Chronic Condition Data Warehouse

Your source for national CMS Medicare and Medicaid research data

CCW Technical Guidance:

Summarizing and Describing Prescription Drug Utilization

January 2018

Version 1.3

This page intentionally left blank.

Revision History

Revision Date	Version Number	Description
01/2018	1.3	MBSF A/B references updated to include C/D; updated data source information to reflect CME instead of EDB.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF CODE EXAMPLES

Chapter 1: Introduction

The Centers for Medicare & Medicaid Services (CMS) launched the Chronic Condition Data Warehouse (CCW), a research database, in response to the Medicare Modernization Act of 2003 (MMA). Section 723 of the MMA outlined a plan to improve the quality of care and reduce the cost of care for chronically ill Medicare beneficiaries. An essential component of this plan was to establish a data warehouse that contains Medicare claims data and assessments, linked by beneficiary, across the continuum of care. Part of the MMA (section 101) established a prescription drug benefit, which is a voluntary benefit offered though the Medicare Part D program. Medicare beneficiaries who are deemed eligible for full Medicaid benefits are automatically enrolled in Part D. Beginning in 2006, beneficiaries were able to purchase Part D coverage through private plans. Medicare Part D data became available to researchers in 2008.

The CCW Medicare administrative claims files are provided to academic researchers and certain government agencies, which have been approved under a Data Use Agreement (DUA) to obtain Medicare administrative data for research purposes. The CCW Medicare data contain identifiable information, and are subject to the Privacy Act and other Federal government rules and regulations (see ResDAC web site for information on requesting Medicare data http://www.resdac.org/). To obtain a DUA to use the Part D data, researchers are required to select variables based on minimum data necessary and justify the need for each variable.

Part D enrollment and utilization counts are published on the CCW website (see Table F1, for example, at https://www.ccwdata.org/web/guest/medicare-tables-reports). Depending on the assumptions researchers make when requesting and compiling variables – different useable sample sizes may be obtained.

The CCW data files were designed to support a variety of research objectives. The Part D User Guide provides background information regarding the Part D benefit and the data variables available for researchers (see the CCW website at: https://www.ccwdata.org/web/guest/user-documentation). This Technical Guidance paper should be used in conjunction with the Part D User Guide.

This paper is designed to help researchers more easily, efficiently, and accurately work with the Part D data files to summarize and describe prescription drug utilization.

Objectives

This Technical Guidance paper is intended to provide guidance to researchers in order to:

1) Specify a study cohort based on Medicare Part D coverage criteria,

2) Link the cohort/enrollment file to the Part D utilization file (known as the Prescription Drug Event File), and the Part D Drug Characteristics File (which describes the drug that was dispensed), and

3) Perform data tabulations to accurately calculate particular study metrics.

CCW is pleased to offer some hints that we have found useful in working with Part D data to obtain the types of information commonly desired from these data files.

Chapter 2: Contents of CCW Part D Data Files

Data for 100% of Medicare-enrolled beneficiaries are available within the CCW. A standard data file available to researchers is the CMS 5% random sample. Criteria for selection of the random 5% are documented in the CCW User Manual, in Chapter 1, at https://www.ccwdata.org/web/guest/user-documentation. Alternatively, CCW data are available upon request for specific researcher-defined cohorts. Medicare Part D enrollment information is available beginning in 2006 as part of the CCW Master Beneficiary Summary File (MBSF).

The CCW contains only Medicare A and B fee-for-service claims, and no managed care encounter information. For Medicare Part D, all prescription drug events (PDEs) are present – regardless of whether the beneficiary is enrolled in managed care or a fee-for-service plan. PDEs are available for all filled prescriptions that are covered as part of the Part D benefit. A small number of non-covered drugs may be included in the PDE File (e.g., supplemental drugs reported by enhanced alternative plans, over-the-counter drugs). CCW obtains the PDE File from CMS late in the year following the year of service in a standard analytic file (SAF) file[1]. Only the final, reconciled transactions are included in the data file. Each prescription fill appears in the data file as a unique PDE.

The PDE includes services for beneficiaries enrolled in Medicare Advantage Prescription Drug Plans (MA-PD; managed care; prescription drug coverage that is integrated with the health care coverage provided under Medicare Part C) and stand-alone prescription drug plans (PDP; offer only prescription drug coverage) – also employer plans. Although service equivalent to a Part D benefit may be offered through Retirement Drug Subsidy (RDS) plans or other creditable plans (e.g., the VA), there are not Part D events data for these enrollees.

CCW processes the PDE and other Part D data to create several derived variables and some supplemental Part D Characteristics data files. Researchers may choose to purchase the "drug characteristics" to supplement the information in the PDE file they obtain, which describe the drug (e.g., brand name, generic name, dosage form and strength).

The data files contain key variables which can be used to join the CCW files. The linkage keys used may vary depending on which files researchers are attempting to join, however for the purposes of this paper the primary linkage will be at the beneficiary (person) – level. It is important to note that several of the Part D data fields are encrypted using DUA-specific criteria. For the purposes of this paper, these encrypted variables include the beneficiary ID (BENE_ID) and a unique key for each prescription drug fill, that is, for each PDE (PDE_ID).

Note: Prior to the 2013 data release, CCW was required to encrypt or use CCW-assigned ID values to protect some of the sensitive Part D data fields. These sensitive fields were those which could be used to identify the prescriber, the pharmacy, the plan, or the plan's

[1] Starting with 2012 data, CCW developed our own version of the SAF file to allow for expedited data delivery.

formulary. We are no longer required to encrypt these variables; however, the CCW Pharmacy Characteristics data files are only released to investigators who agree to forgo the unencrypted pharmacy identifiers on the PDE (and instead will receive the NCPDP ID as the linkage variable with the Pharmacy Characteristics file).

The CCW data extracts are provided to researchers in a user-friendly format. SAS® read-in statements are provided along with the data files requested by the researcher. Two different versions of read-in statements are routinely provided. The SASv6 file (e.g., beneficiary_summary_file_read_v6.sas) contains traditional short SAS® names, which may be appealing for researchers who have worked with historical CMS data. The SASv8 read-in statements (e.g., beneficiary_summary_file_read_v8.sas) take advantage of newer features in SAS®, which allow for longer and more descriptive variable names. The variable names used for examples in this document are the long names found in the SASv8 read-in file.

