Medicare Part D Prescribers Datasets: A Methodological Overview

April 2023

Prepared by:
The Centers for Medicare & Medicaid Services,
Office of Enterprise Data and Analytics

Table of Contents

Table of Contents 2
1. Background 3
2. Key data sources 3
3. Population 3
4. Data Contents 4
4.1. Detailed Data File 4
Part D Prescribers by Provider and Drug 4
4.2. Summary Tables 4
Part D Prescribers by Provider 4
Part D Prescribers by Geography and Drug 5
5. Data Limitations 5
Data Redaction and Suppression 6
6. Additional Information 7
7. Updates 7

1. Background

In an effort to make our healthcare system more transparent, affordable, and accountable, the Centers for Medicare & Medicaid Services (CMS) has prepared a public dataset, the Medicare Part D Prescribers Datasets (herein referred to as the "Part D Prescribers Dataset"), with information on prescription drug events (PDEs) incurred by Medicare beneficiaries with a Part D prescription drug plan. The Part D Prescribers Datasets is organized by National Provider Identifier (NPI) and drug name and contains information on drug utilization (claim counts and day supply) and total drug costs.

2. Key data sources

The primary data source for these data is the CMS Chronic Conditions Data Warehouse, which contains Medicare Part D PDE records received through the claim's submission cut-off date. The submission cut-off date is June 30th following the end of the preceding calendar year. For instance, the 2017 Part D Prescribers Dataset includes PDEs received through June 30, 2018. These data contain 100 percent of Medicare Part D final-action (i.e., all claim adjustments received through the cut-off date have been resolved) PDE records for beneficiaries who are enrolled in the Part D program. Beneficiary counts, claim counts, and total drug costs are summarized from these PDE data. PDE records for over-the-counter drugs (indicated by drug coverage status code = "O"), which may be found in the PDE data due to their inclusion in an approved step-therapy protocols, are excluded from all summarizations. Drug brand names and generic names used in the summarization are obtained by linking the National Drug Codes (NDCs) from PDE records to a commercially available drug information database. A small proportion of PDE records with NDCs that do not match to the drug information database are excluded from all summarizations.

Prescriber demographics are also incorporated in the Part D Prescribers Dataset and include name, credentials, gender, complete address, and entity type from the National Plan & Provider Enumeration System (NPPES). The health care provider's demographic information is collected at the time of enrollment and updated periodically. The demographic information provided in the Part D Prescribers Dataset is based upon information extracted from NPPES as of the end of the subsequent calendar year (e.g., The 2017 Part D Prescribers Dataset includes NPPES information as of the end of calendar year 2018). For additional information on NPPES, please visit https://nppes.cms.hhs.gov/.

3. Population

The Part D Prescribers Dataset is based on beneficiaries enrolled in the Medicare Part D prescription drug program who comprise approximately 77.1% percent of all Medicare beneficiaries. Approximately 49.3% of Part D beneficiaries are enrolled in stand-alone Prescription Drug Plans (PDP) with the remaining 50.7% enrolled in Medicare Advantage Prescription Drug (MAPD) plans.

The Part D Prescribers Dataset is restricted to prescribers who had a valid NPI and who were included on Medicare Part D PDEs submitted by the Part D plan sponsors during the calendar year. The dataset contains information predominantly from individual providers, but also includes a small proportion of organizational providers, such as nursing homes, group practices, non-physician practitioners, residential treatment facilities, ambulatory surgery centers, and other providers.

4. Data Contents

4.1. Detailed Data File

Part D Prescribers by Provider and Drug

The spending and utilization data in the Part D Prescribers Dataset is aggregated to the following:

a) the NPI of the prescriber, and
b) the drug name (brand name in the case of trademarked drugs) and generic name.

Each record in the dataset represents a distinct combination of NPI, drug (brand) name, and generic name. There can be multiple records for a given NPI based on the number of distinct drugs that were filled. For each prescriber and drug, the dataset includes the total number of prescriptions that were dispensed (including original prescriptions and any refills), total 30-day standardized fill counts, total day's supply for these prescriptions, and the total drug cost. To protect the privacy of Medicare beneficiaries, any aggregated records which are derived from 10 or fewer claims are excluded from the Part D Prescribers Dataset. Please see the section on Limitations for additional information about data redactions and suppression in the Part D Prescribers Dataset.

