



Chronic Condition Data Warehouse

Your source for national CMS Medicare and Medicaid research data

CCW Technical Guidance: Getting Started with CMS Medicare Administrative Research Files

December 2017

Version 2.5

This page intentionally left blank.

Revision History

Revision Date	Version Number	Description
12/2017	2.5	Modified formatting of SAS code example on page 37.

TABLE OF CONTENTS

Overview	1
Chapter 1: Contents of CCW Medicare Data Files	1
A. Medicare Enrollment File	1
1. Master Beneficiary Summary File.....	2
B. Claim Types	3
1. Medicare Institutional Claims.....	3
2. Medicare Non-Institutional Claims.....	4
3. Medicare Prescription Drug Event Files.....	5
C. Claim Structure	5
1. Structure of Medicare Institutional Claims	6
2. Structure of Medicare Non-Institutional Claims.....	7
3. Structure of Medicare Part D Events	8
D. Receiving CCW Data.....	8
Chapter 2: Methods	9
A. Sample Selection.....	9
1. Length of Enrollment in the Medicare Benefit	9
2. Type of Medicare Coverage.....	10
3. Age or other Demographic Information	11
4. Aggregating Data to Summarize Coverage Variables	11
B. Examining Patterns of Health Care Utilization.....	15
1. Selecting Utilization Data Files	15
2. Extracting Utilization Information from Claims	16
3. Constructing a Longitudinal Data File.....	28
4. Tabulating or Summarizing Utilization	29
C. Describing Utilization	32
1. Identifying Care for a Particular Diagnosis	32
2. Identifying Care for Individuals with a Particular Condition or Diagnosis	34
3. Identifying Particular Procedures	36
4. Payments for Services	38
Chapter 3: Further Assistance with CMS Administrative Claims Data	49
Appendix A: List of Acronyms and Abbreviations	50

LIST OF TABLES

Table 1. Reference Code Files	7
Table 2. Variable Names for Medicare Enrollment Variables with Monthly Values.....	11
Table 3. Variable Names for Variables Related to Months of Eligibility	14
Table 4. Algorithms for Use in Categorizing Medicare Part A Institutional Claims.....	20
Table 5. Algorithms for Use in Categorizing Medicare Institutional Outpatient Claims	22
Table 6. Algorithms for Use in Categorizing Medicare Non-Institutional Claims	24
Table 7. Algorithms for Use in Categorizing Cost Centers/Areas Within Settings	27
Table 8. Diagnosis, Procedure, and Service Codes Used on Medicare Claims.....	32
Table 9. Variable Names for Payment Fields in Medicare Non-Institutional Claims	40
Table 10. Variable Names for Payment Fields in Medicare Institutional Claims	42

LIST OF CODE EXAMPLES

Code Example 1. Define Sample of Medicare Beneficiaries with Fee-For-Service (FFS) Coverage	12
Code Example 2. Demonstrate Partitioning Medicare Part A Institutional Base Data File into Specific Types of Inpatient Settings	20
Code Example 3. Use Outpatient Base and Revenue Center Data Files to Identify Particular Settings of Care.....	22
Code Example 4. Use Part B Carrier and DME Line Files to Identify Claim Settings	25
Code Example 5. Use Part A Inpatient Revenue Center File to Identify Specific Settings of Care within Inpatient Facility	28
Code Example 6. Use Base and Revenue Center Records to Identify Care Received in Specific Settings and Calculate Population-Level Utilization Statistics.....	29
Code Example 7. Calculate Population-Level Utilization Statistics	30
Code Example 8. Identify Claims for Cohort with Particular Condition of Interest	36
Code Example 9. Find Proportion of Medicare FFS Beneficiaries Who Used Acute Inpatient Care During the Year.....	43
Code Example 10. Identify People Who Experience a Particular Event – a Hospital Readmission	45

Overview

The Chronic Conditions Data Warehouse (CCW) Medicare data are extracted from the Centers for Medicare & Medicaid Services (CMS) enrollment files and fee-for service (FFS) administrative claims submitted for payment to CMS. While the files were designed from the ease-of-use perspective, learning the nuances of any new data source can be challenging. Creating the appropriate analytic dataset can be the key to simplifying analysis of the data.

This document provides technical guidance to approaching Medicare enrollment and administrative data complexities and provides a framework to creating customized, analytic data files. The focus of this document is primarily the Medicare enrollment, Medicare Part A, and Medicare Part B claims delivered by CCW. In addition to narrative technical information, examples of SAS[®] code are shared.

Chapter 1: Contents of CCW Medicare Data Files

Medicare is the primary health insurance program for people age 65 or older, people under age 65 with disabilities, and people of all ages with End-Stage Renal Disease (ESRD) or Lou Gehrig's disease (amyotrophic lateral sclerosis; ALS). Nearly all Medicare beneficiaries receive Part A hospital insurance benefits, which helps cover inpatient hospital care, skilled nursing facility stays, home health and hospice care. Most beneficiaries also subscribe to Part B medical insurance benefits, which help to cover physician services, outpatient care, durable medical equipment (DME) and some home health care. Additionally, many beneficiaries elect to purchase Medicare Part D prescription drug coverage (available since 2006). Beneficiaries may elect to receive FFS Medicare or, as an alternative, enroll in Medicare Part C (Medicare Advantage, or MA). These are private plans similar to HMOs which provide Medicare Part A and Part B services. Many plans offer prescription drug coverage options.

The record layouts for all of the CCW data files may be obtained from the [ccwdata.org](https://www.ccwdata.org/web/guest/data-dictionaries) website, under the Data Dictionaries tab (<https://www.ccwdata.org/web/guest/data-dictionaries>). Throughout this document, when we identify a particular data variable by name we will often identify the specific SAS name which will appear in all capitals.

A. Medicare Enrollment File

Data for 100% of Medicare-enrolled beneficiaries is available from the CCW. The CMS Institutional and Non-institutional data files found in the CCW generally represent Medicare FFS claims only (i.e., managed care encounter information is not available). A few exceptions exist, including coverage of Hospice services. The Medicare managed care benefit is explained in greater detail on the Medicare.gov website (<http://www.medicare.gov/sign-up-change-plans/decide-how-to-get-medicare/your-medicare-coverage-choices.html>).

The Part D prescription drug events (PDEs) data represent events from all beneficiaries participating in the Part D program, regardless of whether the beneficiary is enrolled in managed care or a FFS plan. PDEs are available for all prescription fills that are covered as part of the Part D benefit from 2006 forward (the inception of the benefit), whereas claims data are available from 1999 forward.

1. Master Beneficiary Summary File

The Master Beneficiary Summary File is created annually and contains demographic and enrollment data for all beneficiaries who are alive and enrolled in Medicare for any part of the year. Historical files include the CMS Denominator File, the CCW Beneficiary Summary File (BSF), the Beneficiary Annual Summary File (BASF), and the original pre-March 2017 Master Beneficiary Summary File (MBSF) that used the CMS Enrollment Database (EDB) as its source. Currently, the preferred enrollment data file is the MBSF that uses the CMS Common Medicare Environment (CME) Database as its source. The CCW transition from the EDB to the CME database as the source for the MBSF enrollment and demographic information offers a number of advantages that will better meet researchers' evolving data needs. A description of the rationale and impact of this data conversion is contained in a CCW White Paper, [Master Beneficiary Summary File \(MBSF\): Impact of Conversion from EDB to CME](#).

Like the original MBSF, the current MBSF using the CME enrollment data file is an annual person-level summary file that can be used to determine whether a beneficiary has a sufficient surveillance period (i.e., months of enrollment in the Medicare benefit) for inclusion in the analytic file being created. Variables contained in this file include: 1) the number of months of Medicare Part A, B, C, and D coverage; 2) whether the beneficiary died during the year; 3) Part C or D plan type (if applicable); 4) whether the beneficiary received Part D subsidies; and, 5) other beneficiary demographic and geographic information.

This enrollment data file contains a constellation of enrollment and other person-level variables contained in "segments" – which are separate components of the file that investigators may elect to purchase. Four of the segments are particularly relevant to this document. The current MBSF – Part A/B/C/D segment, also known as the Base Beneficiary Summary File consists of variables that identify monthly Medicare Part A, B, C, and D enrollment status and other key demographic and coverage variables. Within this guidance document we also use information from the two conditions file segments; 1) MBSF- CCW Chronic Conditions segment; and 2) MBSF – CMS Other Chronic or Potentially Disabling Conditions segment. Note that there are two additional segments that can be purchased as part of the MBSF, which are not addressed in this introductory document; these include the NDI (National Death Index – for cause of death; available through 2008) and the Cost & Use (annual summary statistics for each beneficiary) segments.

The MBSF is available in the current layout for 1999 forward.

B. Claim Types

In general, all CMS administrative data files contain key variables which can be used to join the CCW files (e.g., the unique CCW-assigned beneficiary identifier [variable called BENE_ID], the claim identifier [CLM_ID], the claim line/record number [CLM_LINE_NUM], etc.). The linkage keys used may vary depending on which files you are attempting to join. For example, some will link different portions of a claim together by using the claim identification number and claim line number (variables called CLM_ID, CLM_LINE_NUM or LINE_NUM). When medical services provided to a beneficiary are the focus, the primary linkage will be at the person level (i.e., the BENE_ID), after aggregation of the claim level files.

1. Medicare Institutional Claims

The claims from institutional providers and/or settings which are covered by the Medicare Part A benefit appear in the Institutional claims file. In addition, claims for institutional-based services covered by the Medicare Part B benefit (e.g., home health, hospital outpatient) appear in the Institutional claims file. Providers use the 837I standard electronic format (previously known as the UB-04 claim form, or more recently as Form CMS 1450) to submit claims to CMS. Part A claims are processed by CMS contractors known as Medicare Administrative Contractors (MAC), formerly known as Fiscal Intermediaries (FI). For each setting, there is a base file and a revenue center file, which are explained in greater detail in section [C. Claim Structure](#) below. More information regarding the contents and processing of these claim forms is available on the CMS website (http://www.cms.gov/ElectronicBillingEDITrans/15_1450.asp and at <http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/837I-FormCMS-1450-ICN006926.pdf>). These files include the following types of claims:

- **Inpatient** – This is the base file (i.e., the core record for the inpatient claim) with final action claims data – that is, a version of the claim record that reflects services billed by the inpatient hospital for facility costs, processed by the MAC, resubmitted and corrected if necessary, and finalized. Sufficient time must pass before the claims files are considered mature and complete or “final action”. CCW published a white paper regarding Medicare claims maturity, available on the CCW website (<https://www.ccwdata.org/web/guest/ccw-medicare-data-white-papers>).

This file includes ICD-9 or ICD-10 diagnosis and procedure codes, Diagnosis Related Group (DRG) information, dates of service, reimbursement amount, hospital provider, and beneficiary demographic information. Each record in this file is at the claim level.

- **Outpatient** – This is the base file with final action claims data for outpatient services submitted by institutional outpatient providers (e.g., hospital outpatient departments, Rural Health Clinics [RHCs], Federally Qualified Health Centers [FQHCs], renal dialysis facilities, outpatient rehabilitation facilities, comprehensive outpatient

rehabilitation facilities, and community mental health centers). This file includes ICD-9 or ICD-10 diagnosis codes, CMS Healthcare Common Procedure Coding System (HCPCS) codes, dates of service, reimbursement amount, outpatient provider number, and revenue center codes. Each record in this file is at the claim level.

- **Skilled Nursing Facility (SNF)** – This is the base file with final action claims data for claims submitted by SNF providers. This file includes ICD-9 or ICD-10 diagnosis codes, dates of service, reimbursement amount, and SNF provider number. Each record in this file is at the claim level.
- **Hospice** – This is the hospice services base file for final action claims submitted by hospice providers. This file includes the level of hospice care received (e.g., routine home care, inpatient respite care), terminal diagnosis (ICD-9 or ICD-10 diagnosis), dates of service, reimbursement amount, and hospice provider number. Note that claims for all beneficiaries receiving hospice services appear in this file – even hospice services provided to beneficiaries enrolled in managed care plans. Each record in this file is at the claim level.
- **Home Health Agency (HH)** – This is the HH services base file for final action claims submitted by HH providers. This file includes the number of visits, type of visit (e.g., skilled-nursing care, home health aides, physical therapy, speech therapy, occupational therapy, and medical social services), diagnosis (ICD-9 or ICD-10 diagnosis), date(s) of visit(s), reimbursement amount, and HH provider number. Each record in this file is at the claim level.

Note that on October 1, 2015 the conversion from the 9th version of the International Classification of Diseases (ICD-9-CM) to version 10 (ICD-10-CM and ICD-10-PCS) occurred. The Health Insurance Portability and Accountability Act (HIPAA) electronic administrative transactions standards, Version 5010, required that all services provided on or after October 1, 2015 utilize ICD-10 codes. Regardless of when a claim was submitted for payment, services that occurred prior to October 1, 2015, could be billed using ICD-9 codes. This requirement applies to all medical claims in the US, and is not limited to Medicare claims. This means that for claims, when the “from” dates of service (on professional and supplier claims) or dates of discharge/through dates (on institutional claims) are on or after October 1, 2015, then ICD-10 codes must be used.