Chapter 3: Defining a Cohort

Throughout this paper, the assumption is that we begin with a Medicare 1% random sample of beneficiaries who, in 2011, had at least one month of Part D coverage. We are using the CCW Master Beneficiary Summary File Part A/B/C/D segment (also referred to as the Base Master Beneficiary Summary File); then we also use the PDE File for this population for 2011.

A. Denominator Selection

An ideal file to use to make your cohort selection is the CCW Master Beneficiary Summary File (MBSF). The MBSF – Part A/B/C/D segment contains the Medicare Part D enrollment/eligibility status for each Medicare-eligible beneficiary. This information is present regardless of the type of Medicare Part D plan the beneficiary might select (i.e., enrollment data are present for managed care participants as well as those enrolling in stand-alone prescription drug plans). Information is also available for Medicare beneficiaries who did not obtain Part D coverage. A description of the variables contained within each CCW data file can be found on the Data Dictionaries page of the CCW web site (https://www.ccwdata.org/web/guest/data-dictionaries). Most investigators will want to include information from the MBSF A/B/C/D segment related to Medicare Parts A/B/C coverage and demographic information. There is a single row of data for each beneficiary in the population, identified by a unique encrypted beneficiary identification number (BENE_ID).

The additional segments of MBSF are: 1) CCW Conditions, 2) Other Chronic or Potentially Disabling Conditions, and 3) Cost and Use. These segments of the MBSF are easily joined together using the BENE_ID.

Note: When the data files are delivered from CCW, they are already pre-sorted and indexed by BENE_ID.

The MBSF is created annually and contains demographic eligibility and enrollment data for all beneficiaries who are alive and eligible for Medicare for any part of the year. Researchers may use this annual person-level summary file to determine whether a beneficiary has a sufficient surveillance period (i.e., months of Medicare coverage) for inclusion in the study. Variables contained in this file include the number of months of Medicare Part A, B, C , and D coverage; whether the beneficiary died during the year; Part D plan type; whether the beneficiary received Part D subsidies; as well as other beneficiary demographic and geographic information. Refer to the ccwdata.org website for complete record layout (https://www.ccwdata.org/web/guest/data-dictionaries).

1. Length of Part D coverage

From this universe of potential beneficiaries to include in your study, you may wish to limit the population to those who are “at risk” for the events of interest. For example, not all Medicare beneficiaries will elect to purchase Part D coverage. You may wish to limit your cohort to those

who have Part D coverage, perhaps a certain minimum amount of Part D coverage, during the time frame of interest.

Decide whether other considerations are important for your cohort, such as the type of Part D coverage, Medicare Part A/B and fee-for-service coverage, and presence of chronic conditions.

2. Type of Part D Coverage

Do you need to understand the benefit structure for the Part D plan selected by the members of your cohort? If so, you may wish to subset the data to keep members of particular types of plans or organizations. There are some Part D plans that are not required to report benefit details to CMS (e.g., Program of All-inclusive Care for the Elderly [PACE] and employer direct plans). The employer direct plans can be identified in 2007 forward, by the first digit of the Part D contract ID (monthly PTD_CNTRCT_ID_<MM> on the MBSF-ABCD segment; the first digit is "E"). More detailed information regarding the type of benefit offered by each contract that a Part D plan has with CMS is available in the Plan Characteristics data file (see particularly variables such as ORGANIZATION_TYPE or PLAN_TYPE).

Note: A particular plan benefit package is identified using a combination of both the contract and plan benefit package identifiers, which have different variable names in the PDE and Plan Characteristics data files (i.e., PLAN_CNTRCT_REC_ID and PLAN_PBP_REC_NUM in the PDE file, CONTRACT_ID and PLAN_ID in the Plan Characteristics File, and monthly PTD_CNTRACT_ID_<MM> and PTD_PDP_ID_<MM> in the MBSF).

3. Length of Medicare Part A and/or B fee-for-service coverage

Do you need to be able to observe treatment/receipt of care or to accurately ascertain diagnoses and comorbid conditions? If so, you may wish to select beneficiaries who also had fee-for-service (FFS) Medicare A/B coverage. The Base MBSF (A/B/C/D) indicates the type of Medicare coverage received.

4. Presence of Chronic Conditions

Some researchers may wish to bring in additional information, such as clinical data regarding the patient, to specify a cohort. In order to observe health service utilization information for patients, the Medicare beneficiary must have Medicare Part A and/or Part B FFS coverage. Those enrolled in Medicare Advantage (i.e., Medicare managed care plans) do not have FFS claims – therefore no information regarding their care or clinical diagnoses or procedures is available in the CCW Medicare claims files. Researchers may wish to restrict the cohort to those with a certain amount of A/B coverage exposure; rationale described above. Methods to address applying coverage criteria are explained in greater detail in a Technical Guidance paper called "Getting Started with CMS Medicare Administrative Research Files" on the CCW website (https://www.ccwdata.org/web/guest/technical-guidance-documentation).

Researchers wishing to select a cohort based on presence of particular events or medical conditions of interest may wish to merge information from one or both of the Chronic

Condition segments of the MBSF: 1) CCW conditions. CCW has incorporated twenty-seven (27) condition-specific variables which indicate treatment for common and chronic conditions in the Medicare population[2], and 2) Other Chronic or Potentially Disabling Conditions. CMS developed additional condition measures specifically to enhance research of the Medicare-Medicaid dually enrolled population. Medicare enrollment/eligibility and FFS claims data for each beneficiary are available within CCW (note: Medicare claims data are generally not available if the beneficiary is enrolled in managed care). Refer to the MBSF data file documentation on the CCW website for additional information regarding the CCW Conditions and the Other Chronic or Potentially Disabling Conditions file segments.

Coding specifications for the condition definitions are located on the CCW website (https://www.ccwdata.org/web/guest/condition-categories). For more information about using the conditions variables for various purposes, please refer to a CCW paper called "Technical Guidance for Calculating Medicare Population Statistics" on the CCW website at: https://www.ccwdata.org/web/guest/technical-guidance-documentation.

