

## Project Proposal

Alysen Casaccio

### Opioid Treatment in the United States

December 14, 2023

#### Project Subject Area

My project focuses on Medicare providers and opioid prescribing patterns. I have obtained CMS datasets, which include data on Medicare Part D, Medicare Providers, Provider Service Lines, Opioid Prescribing by Geographic Area, and the Opioid Treatment Program.

Although this is likely to evolve as I explore the available data further, I am currently interested in determining if voluntary participation in the Opioid Treatment Program changes the prescribing behaviors of Medicare Providers above and beyond the average for geographic areas or service lines.

#### Data Sources

##### Flat Files

Name: *Opioid Treatment Program Providers*

Description: This is a CSV of all providers who are participants in the Opioid Treatment Program, providing information such as their enrollment date and other identifying information from the Provider Enrollment, Chain, and Ownership System (PECOS)

Shape: 1.4k rows, nine columns

Source: <https://data.cms.gov/provider-characteristics/medicare-provider-supplier-enrollment/opioid-treatment-program-providers>

Name: *CMS Program Statistics – Medicare Deaths*

Description: This is a CSV of all 2021 Medicare mortality rates for beneficiaries of both Original Medicare and Medicare Advantage, demographic characteristics, month of death, and geographic area of residence.

Shape: Seven tables not in machine-readable format, but typically in less than a 12 x 52 frame each.

Source: <https://data.cms.gov/summary-statistics-on-beneficiary-enrollment/medicare-and-medicare-reports/cms-program-statistics-medicare-deaths>

Name: *CDC Drug Overdose Death Rates*

Description: Drug overdose death rates by drug type, sex, age, and race in the United States.

Shape: 6,228 rows, 15 columns

Source: [https://data.cdc.gov/NCHS/Drug-overdose-death-rates-by-drug-type-sex-age-rac/95ax-ymtc/about\\_data](https://data.cdc.gov/NCHS/Drug-overdose-death-rates-by-drug-type-sex-age-rac/95ax-ymtc/about_data)

## API

Name: *Medicare Part D Prescribers – by Provider and Drug*

Description: Information on prescription drugs provided to Medicare beneficiaries enrolled in Part D coverage by physicians, aggregated by NPI and drug

Shape: 25.2m rows, 22 columns

Source: <https://data.cms.gov/provider-summary-by-type-of-service/medicare-part-d-prescribers/medicare-part-d-prescribers-by-provider-and-drug>

Name: *Medicare Part D Prescribers – by Geography and Drug*

Description: Information on prescription drugs provided to Medicare beneficiaries enrolled in Part D coverage by physicians, aggregated by geography and drug

Shape: 114.9k rows, 22 columns

Source: <https://data.cms.gov/provider-summary-by-type-of-service/medicare-part-d-prescribers/medicare-part-d-prescribers-by-geography-and-drug>

## Website Tables

Name: *Wikipedia – United States Drug Overdose Death Rates and Totals over Time*

Description: Tables of drug overdose death rates, by year, by state, and by country

Shape: Four tables with a maximum of 54 rows and 11 columns

Source:

[https://en.wikipedia.org/wiki/United\\_States\\_drug\\_overdose\\_death\\_rates\\_and\\_totals\\_over\\_time](https://en.wikipedia.org/wiki/United_States_drug_overdose_death_rates_and_totals_over_time)

## **Relationships**

The primary relationship I am interested in is prescriber behavior, so the key I can use to create a relationship across datasets is the national provider identity (NPI). The NPI exists in all datasets that refer to specific providers.

Additional keys I may utilize include state and drug name, depending on the robustness of the data I find. The state exists in all datasets that refer to geographic area, and the generic drug names exist across multiple datasets presented here. If I begin to explore data questions related to drug names, I will also include the “Beers List”, which is a list of medications whose clinical guidelines suggest caution when prescribing to patients over the age of 65. This further links this data project to Medicare, or at least to the Medicare-eligible population by age.

## Approach

My approach will be to utilize SQLite as a database due to its suitability in handling local data storage. Prior to loading data, comprehensive data cleaning and transformation will be performed. This includes addressing missing values and standardizing nomenclature or taxonomy where necessary. Planning an organized schema of tables and standard keys will be essential to success.

The data itself focuses on Medicare providers' prescribing patterns and comprehensive prescription data while supplementing this information with geographic and other demographic-focused death rates specific to drug overdose. Some of the datasets will be more comprehensive than needed, so I will plan to use filtering and grouping to better understand the connections and patterns within.

There are of course, always ethically severe implications of working with healthcare data, especially concerning patient privacy and the sensitive nature of opioid prescription data. Strict adherence to data privacy regulations and the identity of specific providers will be maintained. None of the datasets are anticipated to have personally identifiable information at the beneficiary level, but if found, it will be anonymized or aggregated to protect individuals' identities.

The project has many potential challenges, and in my experience with healthcare data, you don't often fully understand the challenges until the data is fully explored. I anticipate that handling such large datasets efficiently may be its own challenge. In addition, while I have some experience with SQL, I haven't set up my own database before, so there is likely to be a learning curve for the more technical maintenance aspects of the project.