2. Medicare Non-Institutional Claims

The Medicare Non-Institutional claims include services covered by the Part B benefit, and consist largely of professional services and DME. Providers use the 837P electronic claim form (also known as the CMS 1500 claim form) to submit bills for services rendered to CMS. All claims are processed by Medicare Administrative Contractors (MACs; previously known as Carriers). More information regarding the contents and processing of these claim forms is available on the CMS website (<http://www.cms.gov/Medicare/Billing/ElectronicBillingEDITrans/HealthCareClaims.html>) or

http://www.cms.gov/Medicare/Billing/ElectronicBillingEDITrans/16_1500.html). These files include the following types of claims:

- **Carrier** – This is the base or main file containing final action claims data for Non-Institutional providers (e.g., physicians, physician assistants, clinical social workers, nurse practitioners, independent clinical laboratories, ambulance providers, and free-standing ambulatory surgical centers) processed by MACs (note: the old file name was physician/supplier). This file includes ICD-9 or ICD-10 diagnosis codes, HCPCS codes, dates of service, reimbursement amount, and non-institutional provider numbers (e.g., UPIN, PIN, NPI). Each record in this file is at the claim level.
- **Durable Medical Equipment (DME)** – These are the Non-Institutional claims for the Durable Medical Equipment Regional Carrier [DMERC] base file containing final action claims submitted by DME suppliers and providers. This file includes ICD-9 or ICD-10 diagnosis codes, HCPCS codes, dates of service, reimbursement amount, and DME provider number (i.e., supplier number, NPI). Each record in this file is at the claim level.

Note that on October 1, 2015 the conversion from the 9th version of the International Classification of Diseases (ICD-9-CM) to version 10 (ICD-10-CM and ICD-10-PCS) occurred. A field in the Carrier and DME data file layout indicates the diagnosis code version that was used on the claim (e.g., variable called IDC_DGNS_VRSN_CD1).

3. Medicare Prescription Drug Event Files

For Part D prescription drug events, one record represents a unique event, which is a single prescription drug fill. Prescription drug fills do not consist of a uniform quantity of medication; rather, the days supply of the medication may be quite variable. For example, fills for the same medication may be for a one week supply or a three month supply. The file includes specific information regarding the drug (i.e., the national drug code, or NDC) and the quantity dispensed. More information regarding the contents and processing of these services is available on the CMS website (<http://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PartDDData.html>). In addition, a technical guidance paper entitled *CCW Technical Guidance: Summarizing and Describing Prescription Drug Utilization* is available on the CCW website (<https://www.ccwdata.org/web/guest/technical-guidance-documentation>).

Note that some drugs are covered under the Medicare Part B benefit, and appear in the Non-Institutional claims file rather than the Part D event file. The Part B covered drugs are generally injectable or infused drugs administered in a medical setting (e.g., chemotherapy and some vaccines).

C. Claim Structure

The claims data files vary in structure, depending on the claim type. See [ccwdata.org](https://www.ccwdata.org) for file record layouts and definitions (<https://www.ccwdata.org/web/guest/data-dictionaries>). One claim type may consist of multiple related files.

1. Structure of Medicare Institutional Claims

For Institutional claims, the **base file** contains the base claim record, while the **revenue center file** contains potentially multiple revenue center records for a corresponding base claim record.

- **Base File** - This file contains claim header information such as claim ID, beneficiary ID, claim type, claim from date, claim through date, weekly processing date, provider ID, claim level payment amount, pass through per diem amount, claim total charge amount, admitting diagnosis, primary diagnosis, and up to ten additional diagnosis code fields and six procedure codes with associated dates, as well as the DRG. Note that beginning with claims files from 2009, the claims (beginning with version J of the layout) allow for 25 diagnosis codes and 25 procedure codes.
- **Revenue Center File** - This file contains the line level procedures (HCPCS) for the claim. Revenue center fields include the claim line number, claim type, DRG, HCPCS, Revenue Center code, Revenue Center date, rendering physician ID, revenue center unit count, revenue center charge amount, and revenue center payment amount. Note that the revenue center payments are totaled and reflected on the base file, therefore most data users will not use the revenue center payment information. Reference the “Medicare Fee-For-Service Institutional Claim” file layout on the [ccwdata.org](https://www.ccwdata.org) website (<https://www.ccwdata.org/web/guest/data-dictionaries>) for more detail.
 - You may only need to use the base file, however, if you are interested in more detailed cost and procedure information you will benefit from the revenue center detail. The revenue centers represent institutional cost centers, for which separate charges are billed. For example, there are revenue centers for emergency department (ED), intensive care, physical therapy, laboratory, pharmacy, blood, imaging, etc. It is common to use the revenue center detail if you are interested in the Outpatient file because the revenue center contains important information to help distinguish between care settings (e.g., clinic vs. dialysis care). Additional information regarding how to use the revenue center information to distinguish between types of outpatient services is presented in [Chapter 2.B.](#), below. The base and revenue center file records for a claim are linked using the claim ID (CLM_ID), and the revenue center lines are numbered sequentially using the claim line number (CLM_LINE_NUM).
 - Additionally, four types of reference code files are available to link to the base claim, including condition, occurrence, span, and value code files. These files are rarely used; however, they contain information regarding special

conditions which may affect payer processing. For example, for inpatient claims, the operating Indirect Medical Education (IME) Amount and the operating Disproportionate Share (DSH) Amount fields are derived using the Value codes reference file. These are important to understand if you are examining geographic variation in payment amounts or variation by facility type. See Table 1 for more detail on Reference Code files.

- Starting in 2010, there is also a Demonstration/Innovation Code File that is disseminated with each claim type, to link with the base claims. This file is similar to the Reference Code Files, in that it provides information regarding any innovation or demonstration projects related to the claim (e.g., the type of project).

Table 1. Reference Code Files

File Name	Description
Condition codes	Codes that indicate a condition relating to an Institutional claim (e.g., insurance related, special condition, student status, accommodation, CHAMPUS, SNF, etc.). Information in the Condition Code File may help identify outlier payment situations (e.g., disproportionate share).
Occurrence codes	Codes that identify a significant event/date relating to an Institutional claim (e.g., accident, medical condition, insurance related, service related, etc.). The corresponding date of the occurrence is listed.
Span codes	Codes that relate to an institution (e.g., exhausted all full/coinsurance days but covered on cost report, hospital prior stay dates, visits occurring in this billing period if different, etc.). The from and through dates during which the situation indicated in the span code are given.
Value codes	Codes indicating value of a monetary condition used by the intermediary to process an Institutional claim (e.g., the wage index to be applied to home health care due to the beneficiary location). The amount indicated by the value code is given.

The CCW website has additional information about these Reference Code files, including descriptions for each variable and the valid values. Refer to the website for source file record layouts and definitions (<https://www.ccwdata.org/web/guest/data-dictionaries>).

2. Structure of Medicare Non-Institutional Claims

A claim in the **base file** corresponds to potentially multiple associated line item records in the **line file**.

- **Base file** - This file contains the overall claim level information such as the claim ID, beneficiary ID, claim type, referring physician, claim from date, claim through date, claim processing date, carrier number, claim payment amount, claim allowed charge amount, and up to eight diagnosis codes. Note that beginning with claims files from 2009, the claims (beginning with version J of the layout) allow for 12 diagnosis

codes.

- **Line file** - This file contains the individual line level information from the claim. This includes the HCPCS code(s), diagnosis code(s), first and last expense dates, line allowed charge amount, line submitted charge amount, and performing provider identifier.

It is common to find the line item detail helpful. For example, if looking for procedures (e.g., immunizations, tests, or other treatments) or for all providers who had a role in caring for the patient, the line files are important. The base and line files for a claim are linked using the Claim ID (CLM_ID), and there may be more than one line file for a claim which is indicated by the claim line number (LINE_NUM).

3. Structure of Medicare Part D Events

For Medicare Part D prescription drug events (PDEs), each record represents a unique drug fill. In other words, the drug has been dispensed at the pharmacy. The PDEs are final action, and the event contains all of the information available at the time of the fill. No additional linkage or processing of the PDE data is necessary to be able to use these Part D files.

D. Receiving CCW Data

Data files are shipped to you on either a USB external hard drive or a DVD/CD. These data files are packaged as encrypted self-decrypting archive (SDA) files (see the CCW User Guide at ccwdata.org for additional information on encryption). The decryption password is sent to you electronically via email. When you receive the data package (via hard drive or DVD/CD), copy them from the shipping media to your local workspace. Note some data shipped on a hard drive can be decrypted on that hard drive, depending on the size of the data files. Using the password provided to you via email, follow the Decryption Instructions enclosed in the data package. Each SDA contains the data file(s), SAS® code and a file transfer summary (.fts) file which can be used to verify the data was read in correctly.

Chapter 2: Methods

This chapter describes methods used with the standard record layout (see ccwdata.org) for Medicare data files that are compiled and delivered by CCW to end users. Although the variable names and values are generally consistent over time, you may need to make slight modifications (e.g., with onset of Part D data in 2006 additional data became available).

It is possible that for a given study, several smaller analytical files may be desirable in order to easily manipulate the data for the desired results.

A. Sample Selection

You have likely received the Master Beneficiary Summary File (MBSF) which includes Medicare enrollment and coverage information for all beneficiaries in your study population. There is a single row of data for each beneficiary. It will be important to determine whether you may need to subset this population for your purposes. For example, you may want to limit your data to the subsample with sufficient FFS coverage to allow for adequate surveillance (opportunity to observe FFS claims). Similarly, some treatment patterns may be more accurately assessed only for people of a certain age, or for males or females.

1. Length of Enrollment in the Medicare Benefit

From the universe of beneficiaries included in your data file, you may wish to reduce the population to those who are “at risk” for the events of interest. Not all Medicare beneficiaries are enrolled in FFS Medicare; however, these are the only types of claims that appear in the Medicare data files (note that in a few exceptions, claims for managed care enrollees will appear, for example with hospice claims). The nuances of the managed care benefit are explained in greater detail in a KnowledgeBase article from the Research Data Assistance Center (ResDAC) (<http://www.resdac.org/resconnect/articles/114>).

For Medicare Part D data, however, all covered prescription drug fills are present, regardless of whether the beneficiary is enrolled in FFS (known as a stand-alone Prescription Drug Plan [PDP]) or managed care (known as a Medicare Advantage – Part D Prescription Drug Plan [MA-PD]). You will need to determine if the analysis should only include those persons with FFS coverage, or a certain minimum amount of FFS coverage, during the time frame of interest.

There may be times when it is appropriate to exclude someone from your sample if the subject died during the surveillance period – and other times when it is important to capture death as an outcome. Note that the death variable in the Medicare data files only indicates deaths which occurred the year of the data file; a death occurring after the end of the calendar year of the file is not noted. You should be cautious when requiring a certain length of Medicare coverage, because coverage is terminated after death. For example, if you were studying acute

myocardial infarction (AMI) and wanted to examine cases for 2015, your denominator (beneficiary sample) should be restricted to those who had at least some Part A coverage (for FFS payment of the inpatient stay), but if you require a full 12 months of coverage you will lose your subjects who did not survive the AMI. Best practice would be to include beneficiaries who had Part A coverage for each month of the study year that they were alive.

For more information about applying coverage restrictions, please refer to a Technical Guidance paper called “CCW Technical Guidance: Calculating Medicare Population Statistics” on the CCW website (<https://www.ccwdata.org/web/guest/technical-guidance-documentation>). Sample code for selecting a sample based on Medicare coverage appears later in this section of the paper.

2. Type of Medicare 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 had FFS Medicare A and B coverage. Those enrolled in Medicare Part C have managed care coverage, and the transactional data regarding services received are not included in the claims data files. The MBSF indicates the type of Medicare coverage obtained.

The Medicare state buy-in variable appears 12 times in the MBSF to represent each month of coverage (variable called MDCR_ENTLMT_BUYIN_IND_01 – 12). The values within this variable indicate whether the beneficiary had Medicare Part A and/or B coverage for the month, and whether there was state buy-in (i.e., Medicaid) for the Part B premium. A limitation of this state buy-in variable is that it does not provide information regarding whether the beneficiary was entitled to full or only partial Medicaid benefits. There is also a summary variable in the MBSF which counts the months of state buy-in. Beginning with 2006 data, the state-reported dual status variable became available. The monthly dual status variable (DUAL_STUS_CD_01-12) describes the level of dual benefits (e.g., none, partial, full). A CCW Technical Guidance Paper provides additional details regarding options for identify populations dually enrolled in Medicare and Medicaid (see: <https://www.ccwdata.org/web/guest/technical-guidance-documentation>).

To determine whether the beneficiary had Medicare FFS or managed care coverage, the HMO indicator variable appears 12 times to represent each month of coverage (variable called HMO_IND_01-12). A summary variable that counts the months of managed care coverage (variable called BENE_HMO_CVRAGE_TOT_MONS) is also available in the MBSF.

The MBSF contains information regarding whether the person was dually eligible for both Medicare and Medicaid services. Beginning with 2006 data, the cost-share group variable describing dual eligibility and enrollment for Medicare Part D subsidies became available. The monthly cost share group variable (CST_SHR_GRP_CD_01-12) describes the extent of the Medicaid subsidy in great detail (e.g., some may have subsidized Part B coverage only - or Part D coverage only – or a Part D low-income subsidy).

See Table 2 for the variable names of interest for the monthly Medicare enrollment data.