B. Aggregating Data to Summarize Coverage Variables

Within the MBSF, there are several variables which describe the Medicare coverage the beneficiary selected. These include the Part D plan's contract and benefit package identifiers, the segment (market area) for the plan, the cost share group and the retiree drug subsidy. Each of these variables has 12 monthly indicators for every beneficiary (refer to Table 1).

Table 1. Part D Monthly Coverage Variables in MBSF

Variables with monthly values	**SAS Variable name(s)** *The last 2 <MM> digits are sequential 01-12*
Contract ID*	PTD_CNTRCT_ID_<*MM*>
Benefit package ID*	PTD_PBP_ID_<*MM*>
Segment ID*	PTD_SGMT_ID_<*MM*>
Cost share group	CST_SHR_GRP_CD_<*MM*>
Retiree drug subsidy	RDS_IND_<*MM*>
Dual Medicare-Medicaid enrollment status	DUAL_STUS_CD_<*MM*>
HMO indicator	HMO_IND_<*MM*>
Medicare buy-in indicator	MDCR_ENTLMT_BUYIN_IND_<*MM*>

* Prior to the 2013 data release, these variables were encrypted.

This level of detail may be important for researchers studying the effect of plan changes during the year; however researchers who are not studying these issues in-depth may wish to

[2] As a historical note, when CCW data first became available there were only 21 CCW conditions. When 2010 data became available, we revised 17 of the original 21 chronic conditions (CCs); the updated 27 CCs are available from 1999 forward. The 21 CCW conditions were previously delivered as either part of the BASF – or the Chronic Condition Summary data file).

generalize the monthly coverage information. For example, it may be sufficient to select a cohort based on having at least one month of Part D coverage, or perhaps you would like to require coverage for the full year (or until time of death).

Researchers may select variables from the MBSF as key study variables of interest – such as whether the beneficiary was dually enrolled in Medicare and Medicaid. For other researchers a more summarized version of this information may suffice (e.g., knowing that the beneficiary was dually eligible for at least one month of the year).

There are many options for summarizing beneficiary coverage information:

- Select a particular month of coverage – Researchers may take a "snapshot" of the monthly coverage information and look at one particular month in the year as an indicator of the coverage status for the beneficiary. For example, the mid-point of the year (i.e., coverage for July, using PTD_CNTRCT_ID_07) or the end of the year (i.e., using PTD_CNTRCT_ID_12).

- Any month of coverage – The Beneficiary Summary Files contains a few summary variables which summarize the number of months of particular types of coverage for each beneficiary during the year. These include the # months of Part D plan coverage (PTD_PLAN_CVRG_MONS), dual eligibility (i.e., enrolled in both Medicare and Medicaid; DUAL_ELGBL_MONS), and retiree drug subsidy (RDS_CVRG_MONS). Using these variables, it is simple to obtain denominator counts which take into consideration the number of beneficiaries who EVER had Part D coverage during the year.

- Use a "majority of months" perspective or count member months of enrollment - consider every month of Part D coverage for every beneficiary, accumulate this information for each covered month. Depending on study objectives, you may either make a determination regarding the type of coverage the beneficiary had for the majority of months enrolled in Part D or count total member months for rate calculations. An illustrative example of what this type of analytic code might look like is presented in a CCW paper called "CCW Technical Guidance: Calculating Medicare Population Statistics" on the CCW website at: https://www.ccwdata.org/web/guest/technical-guidance-documentation.

Once you have selected a cohort based on the parameters of interest, decide whether you want to keep all of the variables in this file – some may not be of interest for your study.

Note: The MBSF records can become quite wide if you are combining variables from the different segments (i.e., merging A/B/C/D and the Conditions segment). Furthermore, your data file can become very long if you are intending to merge information regarding utilization or information from other data files.

Chapter 4: Examining Prescription Drug Utilization

Once you have selected your cohort, you can merge in the subset of PDE data that is for your population. Note that not all Part D eligible beneficiaries will have prescription drug event data (i.e., if they did not fill a prescription during the year).

The PDE is a large file, with a row of data for each prescription fill. Therefore there is a many-to-one relationship to the MBSF; a single beneficiary may have hundreds of PDEs (Table 2). We recommend extracting only the PDEs for your cohort of interest - however you have specified your study group, as discussed above.

Table 2. Part D File Size

Data File	# records in 1% (2011)
MBSF Base (A/B/C/D)	517,322
PDE file	11,552,645

For the examples presented in this paper, we select our cohort as a random 1% sample of Medicare enrollees in 2011 with at least one month of Part D coverage.

A. Merging PDE file to Cohort File

You will likely wish to obtain demographic and Medicare enrollment information for your cohort from the MBSF-ABCD segment. Keep only the variables of interest, and merge these 2 files by BENE_ID. Before attempting to merge the PDE file with the denominator/cohort file – you will need to sort both data files (by BENE_ID).

Note: When the data files are delivered from CCW, they are already pre-sorted and indexed by BENE_ID (the PDE file is sorted by BENE_ID and SRVC_DT and PDE_ID). Therefore, if you have only subsetted the files (and not performed other types of data processing), then you may not need to sort the files.

For some studies the PDE file will be extremely large, therefore efficient data management is prudent. For example, an alternative to managing the large data file is to partition the PDE data file into smaller files (e.g., monthly files based on service date; files based on types of NDCs; files for people in certain population subsets) – then sort each file and loop through all the small files, extracting only the PDEs of interest. There may also be times when it is appropriate to restrict your enrollment/cohort data file only to PDE users – to more efficiently examine patterns of utilization.

Some researchers may wish to use SAS® PROC SQL (rather than a SAS DATA step) to more efficiently handle this sort and merging process. This type of processing negates the need for a separate sorting step. Some sample code is presented below.

Note*:* For all of the coding examples, the variable name appears in all capital letters if it is directly from the source (CCW) data file; all other derived variable names may be a combination of upper and lower case letters.

Code Example 1. Join Part D Event Data with Pre-defined Study Denominator or Cohort File

This code example assumes you have already created a denominator file using variables of interest from the MBSF-ABCD segment. It also assumes that when you made your CCW data request, you had the "Drug Characteristics" variables appended to the PDE; this is not required – however we show some examples that include variables from this source.