The detail data file (Part D Prescribers Dataset by Provider and Drug) does not include providers with fewer than 11 total claims. Drugs for which an individual provider had fewer than 11 claims are also excluded from the detail data file.

4.2. Summary Tables

Two summary type tables have been created to supplement the information reported in the Part D Prescribers Dataset by Provider and Drug described above: 1) aggregated information at the prescriber-level (i.e. one summary record per NPI) that includes enhanced prescriber demographic information beyond what is provided in the Part D Prescribers Dataset by Provider and Drug; and 2) aggregated drug information at the Geography-level, brand name and generic name level. The aggregated summary tables are not restricted to the redacted data reported in the Part D Prescribers Dataset by Provider and Drug, but are aggregated based on all Medicare Part D PDE data.

Part D Prescribers by Provider

The Part D Prescribers Dataset by Provider (Part D Prescriber Summary Table) contains overall drug utilization (claims, 30-day standardized fill counts and day's supply), drug costs, and beneficiary counts organized by NPI. Drug utilization, drug costs, and beneficiary counts are also included for each of the following sub group classifications:

- Beneficiaries age 65 and older;
- Brand drugs, generic drugs, and other drugs;
- Medicare Advantage Prescription Drug (MAPD) and stand-alone Prescription Drug Plans (PDP);
- Low-income subsidy (LIS) and no low-income subsidy (nonLIS); and
- Opioids, long-acting opioids, antibiotics, and antipsychotics in the elderly.

In addition, beneficiary demographic and health characteristics are provided which include age, sex, race, Medicare and Medicaid entitlement and risk scores.

Part D Prescribers by Geography and Drug

The Part D Prescribers Dataset by Geography and Drug "Part D Geography/ Summary Tables" contain information on number of beneficiaries, number of prescribers, total drug claims, 30-day standardized fill counts and total drug costs for all beneficiaries and for beneficiaries age 65 and older. Aggregate cost share amounts that beneficiaries are responsible for paying also are included for both beneficiaries that receive a low-income subsidy and beneficiaries that do not receive a low-income subsidy. In addition, flag indicators are included to identify drugs as opioid, long-acting opioids, antibiotic and/or antipsychotic. The data are organized by geography level (national or state-level) and then by drug name and generic name.

5. Data Limitations

Although the Part D Prescribers Dataset has a wealth of payment and utilization information about Medicare prescription drug events (PDEs), the dataset also has a number of limitations that are worth noting. First, the information presented in this file does not indicate the quality of care provided by individual clinicians. Second, given that the data contain information only from Medicare beneficiaries with Part D coverage, but clinicians typically treat many other patients who do not have that form of coverage, the data in the Part D Prescribers Dataset may not be representative of a prescriber's entire prescribing pattern, nor be fully inclusive of all prescriptions written by the provider. Additionally, the data in this file are limited to medications covered by the Part D program and drugs statutorily excluded by the Part D program, which may be covered by individual Part D prescription drug plans through supplemental coverage. Since not all Part D plans have supplemental coverage for excluded products, utilization, and cost statistics presented in the data likely underestimates the true use of these products in this population.

The total drug costs included in these data reflect the prescription drug costs incurred by Medicare Part D beneficiaries, including costs that are paid by Medicare, by beneficiaries, and by third-party payers. The Part D prescription drug program is administered by private Part D plan insurers. Medicare pays Part D plans a monthly, risk-adjusted capitation payment for each enrollee. Beneficiaries also pay a monthly premium. In addition, Medicare pays Part D plans additional subsidies to cover reduced cost-sharing for low-income beneficiaries and a portion of the costs for beneficiaries whose drug costs are very high. Following each benefit year, CMS shares risk with plans by reconciling the capitation and various subsidy payments to actual drug cost expenditures determined from PDE records and any manufacturer rebates

or other direct and indirect remunerations received by the plan. Therefore, because the drug expenditures derived from the PDE data comprise only a piece of the payment process, it is not possible to directly attribute total drug costs at the prescriber or drug level to payments from the Medicare Trust Fund. Furthermore, these total drug costs do not reflect any manufacture rebates.

Also, there are known issues in the attribution of PDEs to a specific NPI. Some prescribers' claims may be listed under multiple NPIs, such as an organizational and individual NPI. In this case, users cannot determine a prescriber's actual total because it is not possible to identify the individual's portion when the claim is submitted under their organization. In addition, some of an individual's prescriptions might be erroneously attributed to a different prescriber due to errors that can occur in the transcription of prescriber information at the point-of-sale.