Table 2. Variable Names for Medicare Enrollment Variables with Monthly Values

Variables with monthly values	SAS [®] Variable name(s) <i>The last 2 digits <MM> are sequential 01-12</i>
Medicare buy-in indicator	MDCR_ENTLMT_BUYIN_IND_<MM>
State reported dual status code	DUAL_STUS_CD_<MM>
HMO indicator	HMO_IND_<MM>
Cost share group	CST_SHR_GRP_CD_<MM>

3. Age or other Demographic Information

The Medicare program is the primary health insurance program for people age 65 or older. The majority of older adults are enrolled in Medicare; therefore the Medicare data are often used for inferences regarding medical care for older adults in the U.S. The program also provides insurance coverage for people under age 65 with certain disabilities, and people of all ages with ESRD. The beneficiary population enrolled in Medicare who is under age 65 differs from the general Medicare population and the general U.S. population in ways that are important to consider when selecting a sample for study.

If the objective is to understand medical care provided to all people with Medicare FFS coverage, it may be appropriate to retain all ages in your sample. However, if your objective is to calculate rates for certain metrics, it may be desirable to limit the sample to specific age groups (e.g., those aged 65 or older either at the end of the calendar year, at time of event, or at the time of death, if it occurred during the year of interest). Note that the pre-calculated age variable in the MBSF (AGE_AT_END_REF_YR) determines age at the end of the calendar year (or if the beneficiary died during the year – age at the time of death).

There are additional considerations when using Medicare data for population statistics, such as person-time of enrollment in FFS Medicare, which are more thoroughly discussed in a separate paper titled “CCW Technical Guidance: Calculating Medicare Population Statistics” available on the CCW website (<https://www.ccwdata.org/web/guest/technical-guidance-documentation>).

4. Aggregating Data to Summarize Coverage Variables

A description of the variables included within the MBSF can be found in the *Type of Medicare Coverage* section on the previous page. Using these variables, it is relatively simple to obtain denominator counts which take into consideration the number of beneficiaries who **ever** had Part A or B coverage, managed care coverage, or dual coverage during the year. Note that, on a rare occasion, these monthly counts may exceed the number of months the beneficiary was alive during the year. That is, if a beneficiary died during the year, occasionally

the months of coverage will exceed the months the beneficiary was alive. If precision is important, it is advisable that you truncate coverage at the month of death (or month prior to month of death). A code example is shared below.

Code Example 1. Define Sample of Medicare Beneficiaries with Fee-For-Service (FFS) Coverage

Define a sample of Medicare FFS beneficiaries with a specified duration of A and B FFS coverage, or coverage until the month of death. The first portion of the code combines the Medicare Part A, B, and D segments of the MBSF. This step is not needed if using the MBSF A/B/C/D file, available May 2017 forward. **These CCW files are already pre-sorted and indexed using BENE_ID, therefore the MBSF segments can be merged without the need for a separate sorting step.** Then, we use the monthly coverage variables to require at least 11 months of FFS coverage; the next portion of code demonstrates how the summary variables, which count the number of covered months, can be used. The input data source is the MBSF ABCD segment from 2015 (note: the variables and variable names are generally the same over time; there are two exceptions: 1) the Part D data, which was new in 2006, and 2) when MBSF switched from using the EDB to using the CME as the source, some of the MBSF fields were given updated SAS names).

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.

The following SAS[®] code assumes that you are developing a new data file (called *coverage*) by using Medicare coverage specifications. Then you may use these variables to subset your population (e.g., FFS only).

```

data coverage (keep=BENE_ID SEX_IDENT_CD RTI_RACE_CD
  BENE_AGE_AT_END_REF_YR BENE_DEATH_DT STATE_CODE
  Member_Mos Cov15 Ffs15 Ptd15);

  merge a.MBSF_AB_SUMMARY a.MBSF_D_CMPNTS; /*Note that if you are
  using the legacy MBSF AB file and the MBSF D segment
  (which were disseminated prior to May 2017), you would need to
  merge the AB with the D enrollment data*/

  by BENE_ID;

  /*Determine Medicare Part A and B Full Fee for Service Coverage
  using monthly variables*/

  array MemberMos_AB (12) MDCR_ENTLMT_BUYIN_IND_01 -
  MDCR_ENTLMT_BUYIN_IND_12;
  array MemberMos_noHMO (12) HMO_IND_01 - HMO_IND_12;
  array Member_FFSMos (12) Member_FFSMos01 - Member_FFSMos12;

  do i= 1 to 12;

```

```

if MemberMos_AB(i) in ('3','C') and
  MemberMos_noHMO(i)in('0','4')then Member_FFSMos(i)=1;
else if MemberMos_AB(i) NOT in('3','C') or
  MemberMos_noHMO(i)NOT in('0','4')then Member_FFSMos(i)=0;
  Member_Mos=sum(of Member_FFSMos:);
end;

if (BENE_DEATH_DT=. and Member_Mos in (11,12)) or
  (BENE_DEATH_DT~=. and month(BENE_DEATH_DT) = Member_Mos and
  Member_mos~=0) then Cov15=1;
else Cov15=0;

/*coverage using summary variables*/
if BENE_HMO_CVRAGE_TOT_MONS=0 and ((year(BENE_DEATH_DT)~=2015 and
BENE_HI_CVRAGE_TOT_MONS=12 and BENE_SMI_CVRAGE_TOT_MONS=12) or
(year(BENE_DEATH_DT)=2015 and
BENE_HI_CVRAGE_TOT_MONS=month(BENE_DEATH_DT) and
BENE_SMI_CVRAGE_TOT_MONS=month(BENE_DEATH_DT))) then
  ffs15='Y';
else ffs15='N';

/*Part D coverage */
if ((year(BENE_DEATH_DT)~=2015 and PTD_PLAN_CVRG_MONS=12 ) or
  (year(BENE_DEATH_DT)=2015 and
PTD_PLAN_CVRG_MONS=month(BENE_DEATH_DT))) then
  ptd15='Y';
else ptd15='N';

label
  Cov15= '11 or 12 months FFS no HMO - except for those who died'
  Member_Mos = 'Total Member months of A B and No HMO - per bene'
  ffs15 = '12 months of Part A&B FFS-no HMO, or until death month'
  ptd15 = '12 months of Part D coverage, or until death month';

run;

/*One option for evaluating FFS is to include all benes and use FFS as
a stratification variable - e.g., using cross-tabulations for each
variable*/

proc freq data= coverage;
  tables (ptd15 RTI_RACE_CD SEX_IDENT_CD)*ffs15;
  title 'Demographic description of full compared to non-full FFS
population';
run;

/*alternatively - you can subset the population to only tabulate
results for FFS benes*/
proc freq data= coverage;
  where ffs15='Y';
  tables      ptd15 RTI_RACE_CD SEX_IDENT_CD;
  title 'Demographic description of full FFS population';

```

run ;

With a few minor changes you can easily modify this SAS[®] code to fit your own denominator specifications (e.g., two-month break in coverage; not requiring Part B coverage; only Part A coverage). As of 2006, information regarding Part D coverage is available.

It may be desirable to require full or partial Part D coverage if you wish to examine Part D event data (e.g., PTD_PLAN_CVRG_MONS='11' or PTD_PLAN_CVRG_MONS='12').

This level of detail may be important for those studying the effect of plan changes during the year, or service use for partial years; however investigators 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 sample based on having coverage for the full year (or until time of death). Analysts may select variables from the MBSF as key study variables of interest, such as whether the beneficiary was dually enrolled in both Medicare and Medicaid. For others, 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).

Options for summarizing beneficiary coverage information are described more thoroughly in a separate paper titled “CCW Technical Guidance: Options for Determining Which CMS Beneficiaries are Dually Eligible for Medicare and Medicaid Benefits” on the CCW website (<https://www.ccwdata.org/web/guest/technical-guidance-documentation>).

The MBSF contains a few variables which summarize the number of months of particular types of coverage for each beneficiary during the year. See Table 3 below for the relevant variable names.

Table 3. Variable Names for Variables Related to Months of Eligibility

Variables which count # months of eligibility*	SAS[®] Variable name(s)
Months of health insurance (Medicare Part A) coverage	BENE_HI_CVRAGE_TOT_MONS
Months of Part B coverage	BENE_SMI_CVRAGE_TOT_MONS
Months of Part D coverage**	PTD_PLAN_CVRG_MONS
Months of state buy-in for Medicaid/Medicare dual eligibles	BENE_STATE_BUYIN_TOT_MONS
Months of managed care coverage	BENE_HMO_CVRAGE_TOT_MONS
Months of Medicaid dual eligibility**	DUAL_ELGBL_MONS

* Depending on your study, you may wish to calculate these numbers yourself. Refer to code example above.

** Months of Part D coverage variable (and other Part D variables) available 2006 forward on the MBSF.

Once you have selected a sample based on the parameters of interest, decide whether you want to keep all of the variables in this file as some may not be of interest for your study. Note: that the data file can become very wide if you are intending to merge multiple years of data, or information regarding utilization from Medicare claims data files.

B. Examining Patterns of Health Care Utilization

After you have selected your sample, the next step is to obtain the subset of claims data that is relevant for your analysis. It is important to understand which claims files to use for particular purposes.

Note: not all Medicare enrolled beneficiaries will have used Medicare-paid services in a particular time frame. Some do not use any medical services at all, while others may use services that are paid by a third party (e.g., the Veteran's Administration). Third party claims do not appear in the Medicare data files.

1. Selecting Utilization Data Files

The files are partitioned into payment types and a variety of different payment rules and processes apply in different care settings. Since each of the claims files contains a large number of variables, we recommend extracting only the required variables and observations for your analysis.

There are some types of services where it may be possible to observe either an Institutional or Non-Institutional claim, or both claim types for a service. In general, the professional component of a service (e.g., the physician or therapist care, etc.) appears in the Non-Institutional file (i.e., Part B Carrier), whereas the facility claim for an associated service, when applicable, appears in the Institutional file (e.g., Part A Inpatient or Skilled Nursing Facility [SNF]).

One of the most common examples of a type of service which may appear in either the Institutional or Non-Institutional claims is outpatient clinic-type services for physician or other provider care. If the objective is to understand the ambulatory care provided to a patient (e.g., a physician/clinic visit for a service), the hospital outpatient file, which includes hospital-based clinics, Rural Health Clinics (RHCs), and Federally Qualified Health Centers (FQHCs), should be examined in addition to the Part B Carrier files.

Another common example of a service which may appear on an Institutional or Non-Institutional claim is a surgical procedure, which can be performed on an outpatient basis. Institutional facilities (e.g., hospitals) submit bills using a Part A (Institutional) claim, and free-standing ambulatory surgery centers (ASCs) submit Part B (Non-Institutional Carrier) claims. The professional component of these services (e.g., surgeon and anesthesiologist) for all settings will appear on the Part B Carrier claims file.

2. Extracting Utilization Information from Claims

You have many options for exploring and describing utilization. Numerous variables in the claims files make it possible to classify the types of services received (i.e., care setting and procedures), and the reason (i.e., diagnosis) for the care. There are also many options for the unit of analysis (e.g., number of visits, total hospitalized days, per capita utilization).

The Medicare claims files are very large. It will be important to determine whether you need to include all claims for your sample, regardless of the reason for the medical care, or whether you may be able to satisfy your analytic objectives by querying the data files and extracting only the claims related to receipt of specific types of care, or care for certain conditions.

Some beneficiaries will not have any utilization (claims) data, whereas other beneficiaries will have a very large volume of services. There are several steps which must be taken to accurately organize utilization:

- 1) Verify you are examining utilization for the appropriate claim type (Institutional, Non-Institutional, or both);
- 2) Determine the care setting(s) you need to examine within the Institutional/Non-Institutional claims; and,
- 3) Determine whether you need to explore all care for your sample or look only for certain types of care (i.e., diagnoses, procedures, or any care within a particular setting/level of care).

The information important to your objectives may be located on the base portion of the claim or in the revenue center file (for Part A – Institutional claims) or the line file (for Part B – Non-Institutional claims). Refer to previous section: [1.C. Claim Structure](#). For some research questions, there may be more than one methodology which could be used to address a question. For example, there are numerous ways that Medicare claims could be classified, and our intention is not to be prescriptive, but rather to demonstrate one method which may be helpful in beginning to examine Medicare administrative data files.

- **Utilization from Institutional Claims** - Institutional claims cover a variety of setting types, some of which are classified as Inpatient, while others are Outpatient.

Often the entire inpatient stay will be included on a single claim; however, this is not always the case, particularly for lengthy stays. To be certain you have captured all of the claims associated with a stay, we suggest the following process:

- Begin by sorting the claims by type (e.g., all acute inpatient claims together, all critical access hospital [CAH] claims together), person (BENE_ID), claim from date (CLM_FROM_DT) and claim through date (CLM_THRU_DT).
- An individual stay may be determined by sorting all claims for a beneficiary within a type of service by descending first service date, last service date, and claim ID. Note that acute inpatient services, which is often the type of care one might wish to query if looking for reasons for inpatient care, may consist of both inpatient acute hospitalizations as well as hospitalizations in a critical access hospital; however,

inpatient care in other settings (e.g., inpatient rehabilitation or long term hospitals) may not be desirable to include for this purpose. Refer to Table 4 (below) for an illustration of how to identify the different inpatient settings.

- A new stay may be identified as a record with a missing discharge date that is not a transfer and is not the same as the previous discharge date in the ordered claims. Note that many beneficiaries are not discharged to home after an inpatient stay – rather they continue to use some sort of institutional or community-based post-acute care. A discharge status code of 30 indicates they are “Still a Patient.”