The following SAS code assumes that you are developing a new merged data file (called cohort2011) that is created by linking the cohort you established (e.g., using whatever denominator specifications you deemed appropriate – *vis a vis* the Denominator Selection Section of this paper; source file called *coverage*) with the Part D events for this study population (called PDE_2011). The LEFT JOIN is important in keeping only PDE records for your specific study cohort.

```
proc sql;
  create table a.cohort2011 as
    select c.*, p.*
     from coverage c
        left join a.PDE_2011 p
    on c.BENE_ID = p.BENE_ID;
     quit;
     run;
```

You can easily verify that your new file has included the Part D events for everyone in your cohort by using a sorting process and removing duplicate beneficiaries:

```
proc sort data= a.cohort2011 out=PartD (keep=BENE_ID PROD_SRVC_ID
GNN  GCDF PDE_ID);
      by BENE_ID PDE_ID;
     run;
```

This will allow you to verify that you have the same number in your cohort AFTER merging in the PDE data as you did when you finalized your cohort/denominator.

You may prefer to use a SAS DATA step to merge these files.

The following example produces the same data file as the SQL code above. To merge using a data step, the files may need to be sorted first using BENE_ID. However, since the CCW data files are pre-sorted, the sort can be omitted for the PDE file. Then, when you merge the files, you specify (using the "in" statements and a subsetting "if") your intention to keep only the PDEs for the beneficiaries in your cohort.

```
proc sort data=coverage out=benes;
  by BENE_ID;
       run;


data ccw.cohort2011;
merge benes (in=b) a.PDE_2011 (in=p);
   by BENE_ID;
    if b;
      run;
```

B. Tabulating or Summarizing PDEs

Researchers have many options for exploring and describing Part D prescription drug utilization. Numerous variables in the PDE make it possible to classify the types of drugs used. There are also many options for the unit of analysis (e.g., prescription use for enrolled beneficiaries, use per member months of enrollment, or average quantity of fills per person who used at least one PDE). Some common methods for counting and summarizing prescription drug utilization are presented.

1. Tabulating Prescription Drug Utilization

Some beneficiaries will not have any PDE utilization information, indicating no prescription drug fills during the time frame of interest. Other beneficiaries will have a very large volume of PDEs. The following examples use the PDE_ID variable, which assumes that you do not need to know what drug was used, only that a prescription drug event (fill) took place. You may modify the code, for example, to use NDC (using the variable PROD_SRVC_ID) or generic name (using the variable GNN, which is available in the Part D Drug Characteristics File), if your study requires you to tabulate # distinct drugs or drug products/entities.

a. Total PDE Counts

In the simplest case, researchers may wish to take the number of Part D enrolled in the cohort, and determine the average number of prescription drug fills per enrolled beneficiary. First, take the number in your Part D enrolled cohort from the Denominator Selection in Section IV. A above. Next, one must determine the total PDEs for the cohort. There are several ways to accomplish this task. Perhaps the simplest approach is to obtain record counts from your cohort file (e.g., use the a.cohort2011 file, from Code Example 1 above), which has been linked with the PDE records. This file will have one record for each PDE. This information may be used for your numerator and denominator to calculate the average number of prescription fills per beneficiary (Table 3).

$$\text{Average number of prescription fills} = \frac{\text{Total PDEs for the cohort}}{\text{Number of unique beneficiaries in cohort}}$$

Table 3. PDE Counts, 2011 (1% beneficiary sample)

Measure	N
# PDEs for cohort	11,552,645
# unique beneficiaries in cohort	517,322
Beneficiaries with PDEs	291,163

Another option is for researchers to create a counter variable to tally each of the PDEs so that per beneficiary (or per user) counts can be obtained.

b. Per beneficiary PDE Counts

To more discretely tabulate the per-beneficiary number of prescriptions, beneficiary-specific (i.e., Part D enrollee-level or per capita) counts of PDEs can be obtained.

Code Example 2. Calculate Number of PDEs per Beneficiary

The following SAS code sorts the PDEs for each beneficiary and creates a small data file, so that beneficiary-level prescription fill counts can be obtained. We use the SAS keywords "first" and "last" for this new counter variable called numPDE, and create a trim little file called a.PDEslim2011that is specific to this purpose.

Note that we also use the SAS key word "retain" so that our new variable, numPDE, is NOT automatically re-initialized as SAS works through sequentially processing the data file. While technically not required for this sort of counter variable, it is a good coding practice that will allow users to easily adapt this sample code. This "retain" statement allows for variables that do not appear on the input statement to retain or hold the current value while SAS moves on to process the next row of data.

Finally, we use a PROC MEANS statement to calculate the number of PDEs per beneficiary in our sample.

```
proc sort data= a.cohort2011 out=PartD (keep=BENE_ID PROD_SRVC_ID
GNN  GCDF PDE_ID);
      by BENE_ID PDE_ID;
run;

data a.PDEslim2011;
 set PartD;
  by BENE_ID  PDE_ID;

   retain numPDE;

     if first.BENE_ID then  numPDE =0;
```

```
      if PDE_ID ~=' ' then numPDE + 1;
      if last.BENE_ID then output;

label
   numPDE = 'per bene count of PDEs';

run;

 proc means data=a.PDEslim2011 N mean median max min sum;
  var numPDE;
  title 'average number of prescriptions filled per bene';
 run;
```

Those who prefer to use SQL statements for this type of processing may accomplish the same objectives through using something like the following:

> The following SQL code produces a similar file as the DATA step above. With SQL code, you must have an additional data step to handle the BENE_IDs with null PDE_IDs, that is, those without any PDEs.
>
> Summarize the data as above, using the PROC FREQ and/or MEANS.

```
proc sql;
  create table a.PDEslim2011 as
     select BENE_ID,
      count(PDE_ID) as pde_count
     from PartD
     group by BENE_ID;
     quit;

     *recodes to accommodate beneficiaries without PDEs;
     data a.PDEslim2011;
       if pde_count = . then pde_count=0 ;
       if pde_count  >0 then user='y';
         else user='n';
     run;
```

c. Per Prescription Drug User PDE Counts

Different questions related to Part D utilization can be addressed by limiting the sample to those who have used the benefit (i.e., those who had one or more PDEs during the time frame). This sort of analysis can be easily performed using a slight variation on the PROC MEANS statement, used previously.

