



CMS Chronic Condition Data Warehouse

Technical Guidance for Researchers Summarizing and Describing Prescription Drug Utilization

*A technical guidance paper developed under contract with
the Centers for Medicare & Medicaid Services*

Kathy Schneider, PhD

Brian O'Donnell, PhD

Jean O'Donnell, MS

Christopher Powers, PharmD*

** Christopher Powers is an employee of CMS – Office of Research and Demonstrations (ORDI)*

Buccaneer, A Vangent Company
Software Development & Informatics Division

1401 50th Street, Suite 200
West Des Moines, IA 50266
(515) 645-3010
www.ccwdata.org

December 2010

Table of Contents

- I. Introduction 2
- II. Objectives..... 2
- III. Contents of CCW Part D Data Files..... 3
- IV. Methods..... 4
 - A. Defining Cohort..... 4
 - 1. Denominator Selection 4
 - a. Length of Part D coverage 4
 - b. Type of Part D Coverage..... 4
 - c. Length of Medicare Part A and/or B fee-for-service coverage 5
 - d. Presence of Chronic Conditions 5
 - 2. Aggregating Data to Summarize Coverage Variables 5
 - B. Examining Prescription Drug Utilization 7
 - 1. Merging PDE file to Cohort File..... 7
 - 2. Tabulating or Summarizing PDEs 8
 - a. Tabulating Prescription Drug Utilization 8
 - 1) Total PDE Counts..... 8
 - 2) Per beneficiary PDE Counts..... 9
 - 3) Per Prescription Drug User PDE Counts..... 10
 - b. Describing Prescription Drug Utilization 11
 - 1) Total Prescription Fills by Drug or Generic Name..... 11
 - 2) Per Prescription Drug User Counts by Drug or Generic Name 12
 - c. Describing Experience with the Part D Benefit Structure 15
 - 1) Number of Beneficiaries to Reach Each Phase 16
 - 2) Highest Benefit Phase Reached by Each PDE User 17
 - 3) Time to Reach a Phase 18
 - 4) Time in a Phase 19
 - d. Describing Costs of Prescription Drugs 20
- V. Conclusions 21

Technical Guidance for Researchers

Summarizing and Describing Prescription Drug Utilization

I. 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 through 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 summer 2008.

Part D enrollment and utilization counts are published on the CCW website. 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, which is included with all Part D data files, provides background information regarding the Part D benefit and the data variables available for researchers. This Technical Guidance paper should be used in conjunction with the Part D User Guide (see ccw website at: <http://www.ccwdata.org/datadoc.php>).

Due to the recency of the Part D benefit and availability of the data files for researchers, very few researchers have experience in working with the data. This Technical Guidance 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.

II. Objectives

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

- 1) Specify a study cohort based on Part D coverage criteria,
- 2) Link the cohort/enrollment file to the Part D utilization file (known as the Prescription Drug Event File), 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.

III. Contents of CCW Part D Data Files

Data for 100% of Medicare-enrolled beneficiaries is 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, on p.3 at

http://www.ccwdata.org/downloads/CCW_UserManual.pdf . 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 Beneficiary Summary File.

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 noncovered drugs may be included in the PDE files (e.g., supplemental drugs reported by enhanced alternative plans, over-the-counter drugs). CCW obtains the PDEs from CMS late the following year in a standard analytic file (SAF) file. 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 “characteristics” data files. Researchers may choose to have some “event characteristics” appended to, or accompanying, the PDE file they obtain, which describe the drug (i.e., brand name, generic name, dosage form and strength), and how the beneficiary experienced his/her particular drug benefit at the time the prescription was filled (e.g., benefit phase, utilization management variables).

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).

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.

IV. Methods

A. Defining Cohort

Throughout this paper, the assumption is that we begin with a standard Medicare 5% random sample (i.e., where STRICT_SAMPLE_FLAG='05'). This may not be important to you if you have requested data for some other cohort. We are using both the CCW Beneficiary Summary file and the PDE file for 2007.

1. Denominator Selection

An ideal file to use to make your cohort selection is the CCW Beneficiary Summary File with the appended Part D coverage data elements. There is a single row of data for each beneficiary in the population, identified by a unique encrypted beneficiary identification number (BENE_ID).

