

CMS 2008 Prescription Drug Profiles Public Use File (PUF)

General Documentation

1. Overview of the PUF

This release contains the 2008 Prescription Drug Profiles Public Use File (PUF) drawn from 2008 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 2008 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 *2008 Prescription Drug Profiles PUF* represents 100% of prescription drug claims (subject to suppression) for Medicare beneficiaries covered by Part D for reference year 2008. The *CMS 2008 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 *2008 Prescription Drug Profiles PUF*:

- 2008 Beneficiary Summary File;
- 2008 Part D Plan Characteristics File;
- 2008 Part D Prescriber Characteristics 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

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;

¹ 2008 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 2008 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.

(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 2008 Prescription Drug Profiles PUF* are as follows:

- i. It contains 1,036,935,676 prescription events that belong to 25,288,330 unique Medicare beneficiaries in 2008.
- ii. It contains analytic variables on cost and utilization of the Medicare prescription drug benefit in 2008.
- 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 2008 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 1,254 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

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

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 nine categories including 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. Benefit phase and formulary tier are 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 2008 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 2008 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.

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:

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>.

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 six 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 2008 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, 208,531 events out of 1,037,144,207 (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 1,036,935,676 events, profiles containing a total of 29,273,966 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 25,288,330 beneficiaries (approximately 92 percent of Medicare Part D enrollees in 2008).

Table 1 provides a summary of suppression in the *CMS 2008 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 2008 Prescription Drug Profiles PUF*.

Table 1. Suppression in the CMS 2008 Prescription Drug Profiles PUF

Suppressed Variables	Number of events	Percentage of Events
None	1,007,661,710	97.18%
Tier ID	4,981,925	0.48%
Tier ID, gap coverage	2,678,633	0.26%
Tier ID, gap coverage, prescriber type	14,519,487	1.40%
Tier ID, gap coverage, prescriber type, benefit type	3,978,980	0.38%
Tier ID, gap coverage, prescriber type, benefit type, benefit phase	2,211,165	0.21%
Tier ID, gap coverage, prescriber type, benefit type, benefit phase, coverage type	515,349	0.05%
Tier ID, gap coverage, prescriber type, benefit type, benefit phase, coverage type, plan type	330,787	0.03%
Tier ID, gap coverage, prescriber type, benefit type, benefit phase, coverage type, plan type, drug type	57,640	0.01%
Total	1,036,935,676	100.00%

6. Analytic Utility of the PUF

The *CMS 2008 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 2008 Prescription Drug Profiles PUF*.

Table 2. Summary of the CMS 2008 Prescription Drug Profiles PUF

Variable	Total Events	Total Days Supply ⁽¹⁾	Total Drug Cost (\$) ⁽¹⁾	Total Patient Payments (\$) ⁽¹⁾
Gender				
Male	368,635,108	12,481,047,458	26,652,768,041	4,343,978,432
Female	668,300,568	21,870,991,682	41,814,587,357	6,970,967,684
Age				
Under 65	235,978,902	6,847,511,546	21,428,313,287	1,129,450,721
65 - 69	170,922,588	6,039,873,889	10,854,808,895	2,252,675,550
70 - 74	174,962,475	6,206,166,737	10,635,518,832	2,292,738,143
75 - 79	158,240,409	5,541,086,961	9,309,726,193	2,053,937,233
80 - 84	139,550,521	4,741,762,621	7,933,703,564	1,786,489,714
85 & Older	157,280,781	4,975,637,385	8,305,284,626	1,799,654,754
Drug type				
Suppressed	57,640	1,630,888	7,008,413	747,246
Unknown	21,588,694	554,658,435	1,096,876,483	248,933,563
Brand name	332,370,502	11,202,155,673	51,278,238,757	8,278,303,301
Generic	682,918,840	22,593,594,145	16,085,231,745	2,786,962,006
Plan type				
Suppressed	388,427	11,320,361	96,240,613	7,568,006
MAPD	281,429,669	10,118,509,689	16,262,498,141	3,354,670,313
PDP	751,030,679	24,055,750,061	51,807,679,888	7,866,835,568
Other	4,086,901	166,459,028	300,936,755	85,872,229
Coverage type				
Unknown	14,327	434,589	666,098	70,499
Suppressed	903,776	26,187,765	228,904,020	17,536,286
Dual Eligible	459,364,576	13,215,829,494	33,599,832,971	563,184,206
Low Income Subsidy	59,610,372	1,918,168,919	4,045,851,420	235,299,274
No Subsidy	517,042,625	19,191,418,373	30,592,100,889	10,498,855,850
Benefit phase				
No information	63,292,090	2,839,425,904	4,960,957,291	1,129,158,577
Suppressed	3,114,941	88,740,149	637,636,658	65,183,168
Catastrophic	87,707,528	2,547,575,010	12,220,063,494	119,811,348
Deductible	60,673,465	1,828,785,123	2,851,610,522	588,072,222
Initial Coverage Limit	188,564,388	5,779,185,553	14,445,028,328	3,550,224,512

Variable	Total Events	Total Days Supply ⁽¹⁾	Total Drug Cost (\$) ⁽¹⁾	Total Patient Payments (\$) ⁽¹⁾
Non-covered Drug	8,556,625	254,670,198	207,427,686	42,092,397
Pre-initial Coverage Limit	625,026,639	21,013,657,203	33,144,631,419	5,820,403,892
Drug benefit type				
Suppressed	7,093,921	201,368,708	1,345,021,791	141,323,363
No Plan Benefit Information	62,768,244	2,842,281,263	4,928,543,821	1,138,551,371
Defined Standard	150,260,008	4,433,535,034	10,029,366,758	874,342,528
Actuarially Equivalent	202,870,889	6,059,956,672	14,763,789,963	986,421,708
Basic Alternative	275,737,542	8,936,341,645	18,199,583,277	2,903,901,625
Enhanced Alternative	338,205,072	11,878,555,818	19,201,049,788	5,270,405,521
Prescriber type				
Suppressed	21,613,408	629,105,065	3,393,746,732	423,055,456
Family Medicine	295,253,119	9,854,883,171	15,590,728,966	2,739,946,879
Internal Medicine	393,486,675	13,905,389,803	25,148,313,943	4,667,509,292
Psychiatry & Neurology	35,515,898	1,098,195,568	5,542,830,154	403,942,579
Specialist	35,507,851	1,184,619,316	2,501,384,245	444,806,760
Other	255,558,725	7,679,846,217	16,290,351,359	2,635,685,150
Gap coverage				
Unknown/Missing	191,048	6,055,990	8,954,030	526,994
Suppressed	24,292,041	710,118,486	3,813,678,244	490,995,270
No Gap Coverage	822,619,215	26,873,899,204	54,246,188,044	7,887,403,045
Plan offers Gap Coverage	189,833,372	6,761,965,460	10,398,535,079	2,936,020,807
Total	1,036,935,676	34,352,039,140	68,467,355,398	11,314,946,116

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