Code Example 3. Calculate Number of PDEs per Part D Prescription Drug User

> In the PROC MEANS statement, limit the analysis to beneficiaries who had a PDE. This is accomplished by using a WHERE statement.

```
proc means data=a.PDEslim2011 N mean median max min sum;
      var numPDE;
      where PDE_ID ~=' ';
      title 'average number of PDEs per user';
run;
```

Results from this processing are shown in Table 4below.

Table 4. Per beneficiary and per Part D user PDE counts

Measure	# beneficiaries	2011 random 1%
Average PDE fills per bene	517,322	mean=22.33 median=7 max=726
Average PDE fills per user	291,163	mean 39.68 median=29 max=726

2. Describing Prescription Drug Utilization

a. Total Prescription Fills by Drug or Generic Name

There are many potential ways researchers may wish to count Part D prescription drug utilization. Researchers may be interested in the total number of drug fills (as above). Other options include determining the number of fills by NDC (e.g., distinct national drug codes [NDCs] – all 11 digits or a subset of the digits, such as the first 9 digits), or by generic drug names. Note that a beneficiary may obtain a drug that has the same active ingredients and is represented by many NDCs. For example, hundreds of unique NDC codes may all be considered antibacterials, many of which may contain the same active ingredients. Researchers must determine the appropriate level of aggregation for the study.

- The NDCs are contained in the PROD_SRVC_ID field.
- The generic drug name is contained in the GNN field, and is included in the Drug Characteristics file

Code Example 4. Determine Number of Different Drugs and Refills Obtained per Beneficiary

This SAS code is a variation on the PDE counting code above. Here, we sift through the PDEs and count the occurrences of each NDC. We set up variables to accumulate counts for the distinct NDCs.

First you must sort your data file by NDC (PROD_SRVC_ID) or generic name (GNN), depending on your variable of interest. This is the file (called 'pdendc') we use as our input source for the DATA step below.

```
proc sort data=a.cohort2011 (keep=PROD_SRVC_ID GNN ) out=pdendc;
 by PROD_SRVC_ID;
run;

data a.ndc_fills;
 set pdendc;
 by PROD_SRVC_ID;
  retain ndcfillcnt;

if PROD_SRVC_ID ne '';
  if first.PROD_SRVC_ID then ndcfillcnt=0;
  ndcfillcnt+1;
if last.PROD_SRVC_ID then output;

label ndcfillcnt='Number of fills per NDC';

run;

proc univariate data=a.ndc_fills;
  var ndcfillcnt;
 title 'Number of Fills per NDC';
run;
```

This data file will allow researchers to understand the most commonly observed types of drugs in the PDE file. Results from this processing are shown in Table 5, Table 6, and Table 7, below.

a. Per Prescription Drug User Counts by Drug or Generic Name

Researchers may wish to know which particular prescriptions a beneficiary has received, and the number of refills for each type of drug. This analysis is limited to those beneficiaries who had at least one prescription drug fill.

Code Example 5. Determine Number of Different Drugs and Refills Obtained per Part D Prescription Drug User

The following SAS code is a variation on the NDC counting code above. Here, we process every row of PDE data and count the occurrences of each NDC for each beneficiary. We set up variables to accumulate counts for both the distinct NDCs per beneficiary and the frequency of each NDC (e.g., number of fills for each NDC) for the beneficiary.

First you must sort your data file by BENE_ID and NDC (PROD_SRVC_ID). This is the file (called 'pdefills') we use as our input source for the data step below.

To accurately tabulate the number of NDCs per beneficiary, we need to identify the last of the NDC records for each beneficiary. This is the utility of the "L" variable.

```
proc sort data=a.cohort2011(keep=BENE_ID PROD_SRVC_ID GNN PDE_ID)
out=pdefills;
 by BENE_ID PROD_SRVC_ID;
run;

data a.beneNDC;
 set pdefills;
  by BENE_ID  PROD_SRVC_ID;
   retain numNDC numfills;

    if first.BENE_ID then do;
          numNDC =0; L=0;
       end;
      if PDE_ID NE ' ' then do;

         if first.PROD_SRVC_ID then do;
       numNDC + 1; numfills=0;
        end;
       if PDE_ID NE ' ' then do;
          numfills+1;
             if last.BENE_ID then L=1;
          output;
        end;
       end;

label
   numNDC = 'per bene count of unique NDCs'
   numfills = 'number of fills per bene per NDC'
   L='last NDC record per bene';
run;
```

An example of the values for the new variables produced by the DATA step is shown in Table 5.

Table 5. Example of Data Row – Variables for Number of NDCs and Number of Fills

Obs #	BENE_ID	PROD_SRVC_ID	PDE_ID	numNDC	Numfills	L
1	aaaaaaaaaaaaaaa	12345678910	Xxxxxxxxxxxxx1	1	1	0
2	aaaaaaaaaaaaaaa	12345678910	Xxxxxxxxxxxxx2	1	2	.
3	aaaaaaaaaaaaaaa	12345678910	Xxxxxxxxx1xxxx	1	3	.
4	aaaaaaaaaaaaaaa	22222233333	Xxxx5xxxxxxxx1	2	1	.
5	aaaaaaaaaaaaaaa	22222233333	XxxxxxxQxxxxx	2	2	.
6	aaaaaaaaaaaaaaa	55667788991	xhxxxfxxxxxxxx	3	1	1
7	bbbbbbbbbbbbbbb	44444444444	X1xxxxxxxxxxxx	1	1	0
8	bbbbbbbbbbbbbbb	44444444444	X1xxxxxxxxxxxx	1	2	1
9	ccccccccccccccc	07070707071	Xnxxxxxxxx3xx	1	1	1

For each PDE, there is a combination of values to indicate whether it is a new NDC or a refill of an NDC already obtained. Table 5 illustrates that for the first beneficiary (observations 1-6 indicating 6 rows of data for the 6 PDEs), there were 3 different NDCs dispensed (highest value for numNDC is 3 – see row 6); the first NDC was obtained 3 times (highest value for numfills for NDC 1 is 3 – see row 3).The record where L=0 is the first PDE record for the beneficiary; when L=1 it is the last PDE record for the beneficiary. For NDCs with a single fill, L=1.

