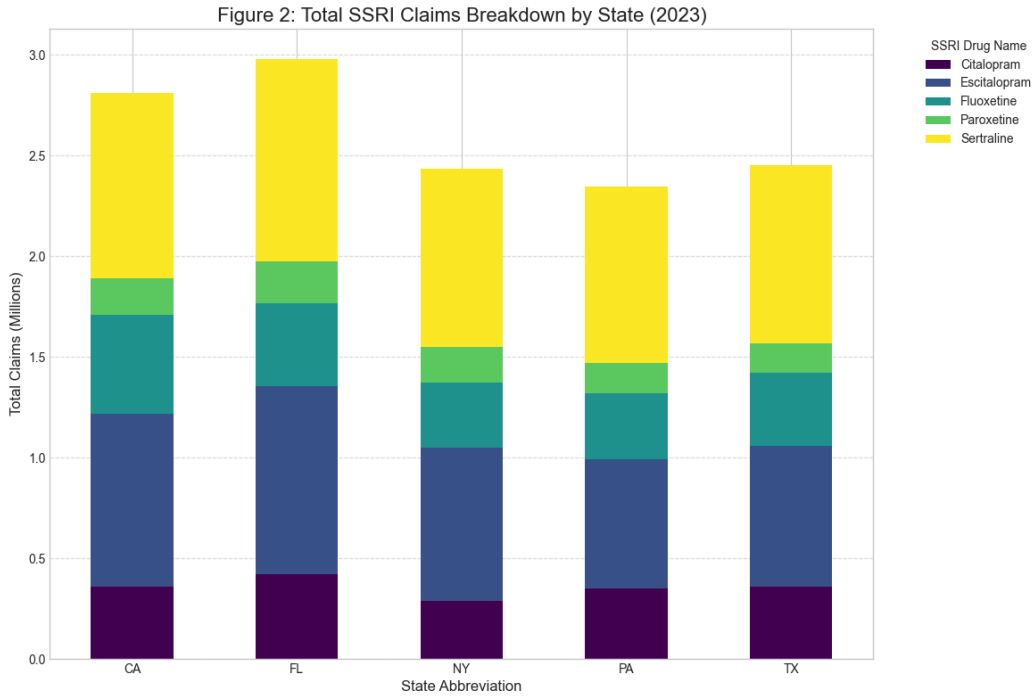
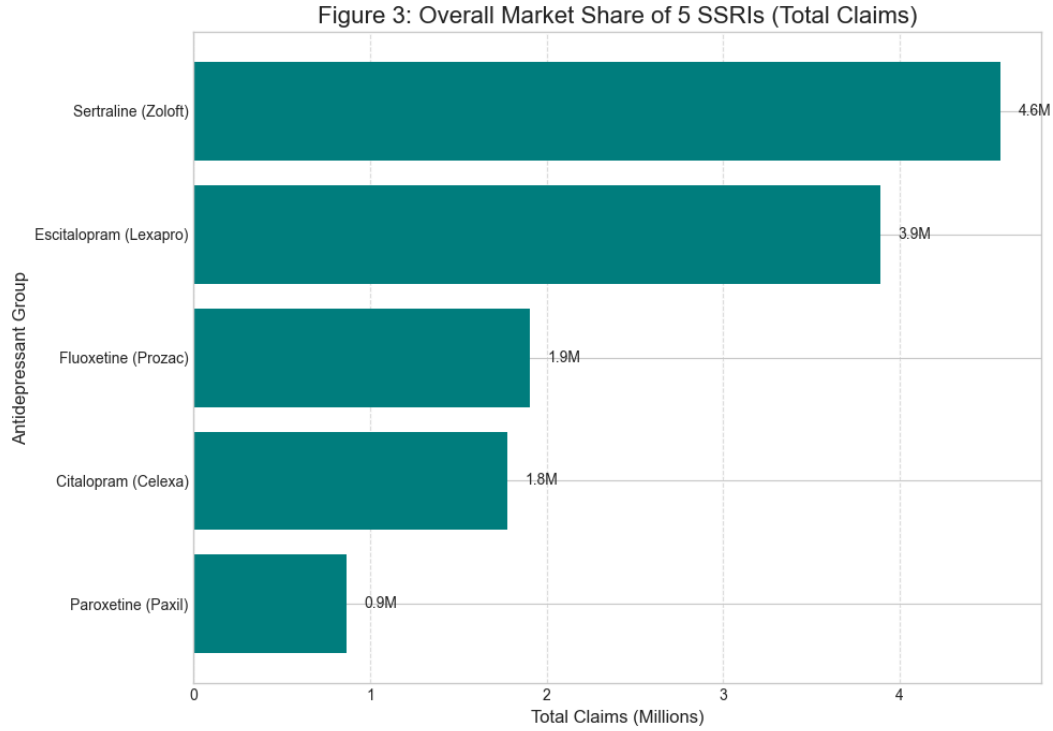


Figure 2, a stacked bar chart, is the primary visualization designed to answer the core research question regarding drug-specific geographical variation. Each state's total volume is represented by a single bar, segmented by the contribution of the five individual SSRIs. The chart visually confirms that Florida (FL) leads with Sertraline (Zoloft) and Escitalopram (Lexapro).



3.2.3 Figure 3. Overall Market Share of 5 SSRIs (Total Claims)\

Figure 3 is a horizontal bar chart presenting the overall market share of the five antidepressant groups across the entire analyzed region. The bars are ordered by volume, providing a clear visual representation of the drugs' relative popularity. This chart reinforces the analytical finding that Sertraline (Zoloft) and Escitalopram (Lexapro) form the dominant cohort of prescribed SSRIs, while the market share for Paroxetine (Paxil) is significantly smaller, illustrating the consensus preference hierarchy across all five states.

4. Conclusion

The core insight derived from the analysis and the Figure 2 visualization is the high intensity prescribing pattern exhibited by Florida (FL). The state not only registered the highest total volume of claims (as shown in Figure 1), but it demonstrated dominance in prescribing frequency for four out of the five analyzed SSRIs (Sertraline, Escitalopram, Paroxetine, and Citalopram). In contrast, California (CA) was the most active prescriber for only one drug, Fluoxetine (Prozac). This confirmed the existence of distinct, measurable geographical variations in drug preference, with the overall market hierarchy remaining consistent: Sertraline and Escitalopram are the dominant agents (as verified by Figure 3).

The established finding of Florida's consistently high prescribing rate across multiple SSRIs provides critical data for public health policy review. This geographical disparity warrants further investigation into regional differences in mental health prevalence or primary care practices within the Medicare population. This analysis provides a robust, quantitative baseline that can inform targeted resource allocation and serve as a necessary benchmark for future longitudinal studies tracking changes in U.S. antidepressant utilization.

5. Future Work

An extension of this project would involve transitioning the analysis from prescription volume to cost analysis, leveraging the Tot_Drug_Cst column within the raw dataset. Future work would focus on comparing the total number of claims (Tot_Clms) against the total cost, enabling the calculation of the average cost per prescription for each state and drug. This shift would provide essential data on cost effectiveness and economic disparities, allowing researchers to determine if high volume states maintain high costs or if they benefit from generic brand utilization. By analyzing cost volume, this approach could offer policy insights into the financial burden placed on the Medicare Part D system by geographical prescribing habits.