Some important variables to consider for classification of services for Institutional claims are described below.

- Claim type – This is the variable on the base claim record that describes the type of claim that was processed (NCH_CLM_TYPE_CD). The data dictionary describes the values (e.g., inpatient hospitals are claim types 60 and 61, skilled nursing facilities are claim types 20 and 30, hospices are claim type 50, and home health agencies are claim type 10).
- Facility type (CLM_FAC_TYPE_CD) – This can be used to identify the type of facility that provided care to the beneficiary. Categories include hospital, SNF, HH, Intermediate Care, dialysis facilities, etc.
- Type of service (CLM_SRVC_CLSFCTN_TYPE_CD) – This can be used to specify the type of service provided to the beneficiary and is used in combination with facility type. This code can be used to help identify rural and FQHCs as well as Community mental health centers (CMHC).
- Bill type – You can derive this variable to classify the type of services that appear on the claim. Bill type is a combination of two variables on the original claim, one describing the facility type (CLM_FAC_TYPE_CD) and a second describing the type of service (CLM_SRVC_CLSFCTN_TYPE_CD). For example, when the type of facility is a hospital (CLM_FAC_TYPE_CD =1) and the type of service is outpatient (CLM_SRVC_CLSFCTN_TYPE_CD =3), then the bill type =13 (hospital outpatient department). An example of how to use the bill type variable appears in Table 5, and sample SAS code for creating this field is in Code Example 3.
- Provider Number (PRVDR_NUM) – The first two digits of a provider number can be used to identify the provider state. The third and fourth digits identify the type of facility. This can be used in combination with a specific claim type to identify payment settings.
- RHCs and FQHCs – These facilities can be identified using the claim facility type code (CLM_FAC_TYPE_CD) and claim service classification code (CLM_SRVC_CLSFCTN_TYPE_CD) in Institutional claims (Outpatient file). Another method for identifying RHCs and FQHCs is to use the last 4 digits of the Provider Number (PRVDR_NUM).
- Revenue Center (REV_CNTR) – This identifies the cost center within a hospital or facility where the service was provided. This can be used to identify services delivered in the ED, intensive care unit (ICU), and intermediate care unit (IMC). Note

that this information can be obtained from the revenue center file, rather than the Institutional base claim file.

Some care types which appear in Institutional claims and could be billed in more than one way are presented below:

- **Emergency Department Care Utilization** - The Revenue Center Files associated with an Institutional claim must be examined to determine whether ED services were received. There are two methods for billing for these services, depending on whether or not the beneficiary was admitted to the hospital. The Inpatient revenue center should be used to identify ED services where the beneficiary was subsequently admitted to the hospital, otherwise the Outpatient revenue center will identify beneficiaries who were discharged from the ED. Often a single ED visit will have bills for both the facility and the provider caring for the patient.
 - Hospital outpatient and Acute Inpatient ED claims are identified by revenue center codes equal to '0450', '0451', '0452', '0456', '0459'.
 - In addition to the ED (facility) codes listed previously, there may be a bill for the professional component of the ED service. The professional component of the ED care can be identified by revenue center code 0981. There may also be a professional component claim for an ED visit, which may be identified in the Part B Carrier claims as Berenson-Eggers Type of Service (BETOS) code 'M3'. BETOS codes are derived from the individual HCPCS code on the claim (see the following section on "Utilization from Non-Institutional Claims" for more details regarding BETOS codes).

A SAS[®] code example for how to identify ED claims appears below in Code Example 6.

- **Utilization from Non-Institutional Claims** – Non-Institutional claims cover a variety of settings including physician office, laboratory, imaging, procedures, ambulatory surgery, and others. Some important variables to consider for classification of services as ambulatory or Physician Office claims are described below.
 - Place of service (LINE_PLACE_OF_SRVC_CD) – This is the Part B (Non-Institutional claim) line item code (i.e., on the line file rather than the Part B base file) indicating the location where the service was provided.
 - Type of service (LINE_CMS_TYPE_SRVC_CD) – This is the Part B line item code indicating the type of service (e.g., medical care, surgery, ambulatory surgical center, etc.).
 - Berenson-Eggers Type of Service (BETOS) codes (variable called BETOS_CD) – This is a reference data set using the CMS HCPCS procedure codes to classify the types of care received. BETOS codes are clinically understood categories that can be used for analysis of patient care. A cross-walk of HCPCS code to BETOS code can be downloaded from the CMS website (http://www.cms.gov/HCPCSReleaseCodeSets/20_BETOS.asp). These codes allow

- data users to distinguish between care provided by physicians and other types of services.
- For example: Claims with BETOS codes of D1A, D1B, D1C, D1D, D1E, and D1F are identified as DME claims. BETOS code D1G indicates DME drug and O1C is parenteral nutrition utilization.
- **Utilization from both Institutional and Non-Institutional Claims** –There are some types of medical utilization which may appear in both Institutional and Non-Institutional claims files. Two scenarios are illustrated below.
 - **Outpatient Physician Care** – Visits with health care professionals in the ambulatory setting may take place in an outpatient facility or a provider office setting.
 - Outpatient care is identified from the Institutional Outpatient claims files. Revenue Center Codes (found on the Revenue Center File) are used to identify the nature of care (e.g., RHC, FQHC). A revenue center code table can be found on the ResDAC website <http://www.resdac.org/cms-data/variables/revenue-center-code>). Alternatively, you may identify the CPT codes used to bill for the types of visits of interest (e.g., 99201-99205, 99211-99215 are used for office care for new or established patients). BETOS codes, described below, may also be referenced to obtain the HCPCS codes of interest.
 - Physician Office Care is a small portion of the claims found in the Part B Carrier files. If you are interested in examining particular types of services, such as face-to-face visits with providers (also referred to as evaluation and management services), you must employ different procedures for two different claim types (i.e., the hospital outpatient/clinic and the provider office). Both the Outpatient Revenue Center file and the Part B Carrier claims file should be used. Some classification examples are shown below, in Table 4.
 - For the Part B Carrier claims, identify the relevant BETOS codes, which are cross-walked to the appropriate HCPCS codes. The BETOS_CD data field is located in the Carrier Line item files (not the base files). BETOS codes can be downloaded from the CMS website (<http://www.cms.gov/Medicare/Coding/HCPCSReleaseCodeSets/BETOS.html>).
 - **Some drugs** are covered under the Medicare Part B benefit - such as intravenous chemotherapy, other infused drugs, and some vaccines. These services may be found using the HCPCS on the revenue center file for Institutional claims, or on the line file for Non- Institutional claims (e.g., many chemotherapy drugs are found in HCPCS J9000-J9999; pneumococcal vaccine may use HCPCS 90669, 90670 or 90732, depending on age of patient and the particular vaccine).

Table 4 below displays algorithms which may be helpful in categorizing services from Institutional claims. There are four categorizations for type of facility (refer to leftmost column of Table 4). If you would like to be more granular in terms of identifying Institutional care settings, the Setting-Description column of the table divides the Part A claims into 9 settings. Additional algorithms for Institutional Outpatient and Non-Institutional claims are illustrated in Table 5 and Table 6; algorithms to identify particular cost centers within a setting are in Table 7.

Table 4. Algorithms for Use in Categorizing Medicare Part A Institutional Claims

Type of facility	Claim Type (NCH_CLM_TYPE_CD)	Setting- Description	Provider Number (PRVDR_NUM)
Inpatient acute care hospital	60,61	Inpatient Prospective Payment System reimbursed hospital (ACUTE)	3rd digit =0
		Critical access hospital (CAH)	3rd & 4th digits = 13
Other inpatient hospital	60,61	Inpatient psychiatric facility (IPF)	3rd & 4th digits = 40,41,42,43,44 or 3rd digit = M or S
		Other hospital type (e.g., children’s hospital or cancer center) (OTH-IP)	All other from claim type 60, 61
Post-acute care	60,61	Long-term care hospital (LTCH)	3rd & 4th digits = 20,21,22
	60,61	Inpatient rehabilitation facility (IRF)	4 digits = 3025 – 3099 or 3rd digit = R or T
	20,30	Skilled nursing facility (SNF)	
	10	Home health (HH)	
Hospice	50	Hospice (HOS)	

The following SAS® code examples illustrate how to identify the claim records for the care settings of interest. For code example 2 we use only claim types 60 & 61 (which are contained in the CCW inpatient base claims file); the example does not apply to SNF, HH or HOS claims.

Code Example 2. Demonstrate Partitioning Medicare Part A Institutional Base Data File into Specific Types of Inpatient Settings

The following SAS® code is designed to subset the claims in the Part A Inpatient Base claims file into smaller files which are specific to the different hospital care settings. For this DATA step we end up with six inpatient setting-specific files: 1) acute, 2) critical access hospital, 3) inpatient psychiatric hospital, 4) long term care hospital, 5) inpatient rehabilitation hospital, and 6) other inpatient hospital.

```

data acute cah ipf ltch irf oth_ip;
  set a.INPATIENT_BASE_CLAIMS_J /*Version K of the RIF claims
  layout became available in 2017, if using, update the file name
  in this example*/ (keep=CLM_ID BENE_ID NCH_CLM_TYPE_CD
    CLM_FROM_DT CLM_THRU_DT NCH_BENE_DSCHRG_DT
  PTNT_DSCHRG_STUS_CD PRVDR_NUM ORG_NPI_NUM AT_PHYSN_UPIN
  AT_PHYSN_NPI OP_PHYSN_UPIN OP_PHYSN_NPI OT_PHYSN_UPIN
  OT_PHYSN_NPI CLM_UTLZTN_DAY_CNT CLM_NON_UTLZTN_DAYS_CNT
  CLM_FAC_TYPE_CD CLM_SRVC_CLSFCTN_TYPE_CD
  CLM_IP_ADMSN_TYPE_CD CLM_SRC_IP_ADMSN_CD CLM_DRG_CD
  CLM_DRG_OUTLIER_STAY_CD NCH_DRG_OUTLIER_APRVD_PMT_AMT
  NCH_BENE_IP_DDCTBL_AMT NCH_BENE_PTA_COINSRNC_LBLTY_AM
  CLM_PASS_THRU_PER_DIEM_AMT CLM_PMT_AMT
  ICD_DGNS_CD1 ICD_DGNS_CD2 ICD_DGNS_CD3 ICD_DGNS_CD4 ICD_DGNS_CD5
  ICD_DGNS_CD6 ICD_DGNS_CD7 ICD_DGNS_CD8 ICD_DGNS_CD9 ICD_DGNS_CD10
  ICD_PRCDR_CD1 ICD_PRCDR_CD2 ICD_PRCDR_CD3 ICD_PRCDR_CD4
  ICD_PRCDR_CD5 ICD_PRCDR_CD6);

  if substr(PRVDR_NUM,3,1) in ('0') then
    output acute;
  else if substr(PRVDR_NUM,3,2) in ('13') then
    output cah;
  else if substr(PRVDR_NUM,3,2) in ('40','41','42','43','44') or
    substr(PRVDR_NUM,3,1) in ('M','S') then
    output ipf;
  else if substr(PRVDR_NUM,3,2) in ('20','21','22') then
    output ltch;
  else if substr(PRVDR_NUM,3,2) in ('30') or substr(PRVDR_NUM,3,1)
    in ('R','T') then
    output irf;
  else output oth_ip;
run;

```

Hospital Outpatient (HOP) claims are considered Institutional data files due to the type of claim used with Medicare, even though these claims are generally paid through the Medicare Part B benefit. All HOP claims, also called Institutional Outpatient claims, consist of claim type = 40; this is used by a variety of institutional providers. It is often helpful to categorize HOP claims according to the type of provider and/or specific settings of care. To accomplish this objective, we create a 2-digit variable called bill type using a combination of two variables on the original claim, one describing the facility type (CLM_FAC_TYPE_CD), a second describing the type of service (CLM_SRVC_CLSFCTN_TYPE_CD). Table 5 below displays algorithms which may be helpful in categorizing services from Institutional Outpatient claims. A coding example follows the table.

Table 5. Algorithms for Use in Categorizing Medicare Institutional Outpatient Claims

Facility Type	Bill Type	Setting- Description*	Bill Type
Hospital outpatient department	85 or 13	Hospital reimbursed under outpatient Prospective Payment System (OPPS)	13
		Critical Access Hospital (CAH – OPD)	85
Outpatient dialysis facility	72	Outpatient dialysis facility (BT 72) (ESRD)	72
Clinic	71,73,77	Federally-qualified health center (FQHC)	73 or 77
		Rural health clinic (RHC)	71
Other Part B institutional services	All other BILL_TYPE values	Part B-covered services for inpatients that have exhausted Part A coverage (NPTA)	12
		Skilled nursing facility (SNF)	22 or 23
		Therapy (THRPY)	74 or 75
		Community mental health center (CMHC)	76
		Other Part B home health services (largely osteoporosis drugs) (BT34)	34
		Laboratory services for non-patients (BT14)	14
		Certain outpatient services in Maryland hospitals (BT83)	83

* Proposed values for these settings appear in parentheses; these are used in the code examples.

The following SAS[®] code examples illustrate how to identify the claim records for the care settings of interest within the HOP setting. For Code Example 3 we use both the outpatient base and outpatient revenue center files. **Note that the CCW claims files are pre-sorted and indexed by BENE_ID and CLM_ID so that they may be easily merged without the need for a separate sorting step.** We wish to use the revenue center record detail in future analyses (which allows us to examine utilization and associated payments more specifically); therefore our output data sets are at the revenue center-level rather than the claim-level.