Beneficiary Summary File – The Beneficiary Summary File 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. This file contains a variable that indicates whether a beneficiary was included in the CMS 5% sample for 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 ([click here](#)).

a. 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. For example:

b. 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., 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 (PTD_CNTRCT_ID_XX on the Beneficiary Summary file; the first digit is “E”). More detailed information regarding the type of benefit offered by the plan is available in the Plan Characteristics data file (see particularly variables such as ORGANIZATION_TYPE or PLAN_TYPE).

c. 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 Medicare A/B coverage. The Beneficiary Summary file indicates the type of Medicare coverage received.

d. Presence of Chronic Conditions

Some researchers may wish to bring in additional information, such as clinical data regarding the patient, to make a cohort determination. In order to observe health service utilization information for patients, the Medicare beneficiary must have Medicare Part A and/or Part B fee-for-service coverage. Those enrolled in Medicare Advantage (i.e., Medicare managed care plans) do not have fee-for-service claims – therefore no information regarding their care or clinical diagnoses or procedures is available in the CCW claims files. Researchers may wish to restrict the cohort to those with a certain amount of A/B coverage exposure (similar to the options section 1, described above; methods to address applying coverage criteria are explained in greater detail in a previous Technical Guidance paper - see the CCW website).

Researchers wishing to select a cohort based on presence of particular events or medical conditions of interest may wish to merge information from the Beneficiary Annual Summary File (BASF – formerly delivered as the Chronic Condition Summary data file) – which has 21 condition categories pre-coded for use. Refer to BASF data file documentation on the CCW website.

Coding specifications for the chronic condition definitions appearing in the BASF are also located on the CCW website ([click here](#)). For more information about applying coverage restrictions, please refer to a previous Technical Guidance paper on the CCW website at: <http://www.ccwdata.org/techguidance.php>

2. Aggregating Data to Summarize Coverage Variables

Within the Beneficiary Summary file, there are several variables which describe the Medicare coverage the beneficiary selected. These include the 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.

Variables with monthly values	SAS Variable name(s) <i>The last 2 digits are sequential 01-12</i>
Contract ID*	PTD_CNTRCT_ID_xx
Benefit package ID*	PTD_PBP_ID_xx

Segment ID*	PTD_SGMT_ID_xx
Cost share group	CST_SHR_GRP_CD_xx
Retiree drug subsidy	RDS_IND_xx
Dual	DUAL_STUS_CD_xx
HMO indicator	BENE_HMO_IND_xx
Medicare buy-in indicator	BENE_MDCR_ENTLMT_BUYING_IND_xx

* Encrypted variables.

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 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 Beneficiary Summary file 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 (PLAN_CVRG_MOS_NUM), dual eligibility (i.e., enrolled in both Medicare and Medicaid; DUAL_ELGBL_MOS_NUM), and retiree drug subsidy (RDS_CVRG_MOS_NUM). 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 previous Technical Guidance paper on the CCW website at: <http://www.ccwdata.org/techguidance.php> (e.g., page 8).

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 record can become very long if you are intending to merge information regarding utilization or information from other data files.

B. 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 beneficiary file; a single beneficiary may have hundreds of PDEs. We recommend extracting only the PDEs for your cohort of interest - however you have specified your study group, as in Section A above.

Data File	# records in 5% (2007)
Beneficiary Summary File	2,581,756
PDE file	55,052,131

For the examples presented in this paper, we select our cohort as Medicare enrollees in 2007 from the 5% sample, with at least one month of Part D coverage.

1. Merging PDE file to Cohort File

Before attempting to merge the PDE file with the denominator/cohort file – you will need to sort both data files (by BENE_ID). For some researchers the PDE file will be extremely large, and special data management activities may be in order. For example, 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.

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 comorbid condition 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 data file only to PDE users – to more efficiently examine patterns of utilization.

The following SAS code assumes that you are developing a new merged data file (called ccw.cohort2007) that is created by linking the cohort you established (e.g., using whatever denominator specifications you deemed appropriate – *vis a vis* Section I of this paper; source file called ccw.PTDcohort2007) with the Part D events for this study population (called ccw.PDE2007). The left join is important in keeping only records for your cohort.

```
proc sql;
  create table ccw.cohort2007 as
  select c.*, p.*
  from ccw.PTDcohort2007 c
  left join ccw.PDE2007 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=ccw.cohort2007 out=benes (keep=bene_id) nodupkey;
  by bene_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 must be sorted first using BENE_ID. Then, when you merge the files, you only capture the PDEs for the beneficiaries in your cohort.

