

Chronic Conditions Warehouse

Your source for national CMS Medicare and Medicaid research data



Chronic Conditions Warehouse Technical Guidance

Getting Started with CMS Medicare Administrative Research Files

SEPTEMBER 2022 | VERSION 2.8

This page intentionally left blank.

Revision Log

Date	Changed by	Revisions	Version
September 2022	K. Schneider	Clarified that NDI file is only available within the CCW VRDC and where to find years of data available	2.8
May 2022	K. Schneider	Adjusted references to conditions data files to reflect the addition of the 30 CCW Chronic Conditions file (MBSF_CHRONIC)	2.7
March 2022	K. Schneider R. Van Gilder D. Happe	Updated SAS code examples to use 2018 files, added information regarding ACO payments, and converted to new document template	2.6
December 2017	K. Schneider R. Van Gilder	Modified formatting of SAS code example on page 37	2.5
October 2017	K. Schneider R. Van Gilder	Updated data source for MBSF to CME and adjusted MBSF SAS variable names when appropriate. Added information re: ICD-10-CM and ICD-10-PCS. Described the new demonstration/innovations reference code file. Updated code examples to use 2015 files. Updated payment variables in Tables 9 and 10	2.4

Table of Contents

Overview	1
Chapter 1. Contents of CCW Medicare Data Files	1
A. Medicare Enrollment Data – Master Beneficiary Summary File (MBSF)	2
B. Medicare Claims Files	2
1. Medicare Institutional Claims	3
2. Medicare Non-Institutional Claims	4
3. Medicare Prescription Drug Event Files	4
C. Claim Structure	5
1. Structure of Medicare Institutional Claims	5
2. Structure of Medicare Non-institutional Claims	6
3. Structure of Medicare Part D Events	7
D. Receiving CCW Data	7
Chapter 2. Methods	8
A. Sample Selection	8
1. Length of Enrollment in the Medicare Benefit	8
2. Type of Medicare Coverage	9
3. Age or Other Demographic Information	10
4. Aggregating Data to Summarize Coverage Variables	10
B. Examining Patterns of Health Care Utilization	13
1. Selecting Utilization Data Files	14
2. Extracting Utilization Information from Claims	14
3. Constructing a Longitudinal Data File	26
4. Tabulating or Summarizing Utilization	27
a. Tabulating Utilization of Care	27
C. Describing Utilization	30
1. Identifying Care for a Particular Diagnosis	30
2. Identifying Care for Individuals with a Particular Condition or Diagnosis	32
3. Identifying Particular Procedures	32
4. Payments for Services	34
a. Responsibility for payments	35
b. Components of payments	36
c. Zero payment or negative payment claims	38
d. Medicare payment reform and claim payment amounts	38
Chapter 3. Further Assistance with CCW Data	48
Appendix A — List of Acronyms and Abbreviations	49

List of Tables

Table 1. Reference code files	6
Table 2. Variable names for Medicare enrollment variables with monthly values	10
Table 3. Distribution of race code — full versus non-full FFS	12
Table 4. Variable names for variables related to months of eligibility	13
Table 5. Algorithms for use in categorizing Medicare Part A institutional claims	18
Table 6. Algorithms for use in categorizing Medicare institutional outpatient claims	19
Table 7. Algorithms for use in categorizing Medicare non-institutional claims.....	21
Table 8. Part B carrier and DME service event categories	24
Table 9. Algorithms for use in categorizing cost centers/areas within settings.....	25
Table 10. Number of OP and IP ED visits	28
Table 11. ED user per beneficiary by race code	29
Table 12. Diagnosis, procedure, and service codes used on Medicare claims.....	30
Table 13. ICD-10_PCS codes for CHF population.....	34
Table 14. Variable names for payment fields in Medicare non-institutional claims.....	36
Table 15. Variable names for payment fields in Medicare institutional claims	37
Table 16. Example — IP value code amount Q1 added to the claim payment amount	40
Table 17. Sample output, carrier line adjustment amounts	41
Table 18. Example adjustment to CLM_PMT_AMT in carrier data	42
Table 19. IP acute payment amounts.....	44
Table 20. All-cause re-admission rate example, using readmit_rate	47

List of Code Examples

Code Example 1. Define a sample of Medicare beneficiaries with fee-for-service (FFS) coverage	10
Code Example 2. Demonstrate partitioning Medicare Part A institutional base data file into specific types of inpatient settings.....	18
Code Example 3. Use outpatient base and revenue center data files to identify particular settings of care	20
Code Example 4. Use Part B carrier and DME line files to identify claim settings	22
Code Example 5. Use Part A inpatient revenue center file to identify specific settings of care within the inpatient facility	26
Code Example 6. Use base and revenue center records to identify care received in specific settings and calculate population-level utilization statistics.....	27
Code Example 7. Calculate population-level utilization statistics	28
Code Example 8. Identify procedures for a cohort with a particular condition of interest	33
Code Example 9. Identify inpatient claims with the Q1 value code.....	39
Code Example 10. Identify carrier lines with Line Other Applied Amount equal to L.....	40
Code Example 11. Medicare FFS beneficiaries who used acute inpatient care, and the associated claim payments	43
Code Example 12. Identify people who experience a particular event — a hospital re-admission	44