Code Example 3. Use Outpatient Base and Revenue Center Data Files to Identify Particular Settings of Care

The following SAS[®] code is designed to associate the claims in Outpatient data files with particular care settings. For this DATA step we end up with 11 HOP setting-specific files.

```
data opps cah_opd fqhc rhc npta snf thrpy cmhc bt34 bt14 bt83;
```

```

merge a.OUTPATIENT_REVENUE_CENTER_J /*Version K of the RIF claims
layout became available in 2017, if using, update the file name
in this example*/ (keep=CLM_ID BENE_ID CLM_LINE_NUM REV_CNTR_DT
REV_CNTR REV_CNTR_UNIT_CNT REV_CNTR_APC_HIPPS_CD
REV_CNTR_PMT_AMT_AMT REV_CNTR_PRVDR_PMT_AMT
REV_CNTR_PTNT_RSPNSBLTY_PMT RNDRNG_PHYSN_NPI  RNDRNG_PHYSN_UPIN)

a.OUTPATIENT_BASE_CLAIMS_J /*update for Version K, as needed*/
(keep=CLM_ID BENE_ID CLM_FAC_TYPE_CD CLM_SRVC_CLSFCTN_TYPE_CD);

by BENE_ID CLM_ID;

/* create the bill type variable to identify settings of care*/

bill_type= CLM_FAC_TYPE_CD ||CLM_SRVC_CLSFCTN_TYPE_CD;

/*partition revenue center records into service settings*/

if bill_type in ('85','13') then
  output OPPS;
else if bill_type in ('72') then
  output cah_OPD;
else if bill_type in ('73','77') then
  output FQHC;
else if bill_type in ('71') then
  output RHC;
else if bill_type in ('12') then
  output NPTA;
else if bill_type in ('22','23') then
  output SNF;
else if bill_type in ('74','75') then
  output THRPY;
else if bill_type in ('76') then
  output CMHC;
else if bill_type in ('34') then
  output BT34;
else if bill_type in ('14') then
  output BT14;
else if bill_type in ('83') then
  output BT83;
run;

```

Part B non-institutional claims, which are either so-called “carrier” claims (claim types 71 and 72) or durable medical equipment (DME) claims (types 81 and 82). The individual claim lines, rather than the base claim, have the detailed information on the nature of services provided and the corresponding payment. For Part B non-institutional services, we based our service categories largely on the BETOS classification scheme that groups HCPCS codes into distinct categories. An algorithm for identifying types of service appears in Table 6; a code example follows. Note that

for the non-institutional claims, we identify ASC services first, since the BETOS codes overlap with other non-ASC categories.

Table 6. Algorithms for Use in Categorizing Medicare Non-Institutional Claims

Type of service	Description*	Values (BETOS codes; BETOS_CD)
ASC (Ambulatory surgical center)†	ASC procedures (PRCASC)	P1A – P9B
	ASC durable medical equipment (DMEASC)	D1A, D1B, D1C, D1D, D1E, D1F
	ASC other service (OTHASC)	All other ASC
DRUG (Part B drug)	Parenteral nutrition (PEN)	O1C
	Chemotherapy (CHEMO)	O1D
	Other Part B drug (OTHDG)	O1E
	Immunization (IMM)	O1G
EM (Evaluation & management)	Physician office (OFF)	M1A, M1B
	Hospital (HOSP)	M2A, M2C
	Emergency room (ER)	M3
	Home or nursing home visit (HNH)	M4A, M4B
	Specialist (SPEC)	M5A, M5B, M5C, M5D
	Consultation (CONSUL)	M6
PROC (Procedure)	Anesthesia (ANES)	P0
	Other – major procedure (MAJOTH)	P1A - P1G
	Major cardiac procedure (MAJCAR)	P2A-P2F
	Major orthopedic procedure (MAJORT)	P3A - P3D
	Eye procedure (EYE)	P4A - P4E
	Ambulatory procedure (AMBLTY)	P5A - P5E
	Minor procedure (MINOR)	P6A - P6D
	Oncology procedure (ONCOL)	P7A, P7B
	Endoscopy (ENDO)	P8A - P8I
Dialysis services (DIALYS)	P9A, P9B	
IMG (Imaging)	Standard imaging (STD)	I1A - I1F
	Advanced imaging (ADV)	I2A - I2D
	Echography (ECHO)	I3A - I3F
	Imaging procedure (IMGPRC)	I4A, I4B
LABTST (Laboratory or test)	Laboratory test (LAB)	T1A - T1H
	Other test (TEST)	T2A - T2D
DME (Durable medical equipment)	DME supplies (DMES)	D1A, D1B, D1C, D1D or D1E
	Orthotic devices (PO)	D1F
	DME drug (DMEDG)	D1G
OTHER (Other Part B non-institutional service)	Ambulance (AMB)	O1A
	Chiropractic (CHIR)	O1B
	Vision, hearing or speech services (VHS)	O1F

	Other/unclassified Part B service (OTH)	Y1, Y2, Z2, and missing
--	---	-------------------------

* Proposed values for these types of service appear in parentheses; these are used in the code examples.

† To classify ASC services we first identify claims where the LINE_CMS_TYPE_SRVC_CD='F'

In a code example below, we illustrate how to partition the Part B line item records into service settings (note: this code does not use the Part B or DME Base claims). We combine information from both the CCW Carrier and DME data files for classifying the service types. Investigators may modify this code to meet their particular research needs. These service types can be further segmented using BETOS Codes, which are derived from HCPCS codes.

Code Example 4. Use Part B Carrier and DME Line Files to Identify Claim Settings

The following SAS® code is designed to identify service types in the Part B Carrier and DME Line files. First, claims from both files are combined. Then, service types are identified; we also illustrate how more granular coding for settings can be employed.

```

data a.PartB;
set a.BCARRIER_LINE_J /*Version K of the RIF claims layout became
    available in 2017, if using, update the file name in this
    example*/ (keep=BENE_ID CLM_ID LINE_1ST_EXPNS_DT
    LINE_LAST_EXPNS_DT ORG_NPI_NUM PRF_PHYSN_NPI PRF_PHYSN_UPIN
    TAX_NUM PRVDR_ZIP LINE_NCH_PMT_AMT LINE_BENE_PMT_AMT
    LINE_COINSRNC_AMT CARR_LINE_MTUS_CNT BETOS_CD HCPCS_CD
    LINE_ICD_DGNS_CD LINE_CMS_TYPE_SRVC_CD)

a.DME_LINE_J /*update for Version K, as needed*/ (keep=BENE_ID CLM_ID
    LINE_1ST_EXPNS_DT LINE_LAST_EXPNS_DT PRVDR_NPI PRVDR_NUM TAX_NUM
    LINE_NCH_PMT_AMT LINE_BENE_PMT_AMT LINE_COINSRNC_AMT
    DMERC_LINE_MTUS_CNT BETOS_CD HCPCS_CD LINE_ICD_DGNS_CD);

length provider_id $12 srvc_1 srvc_2 $6;

/*determine the provider of service - needed to create a new variable
to capture the different variable names and different data options -
NPI or UPIN. For 2009 forward, only NPI is used, not UPIN*/

if PRF_PHYSN_NPI not in ('^','~','') then provider_id=PRF_PHYSN_NPI;
else if PRF_PHYSN_UPIN not in ('^','~','') then
provider_id=PRF_PHYSN_UPIN;

/*this part was from DME -*/
else if PRVDR_NPI not in ('^','~','') then provider_id=PRVDR_NPI;
else if prvdr_num not in ('^','~','') then provider_id=PRVDR_NUM;
else provider_id='';

if CARR_LINE_MTUS_CNT~=. then mtus=CARR_LINE_MTUS_CNT;
else if DMERC_LINE_MTUS_CNT~=. then mtus=DMERC_LINE_MTUS_CNT;

```

```
first_srvc_dt=datepart(LINE_1ST_EXPNS_DT);
last_srvc_dt=datepart(LINE_LAST_EXPNS_DT);

line_bene_pay=LINE_BENE_PMT_AMT + LINE_COINSRNC_AMT;

/*partition line records into service categories using substrings of
the BETOS_CD*/
betos1=substr(BETOS_CD,1,1); betos2=substr(BETOS_CD,1,2);

/*calculate a level 1 service setting for each line item - the srvc_1
variable captures the broad type of service
We identify ASC services first*/

if LINE_CMS_TYPE_SRVC_CD='F' then srvc_1='ASC';
else if BETOS_CD in ('O1C','O1D','O1E','O1G') then      srvc_1='DRUG';
else if betos1='M' then srvc_1='EM';

/*you can add more detailed service settings - we call this a level 2
service setting or srvc_2*/

else if betos1='P' then do;
    srvc_1='PROC';
    if betos2='P0' then srvc_2='ANES';
    else if betos2='P1' then srvc_2='MAJOTH';
    else if betos2='P2' then srvc_2='MAJCAR';
    else if betos2='P3' then srvc_2='MAJORT';
    else if betos2='P4' then srvc_2='EYE';
    else if betos2='P5' then srvc_2='AMBLTY';
    else if betos2='P6' then srvc_2='MINOR';
    else if betos2='P7' then srvc_2='ONCOL';
    else if betos2='P8' then srvc_2='ENDO';
    else if betos2='P9' then srvc_2='DIALYS';
    end;
else if betos1='I' then srvc_1='IMG';
else if betos1='T' then srvc_1='LABTST';
else if betos1='D' then srvc_1='DME';
else if betos1='O' or betos1='Y' or betos1='Z' then srvc_1='OTHER';

format line_1st_expns_dt line_last_expns_dt mmddyy10.;

run;

/* tabulations */
proc freq data= a.PartB;
tables      srvc_1 *srvc_2*betos2/list missing;
title "Part B Carrier and DME Service Events";
run;
```

There are times when it is helpful to understand the types of service received by the beneficiary in terms of particular cost centers or areas of the hospital. The Institutional Base claims do not contain this level of detail; the revenue center data file must be used for this purpose. Revenue center codes identify particular cost centers, such as the emergency department, intensive care unit, operating room, etc. (see the list of revenue center codes: <http://www.resdac.org/cms-data/variables/revenue-center-code>). We share some examples of algorithms that may be helpful in understanding the types of services received by the patient (Table 7, and the code example that follows).

Table 7. Algorithms for Use in Categorizing Cost Centers/Areas Within Settings

Claim Type	Institutional Service setting (or cost)	Revenue center codes*
Inpatient Revenue Center File	Emergency Department	If REV_CNTR in ('0450','0451','0452','0456','0459') then ED (note: some investigators may also wish to include the professional component of the ED care – REV_CNTR = 0981)
	Intensive Care Unit	Else if REV_CNTR in ('0200','0201','0202','0203','0204','0207','0208','0209') then ICU
	Intermediate Care Unit	Else if REV_CNTR in ('0206') then IMC
	Coronary Care Unit	Else if REV_CNTR in ('0210','0211','0212','0213','0214','0219') then CCU
	Acute care - not one of these settings	Else other acute care
SNF Revenue Center File	All non-Swing bed	The base claim record links with a revenue center record with REV_CNTR = '0022'
	Swing Bed	The base claim record does NOT link with a revenue center record with REV_CNTR = '0022'
Outpatient (HOP) Revenue Center File	Outpatient ED Services	REV_CNTR in ('0450','0451','0452','0456','0459') (note: some investigators may also wish to include the professional component of the ED care – REV_CNTR = 0981; professional services will also be found in the Part B claims)

* The base claim records can be linked with the associated revenue center using the (CLM_ID and CLM_LINE_NUM).

For Code Example 5 we use only the Inpatient revenue center file; the example does not apply to all of the settings identified in the table above (e.g., example does not include SNF or HOP revenue center lines). You can modify the code example to read in the appropriate source data to meet your needs.

Code Example 5. Use Part A Inpatient Revenue Center File to Identify Specific Settings of Care within Inpatient Facility

The following SAS® code is designed to subset the claims in the Part A Inpatient revenue center file into smaller files which are specific to care settings. For this DATA step we end up with the following inpatient setting-specific files: 1) emergency department, 2) intensive care, 3) coronary care unit, 4) intermediate care, and 5) all other inpatient revenue centers.

```
data ed icu ccu imc oth_ip_revs;
    set a.INPATIENT_REVENUE_CENTER_J /*update for Version K, as
needed*/;
    rename CLM_THRU_DT=REV_CNTR_DT;

    if REV_CNTR in ('0450','0451','0452','0456','0459') then
        output ed;
    else if REV_CNTR in
('0200','0201','0202','0203','0204','0207','0208','0209') then
        output icu;
    else if REV_CNTR in ('0210','0211','0212','0213','0214','0219')
then
        output ccu;
    else if REV_CNTR in ('0206') then
        output imc;
    else output oth_ip_revs;
run;
```

Once the data are partitioned into settings, the claims files are more compact, and it is simpler to query the relevant files to determine the medical services that have been received by the study population.

3. Constructing a Longitudinal Data File

Before attempting to merge the utilization file with the denominator/sample file, you will need to sort both data files by BENE_ID. Carefully consider the appropriate merging method to use, depending on whether you want to include all beneficiaries, regardless of service use – or if you only want to keep beneficiaries for whom there is utilization information. Note that some beneficiaries will have many claims (i.e., this will require a many-to one merge with claims to beneficiary information), therefore to improve efficiency it is sometimes advisable to defer joining person-level information (e.g., the denominator) with the claims information until the files have been processed and summarized as desired. For some analytic tasks, the utilization file will be extremely large, and you 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.

