

CMS 2010 Prescription Drug Profiles Public Use File (PUF)

General Documentation

1. Overview of the PUF

This release contains the 2010 Prescription Drug Profiles Public Use File (PUF) drawn from 2010 Medicare prescription drug claims.¹ The PUF includes beneficiary demographics (e.g., gender, age), plan characteristics (e.g., plan type, drug benefit type, gap coverage), prescription drug characteristics (e.g., active ingredient, drug class), prescriber characteristics, and payment information (e.g., average patient pay amount).

The *CMS 2010 Prescription Drug Profiles PUF* is an aggregated file in which each record is a *profile* or *cell* defined by the characteristics of Medicare beneficiaries, drugs, plans, and prescribers.

2. Source Data for the PUF

The *2010 Prescription Drug Profiles PUF* represents 100% of prescription drug claims (subject to suppression) for Medicare beneficiaries covered by Part D for reference year 2010. The *CMS 2010 Prescription Drug Profiles PUF* provides various measures of utilization as averages for different groups of Medicare beneficiaries, or *profiles*. Prescription drug events (or Part D Drug Event File) are merged with the following files to construct the *2010 Prescription Drug Profiles PUF*:

- 2010 Beneficiary Summary File;
- 2010 Part D Plan Characteristics File;
- 2010 Part D Prescriber Characteristics File;
- 2010 Part D Formulary File;
- RxNorm, available from the U.S National Library of Medicine, National Institutes of Health;²
- National Formulary, available from the Department of Veterans Affairs;³ and
- Results of the CMS-RxHCC risk-adjustment model.⁴

¹ 2010 is the year of the date in which the prescription was filled.

² RxNorm is available at <http://www.nlm.nih.gov/research/umls/rxnorm/>. The version used for the CMS 2010 Enhanced PDE PUF is the October 2010 release.

³ National Formulary is available at <http://www.pbm.va.gov/NationalFormulary.aspx>.

⁴ See https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk_adjustment.html.

Then, the information is aggregated into profiles defined by 13 variables: (1) Gender; (2) Age category; (3) Drug RxNorm Concept Unique Identifier RxCUI; (4) Drug major class; (5) Drug class; (6) Drug type; (7) Plan type; (8) Coverage type; (9) Benefit phase; (10) Drug benefit type; (11) Prescriber taxonomy; (12) Plan gap coverage indicator; and (13) Tier identifier. Finally, number of events and average utilization measures for each combination of profile variables are calculated.

3. Content of the PUF

The most important aspects of the *CMS 2010 Prescription Drug Profiles PUF* are as follows:

- i. It contains 1,110,089,154 prescription events that belong to 27,525,541 unique Medicare beneficiaries in 2010.
- ii. It contains analytic variables on cost and utilization of the Medicare prescription drug benefit in 2010.
- iii. Every profile in the PUF reflects data for at least 11 beneficiaries in the population. To accomplish this, some profiles are coarsened (see below) into broader ones by suppressing selected variables.
- iv. The PUF does not provide the unique beneficiary count variable associated with each profile but provides a categorized version of that variable.
- v. The fact that the PUF does not provide beneficiary-level information reduces concern about the privacy of the beneficiaries. Nevertheless, the PUF is tested rigorously to ensure that it can be released to the public without compromising beneficiaries' privacy.

4. Variables of the PUF

The *CMS 2010 Prescription Drug Profiles PUF* contains 19 variables of two types: (1) variables that define the profiles (e.g., age, gender, drug name) and (2) variables that summarize the cost and utilization measures of events. Thirteen of the variables define profiles or cells:

1. Gender (BENE_SEX_IDENT_CD): The beneficiary's gender, (1) male or (2) female.
2. Age (BENE_AGE_CAT_CD): The beneficiary's age, reported in six categories: (1) under 65, (2) 65-69, (3) 70-74, (4) 75-79, (5) 80-84, (6) 85 and older.
3. Drug RxNorm Concept Unique Identifier (RXNORM_RXCUI): This is a set of 1,229 possible numeric identifiers including missing/unknown. Each RXNORM_RXCUI identifies a unique drug name which is separately provided.