```

proc sort data=ccw.PTDcohort2007 out=benes;
  by bene_id;
run;

proc sort data= ccw.PDE2007 out=PDE;
  by bene_id;
run;

data ccw.cohort2007;
merge benes (in=b) PDE (in=p);
  by bene_id;
  if b;
run;

```

2. 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). Some common methods for counting and summarizing prescription drug utilization are presented.

a. 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), if your study requires you to tabulate # distinct drugs or drug products/entities.

1) 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 Section 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 (see `ccw.cohort2007`, from section 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.

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

Table 2.a.1. PDE Counts

	2007 random 5%
<i># PDEs for cohort</i>	48,466,056
<i># unique beneficiaries in cohort</i>	1,306,431
<i>Beneficiaries with PDEs</i>	1,191,661

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.

2) Per beneficiary PDE Counts

To more discretely tabulate the per-beneficiary number of prescriptions, beneficiary-specific counts of PDEs can be obtained.

The following SAS code sorts the PDEs for each beneficiary and creates a small data file, so that bene-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 `ccw.PDEslim2007` that 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.

```
proc sort data= ccw.cohort07 out=PartD (keep=bene_id prod_srvc_id gmn
gcdf pde_id);
  by bene_id pde_id;
run;
```

```
data ccw.PDEslim2007;
  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 freq data=ccw.PDEslim2007;
  tables numPDE;
  title 'number of prescriptions filled';
run;

proc means data=ccw.PDEslim2007 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 ccw.PDEslim2007 as
  select bene_id,
         count(pde_id) as pde_count
  from PartD
  group by bene_id;
quit;

*recodes to accommodate beneficiaries without PDEs;
data ccw.PDEslim2007;
  if pde_count = . then pde_count=0 ;
  if pde_count >0 then user='y';
  else user='n';
run;

```

3) 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.

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

```

proc means data=ccw.PDEslim2007 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 the table below.

Table 2.a.3. Per Beneficiary and Per User PDE Counts

<i>Measure</i>	<i># beneficiaries</i>	<i>2007 random 5%</i>
<i>Average PDE fills per bene</i>	1,306,431	<i>Mean=37.098; median=28; max=842</i>
<i>Average PDE fills per user</i>	1,191,661	<i>Mean 40.67; median=31; max=842</i>

b. Describing Prescription Drug Utilization

1) 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 in #1 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), 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

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=ccw.cohort07 (keep=prod_srvc_id gnn) out=pdendc;
  by prod_srvc_id;
run;

```

```

data ccw.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=ccw.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 2.b.2.a, below.

2) 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.

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 'NDCs') 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.

```

data ccw.beneNDC;
  set pdendc;
  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;
    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;
```

The output of the data step is shown below.

Table 2.b.2.a. Example of Number NDC and Number Fills Variable Listing

Obs #	Bene_id	Prod_srvc_id	PDE_ID	numNDC	Numfills	L
1	aaaaaaaaaaaaaaaa	12345678910	Xxxxxxxxxxxxxx1	1	1	0
2	aaaaaaaaaaaaaaaa	12345678910	Xxxxxxxxxxxxxx2	1	2	.
3	aaaaaaaaaaaaaaaa	12345678910	Xxxxxxxxxx1xxxx	1	3	.
4	aaaaaaaaaaaaaaaa	22222233333	Xxxx5xxxxxxxxx1	2	1	.
5	aaaaaaaaaaaaaaaa	22222233333	XxxxxxxQxxxxx	2	2	.
6	aaaaaaaaaaaaaaaa	55667788991	xhxxxfxxxxxxxx	3	1	1
7	bbbbbbbbbbbbbbb	44444444444	X1xxxxxxxxxxxxx	1	1	0
8	bbbbbbbbbbbbbbb	44444444444	X1xxxxxxxxxxxxx	1	2	1

Per user number of NDCs can be tabulated.

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

Results are shown in the table below.

Table 2.b.2.b. Number NDC and Per User NDC and fills

Measure	n	2007 random 5%
Average # of fills per NDC – total population	29,121 different NDCs	Mean=1,664.3; median=25; max=384,615
Average # different NDCs (11-digit) per user	1,191,661 beneficiaries	Mean=11.7; median=10; max=177

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. 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 ccw.Drugs07 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 ccw.cohort07

    group by bene_id;
quit;

*recodes for non-users;
data Drugs07;
    set ccw.Drugs07;
    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=Drugs07 n mean median max min sum;
    var pde_count drug_count gross_cost_sum;
    title 'average utilization and cost per bene';
run;

proc means data=Drugs07 n mean median max min sum;
    var pde_count drug_count gross_cost_sum;
    where user='y';
    title 'average utilization and cost per user';
run;

```

**Table 2.b.2.c. Number Drug Entities (GNN; generic drug products)
Per Beneficiary and Per User**

Measure	n	2007 random 5%
<i>Average # different drug products (GNNs) per bene</i>	<i>1,306,431 beneficiaries</i>	<i>Mean=8.3 median=7; max=103</i>
<i>Average # different drug products (GNNs) per user</i>	<i>1,191,661 beneficiaries</i>	<i>Mean=9.2; median=8; max=103</i>

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 included in the PDE file.

c. 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. 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.

First, since the data values contain intelligence but are not ordinal, 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 format statement is used to classify the ordinal benefit phase variable (called “bpord”) 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 ccw.cohort2007 (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;
```

```
proc freq data =phase;
  tables bpord;
  format bpord order.;
run;
```

This data file contains information for every benefit phase that each beneficiary experienced.