An alternative method of managing a large data file is to partition the utilization data file into smaller files (e.g., files based on settings – as we have illustrated; types of services or procedures; monthly files based on service date) and then sort each file and loop through all

the small files, extracting only the utilization records of interest. There may also be times when it is appropriate to restrict your data file only to Medicare service users to more efficiently examine patterns of utilization.

Multiple years of claims data may be required in order to determine whether an event occurred or not. It will be especially important to be efficient with regard to the number of variables to retain.

4. Tabulating or Summarizing Utilization

You have many options for exploring and describing utilization. Numerous variables in the claims files make it possible to classify the types of services received (i.e., care setting and procedures), and the reason (i.e., diagnosis) for the care. There are also many options for the unit of analysis (e.g., number of visits, total hospitalized days, per capita utilization). Below some common methods for counting and summarizing utilization are presented.

a. Tabulating Utilization of Care

As discussed above, to examine patterns of medical utilization, it is sometimes necessary to combine information from different claim types so that a complete assessment of the care taking place in a particular setting is obtained.

For this example we are seeking to identify all ED care – both the claims that are obtained from inpatient hospitals and care billed as HOP (i.e., the latter occurs when the patient was not also admitted immediately to the acute care facility). Various tabulations are performed.

Code Example 6. Use Base and Revenue Center Records to Identify Care Received in Specific Settings and Calculate Population-Level Utilization Statistics.

The following SAS® code is designed to merge data for ED care, which may appear in two claim types and care settings. The subset of Part A Inpatient revenue centers for ED services (using the file created in the DATA step for code example 5, called ed) are merged with the subset of Part A Outpatient revenue centers for ED services (created as the first step in this coding example, called op_ed). The input files are sorted by BENE_ID and date of service.

```
data op_ed;
  set a.OUTPATIENT_REVENUE_CENTER_J /*update for Version K, as
  needed*/ (keep=CLM_ID BENE_ID CLM_LINE_NUM REV_CNTR_DT REV_CNTR
  REV_CNTR_UNIT_CNT REV_CNTR_PMT_AMT_AMT) ;

  where substr(REV_CNTR,1,3)='045' ;
run;

/*sort the 2 input files which contain ED revenue lines, then merge*/
proc sort data = op_ed;
  by BENE_ID REV_CNTR_DT;
```

```
run;

proc sort data= ed;
  by BENE_ID REV_CNTR_DT;
run;

data all_ed;
  set op_ed (in=o) ed (in=i);

  if o then op_ed=1;

  if i then ip_ed=1;
run;

proc sort data=all_ed;
  by BENE_ID;
run;

/*create a small data file which counts per bene ED use by IP and
Outpatient settings*/

proc means data=all_ed noprint;
  by BENE_ID;
  output out=bene_ed(drop=_freq_ _type_) sum(op_ed ip_ed)=;
run;

proc freq data=all_ed;
  tables op_ed ip_ed;
  title ' number of OP and IP Ed visits';
run;
```

We demonstrate how to calculate per capita use statistics (i.e., per enrolled beneficiary) and other utilization statistics – both in aggregate and for population subsets (e.g., by sex or race).

Code Example 7. Calculate Population-Level Utilization Statistics

The following SAS® code is designed to illustrate several types of population (and subpopulation) utilization statistics, which requires bringing in data from the **coverage** data file (created in code example 1) and the **bene_ed** file created in the previous code example. Then, we show how a macro can be used to calculate the total number of beneficiaries who used ED services – and to subset these statistic for desired population subgroups.

```
data bene_ed_cov;
  merge coverage bene_ed;
  by BENE_ID;

  if first.BENE_ID then
    bene_cnt=1;
```

```
all_ed=sum(op_ed,ip_ed);

if op_ed > 0 then
    op_user=1;

if ip_ed > 0 then
    ip_user=1;

if all_ed > 0 then
    ed_user=1;
run;

/*macro is designed to do various tabulations - first overall, then by
population subgroup - from the coverage file, which the investigator
can specify */
%macro summ_ed (var= );

proc sort data=bene_ed_cov;
    by &var;
run;

proc means data=bene_ed_cov noprint;
    by &var;
    where ffs15='Y';
    output out=&var._summ(drop=_freq_ _type_) sum(op_ed ip_ed
        all_ed op_user ip_user ed_user bene_cnt)=;
run;

data &var._summ;
    set &var._summ;
    op_1000bene=op_ed/bene_cnt * 1000;
    ip_1000bene=ip_ed/bene_cnt * 1000;
    all_1000bene=all_ed/bene_cnt * 1000;
    op_user_pct=op_user/bene_cnt * 100;
    ip_user_pct=ip_user/bene_cnt * 100;
    ed_user_pct=ed_user/bene_cnt * 100;

label
    op_1000bene='OP ED visits per 1000 beneficiaries'
    ip_1000bene='IP ED visits per 1000 beneficiaries'
    all_1000bene='All ED visits per 1000 beneficiaries'
    op_user_pct='Percentage of beneficiaries with an OP ED
visit'
    ip_user_pct='Percentage of beneficiaries with an IP ED visit'
    ed_user_pct='Percentage of beneficiaries with either an OP or IP
ED visit';
run;

proc print data=&var._summ;
    title "ED Visits";
run;
```

```
%mend summ_ed;

/*summary statistics overall, then for different demographic
variables*/

%summ_ed (var=RTI_RACE_CD);
%summ_ed (var= SEX_IDENT_CD);
```

C. Describing Utilization

1. Identifying Care for a Particular Diagnosis

Health care claims vary in the kinds of information they require providers to submit about a patient’s diagnoses and/or procedures obtained during a health care encounter. Table 8 below summarizes the types of information available in the CCW files:

Table 8. Diagnosis, Procedure, and Service Codes Used on Medicare Claims

Type of codes	Part A	Part B Institutional	Part B Non-inst.
ICD-9 or 10 diagnosis code	X	X	X
Diagnosis-related group (DRG)	X		
Revenue center code	X	X	
ICD-9 or 10 procedure code	X		
Current Procedural Terminology (CPT) code	X	X	
Healthcare Common Procedure Coding System (HCPCS) code		X	X
Berenson-Eggers type of service (BETOS) code			X
Ambulatory payment classification (APC) code		X	

Medicare claims use the ninth version of the International Classification of Diseases (ICD-9; or after October 1, 2015 ICD-10) to classify all diagnoses, which identify the condition(s) for which a patient is receiving care. Claims data generally allow providers to specify numerous diagnosis codes (up to 25 codes for Part A claims and up to 12 codes for Part B claims beginning with claims files from 2009, when the version J data file layout was implemented; previously there were only up to 10 diagnosis codes), with one diagnosis identified on the claim as the principal or primary diagnosis. The diagnosis codes appear on the base claims.

Medicare assigns hospital discharges to diagnosis-related groups (DRGs), a classification system that groups similar clinical conditions and procedures. The beneficiary's principal diagnosis and secondary diagnoses, as well as any procedures furnished during the stay, are used to determine the appropriate DRG. CMS reviews the DRG definitions annually. The agency switched to a modified system, called Medicare Severity Diagnosis Related Groups (MS-DRGs) on October 1, 2007. Any claims that CMS received on or after that date are classified using MS-DRGs. Both DRGs and MS-DRGs appear in the same data field (CLM_DRG_CD) in the base portion of the claims.

Medicare pays for inpatient hospital care using case-mix groups called DRGs. Medicare uses other forms of case-mix groups to pay for skilled nursing facility care (resource utilization groups, or RUGs) and home health (home health resource groups, or HHRGs). The RUG for SNF claims appear in the HCPCS_CD field (when the REV_CNTR is 0022, then the first 3 digits of the HCPCS_CD are the RUG). The HHRG for a particular revenue center is located in the data field called the ambulatory payment classification (APC) or Health Insurance Prospective Payment System (HIPPS) code (claims variable called REV_CNTR_APC_HIPPS_CD in the revenue center file).

Revenue centers are distinct cost centers within an institution provider that can each submit separate charges. For example, most hospitals have distinct revenue centers for the emergency department, intensive care unit, physical therapy, laboratory, pharmacy, imaging, and so on. Revenue center codes are helpful for identifying different areas of the facility where the patient received care – and also other types of care which may affect payment (such as blood transfusions or laboratory tests). For the Hospital Outpatient setting, More than 90% of the records have a CMS Healthcare Common Procedure Coding System (HCPCS) code (variable called HCPCS_CD, which appears in the revenue center file); in general, rural health centers (RHCs) and Federally-qualified health centers (FQHCs) are less likely to have HCPCS codes than other records (54.6% and 43.7% of revenue center lines, respectively).

For Part A inpatient hospital claims, providers use ICD-9 (or after October 1, 2015 ICD-10) procedure codes to describe the specific procedures that they performed; these are on the base claim. For Part A claims that do not involve inpatient care and for Part B claims, providers use HCPCS codes to describe the services rendered; these codes appear in the revenue or line item files (not the base files).

A HCPCS_CD typically does not appear for revenue center records for hospitalization claims; however the HCPCS_CD is nearly always populated for home health and hospice claims (e.g., slightly more than 90% of the time; compared to less than 1% of the time for inpatient claims).

There are two levels of HCPCS codes. The first level are codes from version 4 of the Current Procedural Terminology (CPT-4), which is a numeric coding system maintained by the American Medical Association (AMA). The CPT consists of descriptive terms and identifying codes that

are used primarily to identify medical services and procedures furnished by physicians and other health care professionals. The AMA makes decisions about adding, deleting, or revising CPT codes. The second level consists of codes for procedures that are not included in the CPT codes; these are primarily used for non-physician services, such as ambulance services or durable medical equipment.

The Berenson-Eggers type of service (BETOS) classification scheme maps HCPCS codes into seven major categories (physician evaluation and management, procedures, imaging, tests, durable medical equipment, other services, and exceptions/unclassified services), with additional sub-categories within each. You can find a complete list of BETOS codes on the ResDAC website (<http://www.resdac.org/cms-data/variables/Line-NCH-BETOS-Code>). These codes appear in the line item files (not the base claims; variable is called BETOS_CD).

Ambulatory Payment Classifications (APCs) are groupings of HCPCS codes that have similar clinical characteristics and costs. Medicare uses APCs to pay for many outpatient hospital services, making a single payment for all codes within an APC instead of paying for each one separately. These codes appear in the revenue center files (variable is called REV_CNTR_APC_HIPPS_CD). For more information about APCs, see:

<http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/HospitalOutpaysysfctsh.pdf>.

As an investigator, you may be interested only in certain types of care. For example, perhaps you only want to examine hospitalizations related to hip fracture or heart failure. Services related to these conditions can be identified using the DRGs. Since the DRGs group conditions for the purposes of payment, the categories may be too broad for some purposes.

If you wanted to more specifically identify services for a particular cause or diagnosis, the most common way to examine reasons for care is to look at the primary (or principal) diagnosis on the claim (PRNCPAL_DGNS_CD, which is equivalent to ICD_DGNS_CD1). This is the medical or surgical reason that hospitalization or skilled nursing care was required. The diagnosis code is a more granular system than DRG (i.e., many diagnosis codes may fall into a single DRG).

Additional diagnoses are present if there are complicating factors (e.g., may not be the major reason for the stay/visit, but have a bearing on the condition of the patient). These are called secondary diagnosis codes (i.e., ICD_DGNS_CD2 - ICD_DGNS_CD25). It is common to see a condition such as diabetes listed as a secondary diagnosis – a co-occurring condition which must be managed, rather than the major cause of the hospitalization.

2. Identifying Care for Individuals with a Particular Condition or Diagnosis

Your analysis may seek to determine whether the medical care received indicated that a medical condition was present, or whether some routine types of care were provided. To

reduce the risk of false positives (i.e., erroneously stating that the patient was receiving treatment for a particular condition, if in fact the condition was not present), there may be times when it is prudent to require the presence of more than one claim to make this determination. In the CCW, for example, varieties of common and chronic conditions in the Medicare population are pre-coded and appear in the MBSF – either the Chronic Conditions segment (historically referred to as the Chronic Conditions Summary File) or the CMS Other Chronic or Potentially Disabling Conditions segment. The twenty-seven (27) predefined CCW common and chronic condition categories were defined by CMS, in collaboration with the Research Data Assistance Center (ResDAC) and Buccaneer¹. Much of the supporting documentation is available on the CCW website (e.g., chronic condition definitions, data dictionary, etc.). See www.ccwdata.org. CMS developed 34 additional measures for Other Chronic or Potentially Disabling Conditions specifically to enhance research of the Medicare-Medicaid dually enrolled population. These additional conditions include variables for mental health and tobacco use conditions, developmental disorder and disability-related conditions, and other chronic physical and behavioral health conditions.

The algorithms for these conditions are very precise regarding the number of claims, the specific types of services, and the number of years of data which must be examined to make a determination regarding whether a person was likely receiving care for a particular condition. Note that the use of this information assumes that if a claim was processed with a particular diagnosis code, the patient was receiving care for that particular condition. It is important to include in analyses only those diagnosis codes of interest, and exclude (if desired) any screening or preventive care codes that do not represent a definitive diagnosis. Refer to the CCW condition algorithm documentation on the website (<https://www.ccwdata.org/web/guest/condition-categories>). These definitions are not prescriptive; you may use whatever claims-based algorithms are appropriate for your analysis.