4. Drug major class (DRUG_MAJOR_CLASS): This is a set of 30 possible alpha-numeric codes including missing/unknown. It indicates the major class of the drug, such as cardiovascular medications and central nervous system medications.
5. Drug class (DRUG_CLASS): This is a set of 263 possible alpha-numeric (two-letter plus three-digit) codes. It indicates the class of the drug, such as antidepressants and analgesics.
6. Drug type code (PDE_DRUG_TYPE_CD): This is a set of four possible codes including missing/unknown. It indicates the type of a drug: (0) unknown, (1) brand name drug, (2) generic drug, and (*) suppressed.
7. Plan type (PLAN_TYPE): The beneficiary's plan type, reported in four possible codes: (1) Medicare Advantage Prescription Drug (MAPD), (2) Prescription Drug Plan (PDP), (3) other, and (*) suppressed. The plan type is assigned to each event using contract specific identifiers.
8. Coverage type (COVERAGE_TYPE): The beneficiary's coverage type, reported in five categories: (0) unknown, (1) dual eligible, (2) receiving low income subsidy, (3) no subsidy, and (*) suppressed. The coverage is determined by matching the month of the prescription drug event with the monthly cost share group information for the beneficiary.
9. Drug benefit phase (BENEFIT_PHASE): This is the benefit phase for the event, reported in seven categories: (0) no information, (1) catastrophic, (2) deductible, (3) initial coverage limit (ICL), (4) non-covered drug, (5) pre-initial coverage limit (Pre-ICL), and (*) suppressed.
10. Drug benefit type (DRUG_BENEFIT_TYPE): Indicates the type of Part D benefit structure used by the plan benefit package, reported in six categories: (0) no plan benefit information, (1) actuarially equivalent, (2) basic alternative, (3) defined standard, (4) enhanced alternative, and (*) suppressed.
11. Prescriber taxonomy (PRESCRIBER_TYPE): The prescriber's taxonomy, or type, is reported in six categories: (1) family medicine, (2) internal medicine, (3) psychiatry & neurology, (4) specialist, (5) other, and (*) suppressed. The first three are sub-classes of "allopathic & osteopathic physicians" and the fourth is a sub-class of "other service

providers.” These are based on the National Uniform Claims Committee (NUCC) taxonomy codes.⁵

12. Gap coverage indicator (GAP_COVERAGE): Indicates whether the plan includes in its supplemental benefits some coverage in the coverage gap, reported in four categories: (0) unknown, (1) plan does not include gap coverage, (2) plan offers gap coverage, and (*) suppressed.
13. Formulary tier code (TIER_ID): This is the minimum cost sharing tier in which the drug was placed in the sponsor’s formulary, reported in eight categories including unknown and suppressed.

Age, gender, and coverage type are calculated from the Beneficiary Summary File. Plan type, drug benefit type, and gap coverage are calculated from the Part D Plan Characteristics File. Formulary tier is calculated from the Part D Formulary File. Benefit phase is calculated from the Part D Drug Event File. Prescriber taxonomy/type is calculated from the Part D Prescriber Characteristics File. These files are available from the Chronic Condition Data Warehouse.⁶

There are two additional external data sources that are utilized in the creation of the *CMS 2010 Prescription Drug Profiles PUF*: RxNorm (available from the U.S National Library of Medicine, National Institutes of Health) and National Formulary⁷ (available from the Department of Veterans Affairs). Both sources include information on drugs at the National Drug Code (NDC) level.⁸ NDCs are unique identifiers for drugs mandated by the Food and Drug Administration (FDA). RxNorm includes detailed information on the drug characteristics (e.g., name, ingredients, strength, dose form) at the NDC level. The National Formulary database includes information on the class of the drug at the NDC level. Hence, the information from these two data sources is included in the *CMS 2010 Prescription Drug Profiles PUF* using the NDC of each event. The variables RXNORM_RXCUI and PDE_DRUG_TYPE_CD are gathered from RxNorm and DRUG_MAJOR_CLASS and DRUG_CLASS are gathered from the Veterans Affairs National Drug File (VA-NDF).

An NDC code in the Part D Drug Event File may or may not exist in RxNorm or VA-NDF. When an NDC code is not found in RxNorm, the variables that are gathered from RxNorm are set to missing. Similarly, when an NDC is not found in VA-NDF, the class of the drug is set to missing.

⁵ For a current list of NUCC provider taxonomy codes and descriptions see the “Code Sets” link at <http://www.nucc.org>.

⁶ See <http://www.ccwdata.org/data-dictionaries/index.htm>.

⁷ National Formulary is available at <http://www.pbm.va.gov/NationalFormulary.aspx>.

⁸ NDC is maintained and published by the US. Food and Drug Administration. Available at: <http://www.fda.gov/Drugs/InformationOnDrugs/ucm142438.htm>.