Table 2.c. Benefit Phases for PDEs

% PDEs in each Benefit Phase n=48,466,056 PDEs	
<i>No phase (phases not applicable)</i>	1.4%
<i>Deductible</i>	5.4%
<i>Pre-ICL</i>	63.9%
<i>ICL</i>	20.8%
<i>Catastrophic</i>	8.5%

1) Number of Beneficiaries to Reach Each Phase

Researchers will want to keep only the first prescribing event in each benefit phase for each beneficiary in your cohort. Beneficiaries may be counted in multiple phases. Researchers may be as granular as they would like in defining these phases.

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.

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 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;

```

Table 2.c.1. Benefit Phases for Each Beneficiary

% Beneficiaries reaching each Benefit Phase*	
n=1,191,661 beneficiaries (with 2,189,386 bene: phase combos)	
<i>No phase (phases not applicable)</i>	3.7%
<i>Deductible</i>	21.1%
<i>Pre-ICL</i>	51.2%
<i>ICL</i>	18.8%
<i>Catastrophic</i>	5.2%

*Beneficiaries may fall into multiple phase categories during the year

2) 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.

Table 2.c.2. Highest Benefit Phase for each PDE User

Highest (or worst) benefit phase for each beneficiary with PDEs n=1,191,661 beneficiaries	
<i>No phase (phases not applicable)</i>	<i>0.6%</i>
<i>Deductible</i>	<i>5.2%</i>
<i>Pre-ICL</i>	<i>59.7%</i>
<i>ICL</i>	<i>25.1%</i>
<i>Catastrophic</i>	<i>9.5%</i>

3) 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):

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 here.

```
data ICL;
set phase;
  where bp=3;
  by bene_id bp srvc_dt;
  if first.bene_id;
  gaptime=intck('day',mdy(1,1,2007),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 2.c.3. Month to Reach Coverage Gap (ICL)

<i>GAPTIME</i>	<i>Frequency</i>	<i>Percent</i>	<i>Cumulative Frequency</i>	<i>Cumulative Percent</i>
1	3018	0.76	3018	0.76
2	13239	3.33	16257	4.09
3	24157	6.07	40414	10.16

4	34721	8.73	75135	18.89
5	40876	10.28	116011	29.17
6	40214	10.11	156225	39.28
7	45295	11.39	201520	50.67
8	44174	11.11	245694	61.77
9	40658	10.22	286352	72.00
10	41239	10.37	327591	82.36
11	36691	9.23	364282	91.59
12	33452	8.41	397734	100.00

4) Time in a Phase

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

We begin with the input data file created in the code for 2.c.1., above. 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.

```

data ccw.bp2007;
  length bp bppdecnt 3 bp_start_dt bp_end_dt 4 bp_days
  days2phase 3;
  set b_phase;

  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(1,1,2007),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 mmdyy10.;
run;

```

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

```
proc sort data=ccw.bp2007 out=temp;
```

```

    by bp;
run;

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

*Note: N=Number of Beneficiaries reaching each phase;

```

The output of this step is shown below.

Table 2.c.4. Time to reach Coverage Gap (ICL) and Time in Gap

----- Benefit Phase 3=Gap -----

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Maximum
BPPDECNT	Number of PDEs in Phase	397734	24.5697250	19.4045785	263.0
BP_DAYS	Number of Days in Phase	397734	108.1280554	65.2247224	365.0
DAYS2PHASE	Calendar Days to Start of Phase	397734	209.2382195	87.6488261	364.0

- 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.

d. 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 PTPAYAMT is the data field which indicates the amount the beneficiary paid for the prescription drug. The TOTALCST 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 section 2.b. of this paper. The TOTALCST field was summed for each beneficiary. Results are as follows:

Table 2.d. Gross Drug Costs

Measure	<i>n</i>	2007 random 5%
<i>Average gross drug costs per beneficiary</i>	<i>1,306,431 beneficiaries</i>	<i>Mean=\$2377 median=\$1391</i>
<i>Average gross drug costs per user</i>	<i>1,191,661 beneficiaries</i>	<i>Mean=\$2606; median=\$1615</i>

V. 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.