Furthermore, please note that administrative claims data will only represent beneficiaries receiving FFS care for a particular condition within the time period of the claims data set used in the analysis. Code Example 1 demonstrates a method for identifying beneficiaries with Medicare FFS coverage.

Patterns of utilization for a group identified as having a condition of interest can also be examined but a cautious approach is recommended. The CCW data files contain all FFS claims; they are not limited to claims for a particular condition. You will want to be careful about attributing health service utilization or the associated costs of care, to a particular condition (e.g., fractures may be related to trauma rather than osteoporosis; a person with diabetes may be hospitalized for a cardiac event rather than the diabetes). It is not unusual for Medicare beneficiaries to have multiple chronic conditions, making attribution difficult (see

¹ 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.

paper describing multiple chronic conditions in the Medicare population:
(<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2748070/>).

3. Identifying Particular Procedures

You may be interested in determining whether certain procedures or interventions were received by the patient (i.e., the institutional revenue center records and non-institutional line items have HCPCS codes). HCPCS codes can be divided into two major categories – the Level 1 (i.e., CPT codes) and Level II HCPCS codes. The CPT codes are proprietary and maintained by the American Medical Association (AMA), whereas the HCPCS Level II codes are maintained by CMS. Documentation regarding HCPCS codes can be found on the CMS website (<http://www.cms.gov/MedHCPCSGenInfo/>) or through a variety of published reference manuals.

Detailed procedure information is important if looking at particular tests or interventions (e.g., cancer screening, immunization, surgical procedures).

Code Example 8. Identify Claims for Cohort with Particular Condition of Interest

The following SAS[®] code is designed to identify a sample of beneficiaries with Heart Failure (variable called CHF; input file = MBSF CCW Conditions segment) who had full/nearly full FFS coverage (using the input file called *coverage*, which was created in Code Example 1) and merging it with the MBSF CCW Conditions file. Note that we did not have to sort the MBSF_CC file since CCW already sorts it by BENE_ID. Then evaluate the procedure codes performed for those beneficiaries during one year, using inpatient claims.

After the initial sample is identified and a file is created containing the BENE_ID for each beneficiary in the sample, inpatient claims are merged to this file for analysis of procedure codes.

```
proc sort data= b.coverage;
    by BENE_ID;
run;

*Restrict to a Heart Failure sample - make finder file of these
BENE_IDS;

data hf (drop=chf);
    merge b.coverage (in=x) a.MBSF_CC (in=y keep=BENE_ID CHF);
    by BENE_ID;
    if x and y and CHF in (1,3);
run;

/*This SQL code statement can be used to replace the prior data
step and sorts*/

proc sql;
    create table hf as select b.BENE_ID
```

```

from a.MBSF_CC as a, b.coverage as b
where a.BENE_ID=b.BENE_ID and a.CHF in (1,3);
quit;

*Merge claims for benes with HF;

proc sort data=a.inpatient_base_claims_j /*update for
  Version K, as needed*/ (keep=CLM_ID BENE_ID)
  out=hfacclms_all nodupkey;
  by BENE_ID CLM_ID;
run;

data hf_aclms;
  merge hf (in=s) hfacclms_all (in=t);
  by BENE_ID;
  if s and t;
run;

proc sort data=hf_aclms out=hfacclms nodupkey;
  by CLM_ID;
run;

proc sort data=a.INPATIENT_BASE_CLAIMS_J /*update for
  Version K, as needed*/ out=inpatient_base_claims_j;
  by CLM_ID;
run;

data hfclms;
  merge hfacclms (keep=CLM_ID in=d)
    inpatient_base_claims_j (in=e);
  by CLM_ID;
  if d and e;
run;

/*create SAS data sets with claim counts for non-missing Procedure
  codes. 25 Procedure Codes are Available beginning with the Version J
  layout*/

%macro p(num=);
proc freq data= hfclms noprint;
  where icd_prcdr_cd&num. ne '';
  tables icd_prcdr_cd&num. / out=prcdr&num.
  (rename=(icd_prcdr_cd&num.=icd_prcdr_cd count=count&num.));
run;

%mend p;
%p(num=1);
%p(num=2);
%p(num=3);
%p(num=4);
%p(num=5);
%p(num=6);
%p(num=7);
%p(num=8);
%p(num=9);

```

```
%p(num=10);
%p(num=11);
%p(num=12);
%p(num=13);
%p(num=14);
%p(num=15);
%p(num=16);
%p(num=17);
%p(num=18);
%p(num=19);
%p(num=20);
%p(num=21);
%p(num=22);
%p(num=23);
%p(num=24);
%p(num=25);

/*combine all Procedure Codes and Counts together;*/

data hf_aggr_prcdr_cds (drop=percent count1-count25);
  merge prcdr1 prcdr2 prcdr3 prcdr4 prcdr5 prcdr6 prcdr7 prcdr8
  prcdr9 prcdr10 prcdr11 prcdr12 prcdr13 prcdr14
  prcdr15 prcdr16 prcdr17 prcdr18 prcdr19 prcdr20
  prcdr21 prcdr22 prcdr23 prcdr24 prcdr25;

  by icd_prcdr_cd;

  count=sum(of count1-count25);
run;

/*create one table with primary counts;*/

data hf_prcdr_cd_cnts;
merge hf_aggr_prcdr_cds prcdr1 (drop=percent);
by icd_prcdr_cd;
run;

proc print data=hf_prcdr_cd_cnts;
title 'frequency of each procedure code';
run;
```

4. Payments for Services

You have many options for exploring and describing payments on the claim. Numerous variables in the claims files make it possible to determine various perspectives on payments. The first is to determine the responsibility for payments: 1) the amount the provider charged, 2) the amount paid by Medicare, 3) the beneficiary cost-sharing amount, and 4) the total amount paid to the provider. Furthermore, the RIF layout was updated in 2017 to add several variables related to CMS payment reform initiatives and demonstrations.

Many of these variables are available on both the base claim and the detailed revenue or line item records for the claims. In general, the base claim total payment is the sum of the revenue center/line record payments². If you are calculating payments for particular types of services or procedures (e.g., the physician evaluation and management portion of the visit rather than total claim amount - which may have also included immunizations and a laboratory component), you may wish to examine only the Part B (Carrier) line item records to obtain the line item payment amounts rather than the aggregated payment amount for the visit (where you would use data from the base claim file). Code examples 3-5 demonstrate how to identify particular services – and whether these are found in the base or revenue/line files.

There are also variables on the Part A institutional claims that indicate components of payment, such as per diem amounts, amounts that are factored-in due to payments for medical education (e.g., indirect medical education; IME), and payments to hospitals that serve a disproportionate share of low-income patients (DSH). Below are some common methods for counting and summarizing payments.

a. Responsibility for Payments

Using the claims data, there are different ways you can examine the payments made for services. For each claim type, these payment variables are present, although the names vary a bit. The four major perspectives are the:

- 1) Provider-submitted charges - Providers include their customary charge for a service when they submit a claim. The charge amount could be much different than Medicare's payment rate. The charge amount does not affect what Medicare pays for a service since the program largely pays for services using prospectively-determined fee schedules where CMS sets the payment rates. For example, inpatient hospitals are paid largely based on diagnosis-related groups (DRGs) – regardless of the number of days of service or the amount of charges the facility submits. The CLM_TOT_CHRG_AMT variable on the Medicare Institutional claim indicates how much the provider charged for a particular claim; on the Non-Institutional claims, the variable is the claim submitted amount (variable called NCH_CARR_CLM_SBMTD_AMT).
- 2) Amount paid by Medicare - This is the total amount that Medicare actually paid for a particular claim, revenue center record, or claim line. Note that for some Part A and B Institutional claims (does not apply to Part B non-institutional), a portion of Medicare's payment amount may be an outlier payment (i.e., where Medicare makes additional payments for cases where the provider incurs unusually high costs. For all Institutional claims, the CLM_PMT_AMT is the actual Medicare paid amount; however for inpatient claims, the per diem amount multiplied by the utilization days, must be added to the CLM_PMT_AMT (refer to Table 10).

² For Inpatient claims, the revenue lines do not contain reliable payment information since they do not have a bearing on payment – nor do they factor in outlier payments, when applicable; therefore payments for these revenue lines do not necessarily sum to the total claim amounts. We recommend using ONLY the payment information on the base claim for this claim type.

- 3) Cost of the service borne by the beneficiary - Medicare requires beneficiaries to pay part of the cost of many services through deductibles, copayments, or coinsurance³. For inpatient Medicare services, the beneficiary may pay an additional deductible for blood products. Additional information regarding services covered by Medicare and beneficiary costs, refer to Medicare.gov (<https://www.medicare.gov/your-medicare-costs/index.html>). A variety of different variables in the files are used to identify these components of payment (refer to Table 9 and Table 10).
- 4) Total amount paid to provider - This is also referred to as the total Medicare allowed amount. It is the sum of Medicare’s payment amount, any beneficiary liability for cost sharing (deductibles, copayments, or coinsurance), and any third-party payment (i.e., other insurance) for a given service.–In the Non-Institutional claims, this amount is called the NCH_CARR_CLM_ALOWD_AMT. However, in Institutional claims, you must manipulate the data fields to determine the total amount paid to the provider (e.g., by comparing the primary payer paid amount field [variable called NCH_PRMRY_PYR_CLM_PD_AMT in the Institutional claims files to the amount paid by Medicare [CLM_PMT_AMT], and adding the beneficiary liability). Refer to Table 10.

Table 9 and Table 10 present the variable names for certain types of payment found on Non-Institutional and Institutional Medicare claims (Version K of the layout), respectively.

Table 9. Variable Names for Payment Fields in Medicare Non-Institutional Claims

Payment Type	Non-Institutional Claims	
	Carrier and DME (Base File)	Carrier and DME (Line Item File)
Medicare Payment	CLM_PMT_AMT	LINE_NCH_PMT_AMT
Beneficiary Payment	CLM_BENE_PD_AMT*	LINE_BENE_PTB_DDCTBL_AMT + LINE_COINSRNC_AMT
Primary Payer Payment	CARR_CLM_PRMRY_PYR_PD_AMT	LINE_BENE_PRMRY_PYR_PD_AMT
Total payment due to the provider	NCH_CARR_CLM_ALOWD_AMT	LINE_ALOWD_CHRG_AMT (or could sum the 3 types of payment above)
Provider charges	NCH_CARR_CLM_SBMTD_AMT	LINE_SBMTD_CHRG_AMT

* In Version J of the claims layout, this field was not available. You would calculate beneficiary payments as: NCH_CARR_CLM_ALOWD_AMT – (CLM_PMT_AMT + CARR_CLM_PRMRY_PYR_PD_AMT). Alternatively, you could

³ There are some services that do not require beneficiary cost-sharing, such as hospice and home health care.

sum all of the line-level coinsurance amounts (LINE_COINSRNC_AMT) for the claim, and the claim-level deductible (CARR_CLM_CASH_DDCTBL_APLD_AMT). There is not a claim-level coinsurance variable.

b. Components of Payments

For Medicare Institutional services, residential care (i.e., for inpatient, SNF and hospice) is paid a per diem amount that is adjusted for beneficiary case mix (severity/acuity of services needed). For inpatient services, DRG is the basis for payment, whereas for SNF each day of care is paid using resource utilization groups, or RUGs.

As mentioned previously, many of Medicare's payment systems make additional outlier payments for cases where the provider incurs unusually high costs. To qualify, a case must typically have costs that exceed a certain dollar amount, with Medicare paying a percentage of the costs that exceed that threshold. For Part A, Medicare makes outlier payments to all types of inpatient hospitals that are paid prospectively (most acute care hospitals, inpatient psychiatric facilities, long-term care hospitals, and inpatient rehabilitation facilities), as well as home health agencies. Roughly 80 percent of Part A outlier payments are for acute care hospitals. When outliers are present, the amount is already included in the total Medicare payment amount field (CLM_PMT_AMT); note that the portion of the payment that is an outlier can be found on the inpatient base claim in the NCH_DRG_OUTLIER_APRVD_PMT_AMT field.

Medicare payments are described in detail in a series of Medicare Payment Advisory Commission (MedPAC) documents called "Payment Basics" (see: <http://www.medpac.gov/documents/payment-basics>), and also in the Medicare Learning Network (MLN) "Payment System Fact Sheet Series" (see the list of MLN publications at: <https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/MLN-Publications.html>).