The next four variables provide additional information on cost and utilization in the form of averages. These are arithmetic averages in which the numerator is the sum of values of all events in the profile and the denominator is the number of events in the profile. These variables are:

14. Mean prescription drug hierarchical conditional category (RxHCC) score (MEAN_RXHCC_SCORE): This is the average of the RxHCC score associated with beneficiaries with a prescription drug event in the profile. The RxHCC score is created by the CMS-RxHCC risk-adjustment model. The model assigns a risk score to reflect the health status of each beneficiary according to demographic variables and the beneficiary's diagnosis history. A higher risk score correlates to higher estimated costs for a beneficiary. Payments to Medicare Part D Plans are adjusted by CMS according to the risk scores of the beneficiaries.
15. Average days supply (AVE_DAYS_SUPPLY): This is the arithmetic average of the number of days prescribed for the events in the profile. Values greater than 60 are rounded to the nearest day.
16. Average total drug cost (AVE_TOT_DRUG_COST): This is the arithmetic average of the gross cost of the events in the profile. Values greater than 500 are rounded to the nearest dollar.
17. Average payment by patient (AVE_PTNT_PAY_AMT): This is the arithmetic average of the payment amounts by the patients for the events in the profile. Values greater than 100 are rounded to the nearest dollar.

The last two variables in the file are (i) the number of events associated with each profile, or the weight, and (ii) the category for the number of beneficiaries associates with each profile:

18. Number of prescription drug events (PDE_CNT): This is the count of events in the profile. This variable is the denominator for variables MEAN_RXHCC_SCORE, AVE_PTNT_PAY_AMT, AVE_DAYS_SUPPLY, and AVE_TOT_DRUG_COST.
19. Number of unique beneficiaries (categorized) associated with the profile (BENE_CNT_CAT): This is the number of beneficiaries for the events summarized in the profile. Note that this variable categorized into 6 values: (1) 11 – 15, (2) 16 – 20, (3) 21 – 50, (4) 51 – 100, (5) 101 – 500, and (6) more than 500 beneficiaries.

The detailed definitions of these variables are provided in the Data Dictionary & Codebook that accompanies this document.

5. Methodology and Key Assumptions

A key property of the *CMS 2010 Prescription Drug Profiles PUF* is that it provides information for 100% of the Part D Drug events while protecting the privacy and confidentiality of Medicare beneficiaries. This was achieved by de-identifying the file using suppression of variable values for some of the profile variables. For profiles containing fewer than 11 beneficiaries, a suppression algorithm was applied until every profile contains at least 11 beneficiaries. The profile variables were suppressed in the following order: (1) Formulary tier code; (2) Gap coverage indicator; (3) Prescriber type; (4) Drug benefit type; (5) Drug benefit phase; (6) Coverage type; (7) Plan type; and (8) Drug type code.

First, 216,654 events out of 1,110,305,808 (or 0.02 percent) were dropped from the file because these events remained in profiles containing less than 11 beneficiaries even after applying the suppression algorithm. In the remaining data with 1,110,089,154 events, profiles containing a total of 30,628,824 drug events had at least one of the variable values suppressed affecting less than three percent of total drug events. The prescription events in the PUF belong to 27,525,541 beneficiaries (approximately 93 percent of Medicare Part D enrollees in 2010).

Table 1 provides a summary of suppression in the *CMS 2010 Prescription Drug Profiles PUF*. The values for gender, age category, drug name, RxNorm RxCUI, drug major class, and drug class are never suppressed in the *CMS 2010 Prescription Drug Profiles PUF*.

Table 1. Suppression in the CMS 2010 Prescription Drug Profiles PUF

Suppressed Variables	Number of events	Percentage of Events
None	1,079,460,330	97.24%
Tier ID	5,971,573	0.54%
Tier ID, gap coverage	3,093,120	0.28%
Tier ID, gap coverage, prescriber type	14,578,707	1.31%
Tier ID, gap coverage, prescriber type, benefit type	3,725,219	0.34%
Tier ID, gap coverage, prescriber type, benefit type, benefit phase	2,330,474	0.21%
Tier ID, gap coverage, prescriber type, benefit type, benefit phase, coverage type	523,112	0.05%
Tier ID, gap coverage, prescriber type, benefit type, benefit phase, coverage type, plan type	349,216	0.03%
Tier ID, gap coverage, prescriber type, benefit type, benefit phase, coverage type, plan type, drug type	57,403	0.01%
Total	1,110,089,154	100.00%

6. Analytic Utility of the PUF

The *CMS 2010 Prescription Drug Profiles PUF* provides information on the number of Part D events, days supply, total drug costs, and patient payments by the profile variables. Table 2 summarizes the events, days supply, costs, and payments by select variables of the *CMS 2010 Prescription Drug Profiles PUF*.