Using this data file, per user number of NDCs can be tabulated.

```
proc means data=a.beneNDC N mean median max min sum;
 where L=1;
  var numNDC;
 run;
```

Results are shown in Table 6.

Table 6. Number of NDCs and Per User NDC and Fill Totals

Measure	n	2011 random 1%
Average # of fills per NDC	21,145 different NDCs	mean=546.35 median=29 max=66,190
Average # different NDCs (11-digit) per user	291,163 beneficiaries	mean=12.23 median=10 max=111

In our study population, there were 21,145 different NDCs dispensed. An NDC appeared, on average, 546 times in our data file, although one NDC had over 66 thousand PDEs in our sample.

Code Example 6. Calculate Per Capita and Per User Statistics for Prescription Drug Cost and Use

The following SAS code is a variation on the NDC counting code above, except here we use SAS SQL rather than a DATA step, and we count the number of unique drug products, using the generic name of the drug (variable called GNN). During a single SQL submit, we are able to sift through the PDEs and count the occurrences of each GNN for each beneficiary. We also set up a variable to accumulate total drug costs for the PDEs, to demonstrate that multiple processing steps can be efficiently handled in a single PROC.

When using SAS SQL, it is not necessary to sort the input data file first. A DATA step is used to include the null counts for beneficiaries without PDEs, to allow for flexibility in producing summary statistics (i.e., per beneficiary counts or per user counts).

```
proc sql;
           create table a.Drugs11 as
           select
            BENE_ID,
            count(PDE_ID) as pde_count,
            count(distinct GNN) as drug_count,
            sum(TOT_RX_CST_AMT) as gross_cost_sum
           from a.cohort2011

           group by BENE_ID;
           quit;

           *recodes for non-users;
     data Drugs11;
        set a.Drugs11;
           if pde_count = . then pde_count=0 ;
           if drug_count = . then drug_count=0;
           if gross_cost_sum  = . then gross_cost_sum=0 ;

     if pde_count>0 then user='y';
       else user='n';

  run;

proc means data=Drugs11 n mean median max min sum;
           var pde_count drug_count  gross_cost_sum;
           title 'average utilization and cost per beneficiary';
           run;

proc means data=Drugs11 n mean median max min sum;
           var pde_count drug_count  gross_cost_sum;
                 where user='y';
           title 'average utilization and cost per PDE user';
           run;
```

Results are shown in Table 7.

Table 7. Number of Drugs (GNN – Generic Drug Products) Per Beneficiary and Per Part D User

Measure	n	2011 random 1%
Average # different drug products (GNNs) per bene	517,322 beneficiaries	mean=5.24 median=3 max=79
Average # different drug products (GNNs) per user	291,163 beneficiaries	mean=9.32 median=8 max=79

Researchers may wish to be very specific regarding whether prescription drug fills indicated that an "effective" dose was likely used or whether the beneficiary was taking a medication as often as necessary to manage a particular condition (note that these data indicate only that the prescriptions were filled, and are not proof that the beneficiary was taking the medications as directed). The PDE file contains information regarding the quantity of the medication dispensed during the fill (QTY_DSPNSD_NUM) and the number of days supply provided by the fill (DAYS_SUPLY_NUM). Additionally, information regarding the dosage form (GCDF and GCDF_DESC) and strength (STR) are available in the Part D Drug Characteristics File.

3. Describing Costs of Prescription Drugs

The Part D benefit is a risk-adjusted prospective capitated benefit, with Part D plans generally receiving a fixed annual payment for each enrolled beneficiary. Note that a final reconciliation at the end of the year may occur, whereby Medicare may alter the final paid amount paid to plans based on actual payments. The costs shown on the PDEs do not correspond with plan payments (i.e., sum of drug costs for all beneficiaries enrolled in a plan does not equal plan payments).

There are times when it may be helpful to know how much the beneficiary paid for specific prescription drugs, or to understand the gross (total) cost of drugs for a beneficiary during the year. The PTNT_PAY_AMT is the data field which indicates the amount the beneficiary paid for the prescription drug. The TOT_RX_CST_AMT is the gross drug cost, which is the sum of three fields: 1) ingredient cost, 2) dispensing fee, and 3) sales tax.

A sample of some SQL code to calculate the total costs for PDEs was included previously in Code Example 6 of this paper. The TOT_RX_CST_AMT field was summed for each beneficiary. Results are as follows (Table 8):

Table 8. Gross Drug Costs

Measure	n	2011 random 1%
Average gross drug costs per beneficiary	517,322 beneficiaries	mean=$1636 median=$158
Average gross drug costs per user	291,163 beneficiaries	mean=$2907 median=$1480

4. Describing Experience with the Part D Benefit Structure

Researchers may wish to understand how many beneficiaries were exposed to particular phases of the Part D benefit. The BENEFIT_PHASE variable is a CCW derived field that considers the plan benefit structure for each beneficiary at the time of the fill, and indicates where in the benefit cycle each PDE occurred. This variable is in the Part D Event File. For a standard Part D benefit, there are four phases that a beneficiary may

experience: the deductible, pre-ICL, ICL (initial coverage limit – or coverage gap), and catastrophic coverage.

Different data manipulation will be necessary depending on the specific question to be addressed. Also, this data field will occasionally contain unexpected information (e.g., a phase that appears to be out of sequence) because a small number of beneficiaries may switch plans during the year, in which case the benefit phase sequencing "clock" starts over.

Code Example 7. Ordering Part D Benefit Phases

The BENEFIT_PHASE values contain intelligence but are not ordinal; therefore it may be helpful to create a recoded ordinal variable containing benefit phase. There are a couple of different ways to do this – using the first digit of the variable to indicate the starting benefit phase for the fill, or using the second digit of the variable to indicate the ending benefit phase for the fill. For the following examples, we use the ending benefit phase for the fill.

The following SAS code creates an ordinal benefit phase variable called "bpord". A PROC FORMAT statement is used to classify the ordinal benefit phase variable into the four standard benefit phases based on the last digit of the benefit phase variable. For most purposes, researchers may wish to hard code the four phases, rather than maintaining a very granular bpord variable; this would provide for a cleaner way of counting the number of beneficiaries to reach each distinct standard phase.