Table 10. Variable Names for Payment Fields in Medicare Institutional Claims

Institutional Claim Type (on CMS 1450 forms)							
Data File	Inpatient	SNF	HOS	HH		HOP	
Payment Type	Base File	Base File	Base File	Base File	Revenue Center	Base File	Revenue Center
Medicare payment	CLM_PMT_AMT + (CLM_PASS_THRU_PER _DIEM_AMT x CLM_UTLZTN_DAY_CN T)	CLM_PMT_AMT	CLM_PMT_AM T	CLM_PMT_AM T	REV_CNTR_PMT_ AMT_AMT	CLM_PMT_AMT	REV_CNTR_PMT_ AMT_AMT
Beneficiary payment	CLM_IP_TOT_DDCTN_ AMT (or could sum NCH_BENE_IP_DDCTBL _AMT + NCH_BENE_PTA_COIN SRNC_LBLTY_AM + NCH_BENE_BLOOD_D DCTBL_LBLTY_AM)	CLM_IP_TOT_DDCTN_ AMT (or could sum NCH_BENE_IP_DDCT BL_AMT + NCH_BENE_PTA_COI NSRNC_LBLTY_AM + NCH_BENE_BLOOD_ DDCTBL_LBLTY_AM)	none	none	REV_CNT_PTNT_R SPNSBLTY_PMT	NCH_BENE_PTB_D DCTBL_AMT + NCH_BENE_PTB_C OINSRNC_AMT + NCH_BENE_BLOO D_DDCTBL_LBLTY_ AM	REV_CNTR_PTNT_ RSPNSBLTY_PMT (or could sum REV_CNTR_CASH_ DDCTBL_AMT + REV_CNTR_COINS RNC_WGE_ADJST D_C)
Primary Payer payment	NCH_PRMRY_PYR_CL M_PD_AMT	NCH_PRMRY_PYR _CLM_PD_AMT	NCH_PRMRY_ PYR_CLM_PD_ AMT	NCH_PRMRY_ PYR_CLM_PD_ AMT	none	NCH_PRMRY_PYR _CLM_PD_AMT	REV_CNTR_1ST_M SP_PD_AMT + REV_CNTR_2ND_ MSP_PD_AMT
Payment due to the Provider	n/a (could sum 3 above fields)	n/a (could sum 3 above fields)	n/a (could sum 3 above fields)	n/a (could sum 3 above fields)	n/a (could sum 3 above fields)	n/a (could sum 3 above fields)	n/a (could sum 3 above fields)
Provider charges	CLM_TOT_CHRG_AMT	CLM_TOT_CHRG_A MT	CLM_TOT_CHRG _AMT	CLM_TOT_CHRG _AMT	REV_CNTR_TOT_C HRG_AMT	CLM_TOT_CHRG_ AMT	REV_CNTR_TOT_C HRG_AMT

c. Zero payment or negative payment claims

Unless otherwise requested, investigators receive final action claims from CCW. The final action claim is the version of the claim where all adjustments to earlier claims have been resolved and CMS's final action on the claim is accurately recorded. Nonetheless, if your objective is to examine payment information, depending on your perspective, you may wish to exclude particular claims or lines of the claim. For example, it is possible that for some final action claims in the Non-Institutional settings (i.e., Carrier or DME claims), particular lines may be denied. Although this is already factored in to the base claim payment fields, if you are examining payments using line item records, we recommend that you remove all denied items (i.e., those where the line processing indicator code [LINE_PRC SG_IND_CD] does not indicate the line item was accepted; that is, keep only those where LINE_PRC SG_IND_CD = A (allowed), R (reprocessed/adjusted), or S (secondary payer). Another example is for Home Health claims, where providers may submit a "request for anticipated payment", also referred to as a "RAP" claim, which CMS uses to make interim payments. Final payments should appear on a non-RAP claim. These can be identified where the type of facility is Home Health (CLM_FAC_TYPE_CD=3), and CLM_SRVC_CLSFCTN_TYPE_CD= 2 or 3 (Inpatient or Outpatient PTB including HH bill classification), and CLM_FREQ_CD=2 (Interim claim).

There are times when the final action claim has a negative payment amount. This situation may occur, for example, when the beneficiary deductible and coinsurance may have exceeded the total payment amount due to the provider.

There are times when the final action claim is for zero payment (\$0). There are a variety of reasons Medicare may not have paid the provider for the service. For example, there could be a primary payer other than Medicare, or Medicare may have denied payment to the provider. A variable on the base claim can provide information regarding reason for nonpayment of the claim (see CLM_MDCR_NON_PMT_RSN_CD in the Institutional claims, or the CARR_CLM_PMT_DNL_CD in the Non-Institutional claims).

Some common methods, and sample code, for examining utilization and payments are shared below.

- Per Capita and Per User rates

Code Example 9. Find Proportion of Medicare FFS Beneficiaries Who Used Acute Inpatient Care During the Year

Consider whether you want to restrict the claims file to a particular population subset (e.g., only those with FFS coverage, or only those of a certain age). For this example, we examine acute utilization and payments only for beneficiaries with full FFS.

The following SAS[®] code is designed to identify all beneficiaries enrolled for the full year (or until time of death), and all those with an acute inpatient claim. Then we tabulate per beneficiary and per user payments.

The input data sources were created above, and consist of the *coverage* file (from code example 1 - to specify inclusion of beneficiaries enrolled in FFS, or to be able to identify the demographic characteristics of users) and the acute base claims file *acute* (from code example 2). Files are sorted, and then merged. Members of our population with and without acute IP care are identified. Finally, PROC MEANS is used to summarize payments for each record, and then another data step calculates metrics for the entire cohort.

```
/*sort the 2 input files, then merge*/

proc sort data= acute;
  by BENE_ID CLM_ID;
run;

proc sort data= b.coverage out=coverage;
  by BENE_ID;
run;

data acute_cov;
  merge coverage  acute;
  by BENE_ID CLM_ID;

run;

/* identify full FFS beneficiaries and acute inpatient care users*/
data b.acute;
  set acute_cov;
  by BENE_ID;

  if first.BENE_ID then
    bene_cnt=1;

  if first.BENE_ID and CLM_PMT_AMT~=' ' then
    user_cnt=1;

  where ffs15='Y';
run;

/*creating a file consisting of a summarized dataset*/
proc means data=b.acute noprint;
  output out=acute_summ(drop=_freq_ _type_) sum(CLM_UTLZTN_DAY_CNT
    CLM_PMT_AMT NCH_BENE_IP_DDCTBL_AMT bene_cnt user_cnt)=;
run;

/*Per User and Per Capita Days & Payments*/
data acute_summ;
  set acute_summ;
```

```
format CLM_UTLZTN_DAY_CNT 10.;
acute_pmt_per_user=round(CLM_PMT_AMT/user_cnt, .01);
acute_pmt_per_cap=round(CLM_PMT_AMT/bene_cnt, .01);
label
    acute_pmt_per_user='Per User Acute Payments'
    acute_pmt_per_cap='Per Capita Acute Payments'
    clm_utlztn_day_cnt='IP days'
    clm_pmt_amt='Total IP Payments'
    nch_bene_ip_ddctbl_amt='Total IP bene deductible amounts';
run;

proc print data=acute_summ;
run;
```

- Rates of Events – such as Acute Inpatient Hospital Readmissions

If you are examining all acute inpatient hospital stays, we recommend combining two of the Part A Inpatient settings identified earlier in the code examples – the acute and CAH. The code example below can be used to determine whether a patient with more than one IP claim may have been transferred from one facility to another as part of the same continuous hospitalization, or whether the claim constitutes a separate admission or readmission.

Code Example 10. Identify People Who Experience a Particular Event – a Hospital Readmission

Several data processing steps are necessary to accomplish this task.

First we identify all beneficiaries with an acute inpatient or CAH claim, and then sort the claims in date of service order to identify all inpatient stays. Transfers to different facilities are identified, so that they are not confused with readmissions (i.e., where there is a break in the acute stay, during which time a lower level of care may be provided, followed by another inpatient admission).

The following SAS® code is designed to combine all acute and CAH inpatient hospitalizations into distinct “stays”.

The input data sources were created above (in code example 2), and consist of the **acute** file (IP acute base claim records) and the CAH base file **cah**. Files are sorted, and then merged to identify transfers.

```
proc sort data=acute out=b.acute;
by BENE_ID descending CLM_FROM_DT descending CLM_THRU_DT CLM_ID;
run;

proc sort data=cah out=b.cah;
by BENE_ID descending CLM_FROM_DT descending CLM_THRU_DT CLM_ID;
run;

data IPstays(drop= temp_dt temp_prov);
set b.acute b.cah;
by BENE_ID descending CLM_FROM_DT descending CLM_THRU_DT CLM_ID;
```

```
retain stays temp_dt temp_prov;

if first.BENE_ID then do;
    stays=1; temp_dt=CLM_FROM_DT; temp_prov=PRVDR_NUM;
end;

else do;

if ~(NCH_BENE_DSCHRG_DT=. or
(temp_dt=CLM_THRU_DT and temp_prov~=PRVDR_NUM)) then stays+1;

if (temp_dt=CLM_THRU_DT and temp_prov~=PRVDR_NUM) then
transfer=1;
end;

temp_dt=CLM_FROM_DT; temp_prov=PRVDR_NUM;

run;

proc means data=IPstays noprint;
    by BENE_ID;
    output out=bene_IPstays(drop=_freq_ _type_) max(stays)=
sum(transfer)=;
run;

proc freq data=bene_IPstays;
    tables stays transfer;
    title "number of IP stays and transfers";
run;

proc sort data= IPstays out=b.IPstays;

    by bene_id descending stays CLM_FROM_DT CLM_THRU_DT
descending CLM_ID;
run;
```

The following SAS[®] code uses the *IPstays* file (created above, in Code Example 10) as the main input source for determining whether the patient experienced a hospital readmission.

Once the data are sorted into stays, a determination can be made regarding whether an acute hospitalization is a readmission. Investigators may specify their own time interval for a readmission; here we use 30 days.

```
/*Merging Stay file with Coverage */
proc sort data=coverage out=b.coverage;
    by BENE_ID;
run;

proc sort data=IPstays;
    by BENE_ID;
run;
data IPstays2;
    merge b.IPstays (in=x) b.coverage;
    by BENE_ID;
```

```
        if x;
run;

proc sort data=IPstays2;
    by BENE_ID stays;
run;

    /*Calculating readmissions - this example uses a 30 day interval*/

data readmits;
    set IPstays2;
    by BENE_ID;
    retain tempdate tempstd tempprov tempdrg;
    admit_clm=0;

if first.BENE_ID then do;

if PTNT_DSCHRG_STUS_CD not in ('20','07') and
(BENE_DEATH_DT=' ' or BENE_DEATH_DT-CLM_THRU_DT>30)

then do;
    admit_clm=1;
    readmit=0;
end;

tempdate=CLM_FROM_DT;
tempstd=CLM_PMT_AMT;
tempprov=PRVDR_NUM;
tempdrg=CLM_DRG_CD;
end;

else do;

    if PTNT_DSCHRG_STUS_CD not in ('20','07') and
year(CLM_THRU_DT)=2015 then do;
        admit_clm=1;
        readmit=0;
        if .<intck('day',CLM_THRU_DT,tempdate)<=30 then do;
            readmit=1; readmit_pmt=tempstd; readmit_prov=tempprov;
            readmit_drg=tempdrg;
        end;
    end;

tempdate=CLM_FROM_DT; tempstd=CLM_PMT_AMT; tempprov=PRVDR_NUM;

tempdrg=CLM_DRG_CD;

end;

output;

run;
```

```
/*macro is designed to do various tabulations - first overall, then  
by population subgroup - from the coverage file, which the user can  
specify */
```

```
%macro summ_read (var= );  
proc sort data=readmits;  
    by &var;  
run;  
  
proc means data=readmits noprint;  
    by &var;  
    where ffs15='Y';  
    output out=&var._summ(drop=_freq_ _type_) sum(admit_clm  
    readmit)=;  
run;  
  
data &var._summ;  
    set &var._summ;  
    readmit_rt=readmit/admit_clm * 100;  
    label readmit_rt='Readmission Rate';  
run;  
  
proc print data=&var._summ;  
    title "Readmission rates";  
run;  
  
%mend summ_read;  
%summ_read (var=RTI_RACE_CD);  
%summ_read (var= SEX_IDENT_CD);
```

Chapter 3: 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
AMA	American Medical Association
ASC	Ambulatory Surgery Center
BASF	Beneficiary Annual Summary File
BETOS	Berenson-Eggers Type of Service
CAH	Critical Access Hospital
CCU	Coronary Care Unit
CCW	Chronic Conditions Data Warehouse
CMS	Centers for Medicare & Medicaid Services
CPT	Current Procedural Terminology
DME	Durable Medical Equipment
DMERC	Durable Medical Equipment Regional Carrier
DRG	Diagnosis Related Group
DSH	Disproportionate Share Hospital
E&M	Evaluation and Management
ED	Emergency Department
ESRD	End Stage Renal Disease
FFS	Fee-for-Service
FI	Fiscal Intermediary
FQHC	Federally Qualified Health Center
HCPCS	Healthcare Common Procedure Coding System
HH	Home Health Agency
HOP	Hospital Outpatient
HMO	Health Maintenance Organization
ICD-9	International Statistical Classification of Diseases and Related Health Problems (9 th version)
ICD-10	International Statistical Classification of Diseases and Related Health Problems (10 th version; in effect starting 10/1/2015)
ICU	Intensive Care Unit
ID	Identification Number
IMC	Intermediate Care Unit
IME	Indirect Medical Education
IPF	Inpatient Psychiatric Facility
LTCH	Long Term Care Hospital
MA	Medicare Advantage
MAC	Medicare Administration Contractor
MA-PD	Medicare Advantage – Prescription Drug Plan
MBSF	Master Beneficiary Summary File
NDC	National Drug Code
NDI	National Death Index
NPI	National Provider Identifier

PDE	Prescription Drug Event
PDP	Prescription Drug Plan
ResDAC	Research Data Analytic Center
RHC	Rural Health Clinic
RIF	Research Identifiable File
SDA	Self-Decrypting Archive
SNF	Skilled Nursing Facility