If users attempt to link data from these files to other public datasets, please be aware of the particular Medicare populations included and timeframes used in each file that will be merged, as well as the identifiers used to merge data. For example, efforts to link the Part D Prescriber data to the Medicare Physician and Other Practitioners Datasets would need to account for the fact that some beneficiaries who have fee-for-service (FFS) Part B coverage (and are thus included in the Physician and Other Practitioners Datasets) do not have Part D drug coverage (and thus not represented in the Part D Prescribers Datasets). At the same time, some beneficiaries that have Part D coverage (and are thus included in the Part D Prescribers Dataset) do not have FFS Part B coverage (and thus not included in the Physician and Other Practitioners Dataset). Another example would be linking to data constructed from different or non-aligning time periods, such as publicly available data on physician referral patterns, which is based on an 18-month period. Users attempting to merge data from the Part D Prescribers Dataset to publicly available Open Payments data on financial relationships should be aware that NPIs are not available in the Open Payments data and thus merges must be conducted using text-string identification fields such as name and address.

Data Redaction and Suppression

As previously stated, the Part D Prescribers detail data file does not include drugs with fewer than 11 Part D drug claims, so users should be aware that summing data in detail file will underestimate the true Part D totals. In addition, in the detail file as well as the summary tables, beneficiary counts, claim counts, 30-day fill counts, drug costs, and day's supply are suppressed if the value is between 1 and 10 and also may be removed for counter-suppression purposes. Since total claim counts are available on the files and some subgroups (e.g., brand, generic, and other) sum to the total claim count, if one of the sub-group categories is suppressed because it has a claim count between 1 and 10 (primary suppression), then the next lowest claim count sub-group category must be suppressed to prevent disclosure of this primary suppressed value. Since only one sub-group category is suppressed, you can mathematically determine it using the values from the other claim count categories and the total claim count information. To help users understand the reasons for suppression, suppression flag variables are included.

Suppressed values represent values 1 to 10 and are indicated by a "blank" in the data files. When analyzing the data, users should note that excluding the suppressed values will result in estimates that

are different from the true values. If users choose to retain the suppressed values in their analysis, please note that most statistical software packages will treat the "blanks" as "zeroes", resulting in underestimates of the true values. Alternatively, users may assign an imputed value of their choosing, e.g. five (5), for the suppressed value.

6. Additional Information

Other Data Sources: CMS also releases the "Medicare Fee-For-Service Public Provider Enrollment Data" that include provider name and address information from the Provider Enrollment and Chain Ownership System (PECOS). These data are updated on a quarterly basis and are available at data.cms.gov.

Opioid Classification: The opioid and long-acting opioid information presented in the "Part D Prescribers Dataset by Provider" and "Part D Prescribers Dataset by Geography" tables is based upon opioid drugs included in the Medicare Part D Overutilization Monitoring System (OMS), which can change from year to year. Additional information on Medicare Part D OMS are available at https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/RxUtilization.html.

HCCs (hierarchical condition categories): CMS developed a risk-adjustment model that uses HCCs (hierarchical condition categories) to assign risk scores. Those scores estimate how beneficiaries' FFS spending will compare to the overall average for the entire Medicare population. The average risk score varies from year to year; beneficiaries with scores greater than that are expected to have above-average spending, and vice versa. Risk scores are based on a beneficiary's age and sex; whether the beneficiary is eligible for Medicaid, first qualified for Medicare on the basis of disability, or lives in an institution (usually a nursing home); and the beneficiary's diagnoses from the previous year.

The HCC model was designed for risk adjustment on larger populations, such as the enrollees in a Medicare Advantage plan, and generates more accurate results when used to compare groups of beneficiaries rather than individuals. For more information on the HCC risk score, see: https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html.

7. Updates

May 2023 Updates

For data years prior to 2021, long-acting opioid claims were identified using controlled substances including opioids with oral morphine milligram equivalent (MME) data available for download on Centers for Disease Control and Prevention (CDC) Opioid Resource Center's web page at: https://www.cdc.gov/opioids/data-resources/index.html. Beginning with data year 2021, long-acting opioid claims are identified using data from the Medicare Part D OMS.

August 2021 Updates

Please note that data for years 2013 – 2019 was updated in August 2021 to align with the methodology described in this document. Users may observe slight differences from previously released data; these are primarily due to changes in how suppression was applied.