Note that beneficiaries without PDEs are included in this file, to allow for flexibility in calculating rates.

```
data phase;
 set a.cohort2011 (keep=BENE_ID SRVC_DT BENEFIT_PHASE PDE_ID);

if BENEFIT_PHASE='DD' then bpord=11;
else if BENEFIT_PHASE='DP' then bpord=21;
else if BENEFIT_PHASE='DI' then bpord=31;
else if BENEFIT_PHASE='DC' then bpord=41;
else if BENEFIT_PHASE='PP' then bpord=22;
else if BENEFIT_PHASE='PI' then bpord=32;
else if BENEFIT_PHASE='PC' then bpord=42;
else if BENEFIT_PHASE='II' then bpord=33;
else if BENEFIT_PHASE='IC' then bpord=43;
else if BENEFIT_PHASE='CC' then bpord=44;
else bpord=0;

proc format;
 value order low-9='no phase'
        10-19='deductible'
        20-29 ='pre-ICL'
        30-39 ='ICL'
```

```
            40-high='catastrophic';
run;
```

This data file contains information for every benefit phase experienced by each beneficiary. Since we included all beneficiaries in this data file, members of our population without any PDEs are counted as having bpord=0. For several of our analyses, we will want to remove these rows (i.e., where there is a record but not a PDE_ID).

```
/*tabulate after excluding records for beneficiaries without any
PDES*/

proc freq data =phase;
     tables bpord;
     where PDE_ID ~=' ';
     format bpord order.;
run;
```

Results are in Table 9.

Table 9. Benefit Phases for Each PDE

PDEs in each Benefit Phase n=11,552,645 PDEs	#	%
No phase (phases not applicable)	965,625	8.4
Deductible	783,971	6.8
Pre-ICL	6,900,926	59.7
ICL	1,894,985	16.4
Catastrophic	1,007,138	8.7

A total of 8.7% of all PDEs were in the catastrophic coverage phase of the Part D benefit.

b. Number of Beneficiaries to Reach Each Phase

To determine the range of benefit phases experienced by members of our study population, we will want to keep only the first prescribing event in each benefit phase for each beneficiary in the cohort. Beneficiaries may be counted in multiple phases. You may be as granular as you would like in defining these phases; for simplicity we use the four standard phases.

Code Example 8. Determining Part D Benefit Phases for Each Beneficiary

For this code, we must use a more specific benefit phase variable than we used in the prior analyses. Here we will use only the four distinct standard benefit phases. We also eliminate rows of data for beneficiaries without any PDEs (i.e., where there is not a PDE_ID).

Next, the data file is sorted so that the PDEs are in order of service date for each beneficiary. We use the SAS keyword “first” in order to keep the first PDE record for each beneficiary in each of the benefit phases applicable. This file outputs a record for each beneficiary: phase combination. Researchers may also wish to partition the data into smaller files based on the specific benefit phase of interest.

```
data b_phase;
set phase;

if PDE_ID ~=' ';

 if bpord in (0) then BP=0;
   else if bpord in (11) then BP=1;
   else if bpord in (21,22) then BP=2;
   else if bpord in (31,32,33) then BP=3;
   else if bpord in (41,42,43,44) then BP=4;

proc format;
  value b 0='no phase'
         1='deductible'
         2 ='pre-ICL'
         3 ='ICL'
        4-high='catastrophic';
       run;

 proc sort data=b_phase out=small;
  by BENE_ID  bp SRVC_DT;
 run;

  data benephase;
  set small;
  by BENE_ID bp;
  if first.bp;
  run;

 proc freq data=benephase;
  tables bp;
  format bp b.;
  title 'first event each phase';
   run;
```

Results are in Table 10.

Table 10. Each Benefit Phase for Each Beneficiary

Beneficiaries reaching each Benefit Phase* n=291,163 beneficiaries with PDEs 489,635 beneficiary: phase combinations		
No phase (phases not applicable)	38,030	7.8
Deductible	104,273	21.3
Pre-ICL	243,111	49.7
ICL	78,419	16.0
Catastrophic	25,802	5.3

*Beneficiaries may fall into multiple phase categories during the year; %s not meaningful in this example.

78,419 beneficiaries had a PDE in the ICL phase. Note that the numbers and percentages from this code are interesting primarily from a reference standpoint; we use this code primarily to show how to identify the first event in each phase – which allows for calculation of some timing variables (e.g., Code Example 11).

c. Highest Benefit Phase Reached by Each PDE User

Researchers will want to keep only the highest (worst) benefit phase reached by each beneficiary.

For this analysis we use the data set that has been sorted previously (above) – by BENE_ID bp and SRVC_DT.

This time we look at the last event for each beneficiary.

```
data bp_last;
 set small;
 by bene_id ;
 if last.BENE_ID;
run;

proc freq data =bp_last;
  tables bp;
  title 'for last event';
  format bp b.;
 run;
```

The processing above is limited the sample to those with at least one PDE. Results appear in Table 11.

Table 11. Highest Benefit Phase for Each PDE User

Highest (or worst) benefit phase for each beneficiary with PDEs n=291,163 beneficiaries (PTD Users)	**n**	**%**
No phase (phases not applicable)	38,896	9.9
Deductible	18,529	6.4
Pre-ICL	164,917	56.6
ICL	53,019	18.2
Catastrophic	25,802	8.8

A total of 8.8% of beneficiaries who used any Part D drugs reached the catastrophic coverage phase, an additional 18.2% reached the coverage gap (i.e., the ICL).

d. Time to Reach a Phase

Researchers may wish to determine how long it took beneficiaries to use enough prescription drugs to cross into a particular phase of the benefit. Similarly, it may be desirable to know the length of time a beneficiary was considered to be in a particular benefit phase.

- In general, researchers may wish to calculate "time to event" in terms of calendar time (which corresponds to the Part D benefit year). For example, the following code could be used to calculate the time to reach the ICL (initial coverage limit; coverage gap):

Code Example 10. Determining Time to Reach a Benefit Phase (time to coverage gap)

We make use of the ordinal benefit phase variables we created in Code Example 8, and leverage the data file we created during that step (file called *b_phase*).