Table 2. Summary of the CMS 2010 Prescription Drug Profiles PUF

Variable	Total Events	Total Days Supply ⁽¹⁾	Total Drug Cost (\$) ⁽¹⁾	Total Patient Payments (\$) ⁽¹⁾
Gender				
Male	407,083,843	14,569,142,938	30,840,322,824	5,008,445,228
Female	703,005,311	24,172,486,960	46,677,022,027	7,522,015,466
Age				
Under 65	265,185,291	7,960,319,256	24,613,169,044	1,285,635,448
65 - 69	184,585,324	6,909,708,874	12,440,849,950	2,587,928,996
70 - 74	188,676,103	7,126,309,435	12,268,644,937	2,617,498,732
75 - 79	162,999,181	6,064,522,346	10,216,174,105	2,178,045,901
80 - 84	142,604,898	5,145,431,863	8,672,131,054	1,889,596,347
85 & Older	166,038,357	5,535,338,123	9,306,375,760	1,971,755,270
Drug type				
Suppressed	57,403	1,695,931	8,008,517	773,010
Unknown/Missing	18,057,071	433,892,232	717,234,497	175,943,366
Brand name	291,380,901	10,131,055,062	57,906,367,868	8,570,067,435
Generic	800,593,779	28,174,986,673	18,885,733,969	3,783,676,883
Plan type				
Suppressed	406,619	12,330,429	126,375,825	9,360,197
MAPD	342,061,439	12,859,413,919	20,807,630,735	4,123,799,843
PDP	762,130,548	25,649,977,065	56,180,442,062	8,290,804,325
Other	5,490,548	219,908,484	402,896,229	106,496,329
Coverage type				
Suppressed	929,731	27,981,301	317,216,090	21,237,222
Unknown/Missing	19,130	654,449	973,253	113,045
Dual Eligible	496,930,662	14,792,163,989	38,174,416,840	671,060,437
Low Income Subsidy	64,628,904	2,183,465,861	4,818,588,828	269,310,210
No Subsidy	547,580,727	21,737,364,298	34,206,149,840	11,568,739,779
Benefit phase				
Suppressed	3,260,205	97,692,599	882,842,248	80,547,167
No information	75,320,581	3,374,911,295	6,157,820,254	1,290,679,291
Catastrophic	87,375,217	2,611,355,332	14,392,101,862	140,187,858
Deductible	86,358,657	2,746,956,600	3,957,310,522	741,747,431
Initial Coverage Limit	181,794,463	5,744,757,743	15,139,002,783	3,334,667,402

Variable	Total Events	Total Days Supply ⁽¹⁾	Total Drug Cost (\$) ⁽¹⁾	Total Patient Payments (\$) ⁽¹⁾
Non-covered Drug	7,297,816	220,099,968	108,679,621	28,760,602
Pre-initial Coverage Limit	668,682,215	23,945,856,362	36,879,587,561	6,913,870,943
Drug benefit type				
Suppressed	6,985,424	208,312,278	1,785,305,550	164,587,093
No Plan Benefit Information	74,532,350	3,374,517,861	6,094,344,148	1,301,064,183
Defined Standard	125,459,412	3,670,163,341	8,771,526,168	264,796,852
Actuarially Equivalent	331,695,876	10,186,180,814	24,948,556,604	1,448,589,058
Basic Alternative	196,061,975	7,166,952,514	13,791,491,281	3,369,586,299
Enhanced Alternative	375,354,117	14,135,503,089	22,126,121,101	5,981,837,209
Prescriber type				
Suppressed	21,564,131	656,652,624	4,263,051,011	476,720,418
Family Medicine	327,145,574	11,640,145,662	17,870,380,879	3,168,874,034
Internal Medicine	434,639,707	16,244,681,922	30,120,297,668	5,423,664,061
Psychiatry & Neurology	41,886,388	1,324,756,518	6,357,499,957	468,504,038
Specialist	34,710,116	1,207,063,582	2,559,628,198	444,610,084
Other	250,143,238	7,668,329,590	16,346,487,138	2,548,088,059
Gap coverage				
Suppressed	24,657,251	753,703,252	4,847,738,185	571,099,502
Unknown/Missing	718,209,097	24,106,574,094	51,929,795,980	6,206,800,679
No Gap Coverage	178,465,198	6,644,246,903	10,023,688,548	2,789,512,459
Plan offers Gap Coverage	188,757,608	7,237,105,648	10,716,122,138	2,963,048,054
Total	1,110,089,154	38,741,629,898	77,517,344,851	12,530,460,694

(1) Calculated by multiplying number of Part D events by the average amount in the PUF and aggregating.