The following SAS code keeps only the subset of beneficiaries that reached the benefit phase of interest (this example uses the ICL). A small file such as this, which is specific to any benefit phase of interest, can be created. We use the SAS "first" function as well as the "intck" function.

```
/*time to reach the coverage gap -or ICL benefit phase*/
data ICL;
      set b_phase;
      where bp=3;
      by BENE_ID bp SRVC_DT;


      if first.BENE_ID;
      gaptime=intck('day',mdy(1,1,2011),SRVC_DT);
      format gaptime month.;
      label gaptime='month to reach gap';
run;

proc freq data=ICL;
      tables gaptime;
      title 'all who reached gap - timing';
                run;
```

The frequency output is shown below (Table 12).

Table 12. Month to Reach Coverage Gap (ICL benefit phase)

gaptime	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	1291	1.65	1291	1.65
2	3543	4.52	4834	6.16
3	5951	7.59	10785	13.75
4	6602	8.42	17387	22.17
5	8095	10.32	25482	32.49
6	8414	10.73	33896	43.22
7	8066	10.29	41962	53.51
8	8142	10.38	50104	63.89
9	7476	9.53	57580	73.43
10	7661	9.77	65241	83.20
11	6914	8.82	72155	92.01
12	6264	7.99	78419	100.00

In the first month of 2011, 1.65% of beneficiaries who would eventually experience the coverage gap (the ICL) benefit phase had done so. June (i.e., where gaptime=6) was the month when the largest number of people reached the ICL phase.

e. Time in a Phase

A variation on this code will allow researchers to examine the amount of time spent in each benefit phase.

Code Example 11. Determining Duration of Part D Benefit Phases for Each Beneficiary

We begin with the input data file created in Code Example 9, above (file called *small*). The file has already been sorted by beneficiary (BENE_ID) and benefit phase (bp). We create several new variables which key in on the particular dates of certain events for each beneficiary. We use the "retain" statement to keep the resulting dates of interest.

```
/*highest BP per beneficiary*/

     data a.bp2011;
          length  bp bppdecnt 3 bp_start_dt bp_end_dt 4 bp_days
     days2phase 3;
          set small;
          by BENE_ID bp;
          retain bp_start_dt;
```

```
      if first.bp then
            do;
                  bp_start_dt=SRVC_DT;
                  bppdecnt=0;
            end;

      bppdecnt+1;

      if last.bp then
            do;
                  bp_end_dt=SRVC_DT;
                  bp_days=intck('day',bp_start_dt,bp_end_dt)+1;

      days2phase=intck('day',mdy(1,1,2011),bp_start_dt);
                  output;
            end;

      label
            bp='Benefit Phase'
            bppdecnt='Number of PDEs in Phase'
            bp_start_dt='Earliest Service Date of Phase'
            bp_end_dt='Latest Service Date of Phase'
            bp_days='Number of Days in Phase'
            days2phase='Calendar Days to Start of Phase';

      format bp_start_dt bp_end_dt mmddyy10.;
run;
```

Then, to find counts and averages for selected analysis variables we can perform the following:

```
/*tabulations*/
proc sort data=a.bp2011 out=temp;
      by bp;
run;

proc means data=temp maxdec=2;
      by bp;
      var bppdecnt bp_days days2phase;
            run;
```

The output of this step is shown below (Table 13).

Table 13. Time to Reach Coverage Gap (ICL) and Time in Gap

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
bppdecnt	Number of PDEs in Phase	78419	24.16	19.26	1.00	269.00
bp_days	Number of Days in Phase	78419	106.24	64.17	1.00	357.00
days2phase	Calendar Days to Start of Phase	78419	201.75	91.50	0	364.00

- Another way to examine time to a phase (as "time to event") is to look at the number of benefit months before hitting a particular phase; that is, the member months of enrollment before the benefit phase event first occurred. This type of analysis may be helpful if you have a cohort which includes a large population of those newly enrolled in Medicare.

  Readers may refer to a previous CCW Technical Guidance paper (see CCW website) for considerations and examples for constructing member months of enrollment variables.

Chapter 5: Conclusions

Researchers have a variety of hypotheses and objectives. The intent of this paper is not to be prescriptive, but rather descriptive of some useful tools for combining data files and tabulating Part D prescription drug data. The objective is to make it easy for researchers to accomplish their study objectives – and to ensure they are able to do so with a thorough understanding of the data available from the CCW. Using the methods described in this paper, we can gain a better understanding of prescription drug use, and the effect on the population, through appropriate and accurate data analysis techniques.

Chapter 6: Further Assistance with CMS Administrative Claims Data

The Research Data Assistance Center (ResDAC) offers free assistance to those using Medicare data for research. The ResDAC website provides links to descriptions of the CMS data available, request procedures, supporting documentation, workshops on how to use Medicare data and other helpful resources. Visit the ResDAC web site at (http://www.resdac.org) for additional information.

ResDAC is a CMS contractor and requests for assistance in the application, obtaining, or using the CCW data should first be submitted to ResDAC. Investigators can reach ResDAC by phone at 1-888-973-7322, e-mail at resdac@umn.edu, or online at (http://www.resdac.org).

In the event that a ResDAC technical advisor is not able to answer the question, the technical advisor will direct the investigator to the appropriate person. If additional CMS data (data not available from the CCW) is required to meet research objectives, or the investigator has any questions about other data sources, the investigator can review all available CMS data by visiting the ResDAC website and contact ResDAC for further assistance.

www.ccwdata.org
Email: CMSdata@gdit.com
Phone: 1-866-766-1915

Appendix A: List of Acronyms and Abbreviations

Acronym	Definition
CCW	Chronic Condition Data Warehouse
CMS	Centers for Medicare &Medicaid Services
DUA	Data Use Agreement
FFS	Fee-for-Service
ICL	Initial Coverage Limit
MA-PD	Medicare Advantage – Prescription Drug Plan
MBSF	Master Beneficiary Summary File
MMA	Medicare Modernization Act of 2003
NCPDP	National Council for Prescription Drug Programs
NDC	National Drug Code
PACE	Program of All-inclusive Care for the Elderly
PDE(s)	Prescription Drug Event(s)
PDP	Prescription Drug Plan
RDS	Retirement Drug Subsidy
ResDAC	Research Data Analytic Center
SAF	Standard Analytic File
VA	Veteran's Administration