Overview

The Centers for Medicare & Medicaid Services (CMS) launched the Chronic Conditions Warehouse (CCW) in 2005 as a research data ecosystem that contains Medicare fee-for-service (FFS) institutional and non-institutional claims, assessment data, and enrollment/eligibility information from January 1, 1999, forward. CCW designed the data files to support a variety of research objectives. Users may find learning the nuances of any new data source challenging. Creating the appropriate analytic dataset can be the key to simplifying the data analysis.

This document provides technical guidance for approaching Medicare enrollment and administrative data complexities and provides a framework for creating customized, analytic data files. We provide an overall description of the Medicare data files in the [CCW Medicare Administrative Data User Guide](#) on the CCW website. This technical guidance document primarily focuses on the Medicare enrollment, Medicare Part A, and Medicare Part B FFS claims delivered by CCW as Research Identifiable Files or RIFs. In addition to narrative technical information, the CCW team shares some SAS® code examples to illustrate common analytic processes.

Chapter 1. Contents of CCW Medicare Data Files

Medicare is the primary health insurance program for people aged 65 or older, 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 help cover inpatient hospital care, skilled nursing facility (SNF) stays, home health and hospice care. Most beneficiaries also subscribe to Part B medical insurance benefits, which help 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 (managed care, currently known as Medicare Advantage, or MA). These are private plans similar to HMOs, which provide Medicare Part A and Part B services. The MA plans typically offer prescription drug coverage.

The CCW system contains data for 100% of Medicare-enrolled beneficiaries, not only those with a chronic condition. The Medicare institutional and non-institutional data files in the CCW have historically represented Medicare FFS claims. Managed care encounter data is available for people enrolled in Medicare Advantage (MA) plans, starting in 2015. The [Medicare.gov](#) website explains the Medicare managed care benefit in greater detail. Hospice services are always in the FFS claims files, even if the beneficiary elected MA coverage.

CCW data files include all Part D prescription drug events (PDEs) from all beneficiaries participating in the Part D program, regardless of whether the beneficiary enrolled in a managed care or an FFS plan. Furthermore, every prescription fill that the Part D benefit from 2006 forward (the inception of the benefit) covers has a PDE in the data file. Medicare institutional and non-institutional claims data are available from 1999 forward.

CCW removes the Medicare beneficiary Health Insurance Claim numbers (HICs)¹ from the data files delivered to researchers (unless otherwise specified/approved in the Data Use Agreement [DUA]). CCW adds a unique CCW beneficiary identifier (variable called the BENE_ID) to each data file. The unique CCW beneficiary identifier field is specific to the CCW and does not apply to any other identification system or data source.

¹ CMS began using a new Medicare Beneficiary Identifier (MBI) in place of the HIC starting in 2018.

The record layouts for all CCW data files are available on the ccwdata.org website, under the [Data Dictionaries](#) tab. Throughout this document, when we identify a particular data variable by name, we will often identify the specific SAS name, appearing in all capitals.

A. Medicare Enrollment Data – Master Beneficiary Summary File (MBSF)

The CCW team creates the MBSF for each calendar year. The MBSF contains demographic entitlement and enrollment data for beneficiaries who: 1) CMS documents as being alive for some part of the reference year; and 2) enrolled in the Medicare program during the file’s reference year. Reference year refers specifically to the calendar year accounted for in the MBSF. So, for example, the 2018 MBSF covers the year 2018 — which is the reference year. Currently, the enrollment data file CCW disseminates is the MBSF that uses the CMS Common Medicare Environment (CME) database as its source.²

The MBSF is an annual person-level summary file; each record is for a distinct BENE_ID. Researchers may use this file 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 they create. The CCW data files include the following types of variables: 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.

The MBSF file contains many enrollment and other person-level variables organized into file “segments.” These segments are separate components of the file that investigators may request. Within this guidance document, we 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. There are two additional segments that you may request as part of the MBSF: 1) NDI (National Death Index — for the cause of death) segment, and 2) Cost and Use segment (annual summary statistics for each beneficiary). Researchers may only use the NDI files within the CCW Virtual Research Data Center (VRDC). We do not address these in this guidance document; refer to the [CCW Medicare Administrative Data User Guide](#) on the CCW website.

The MBSF is available from 1999 forward.

B. Medicare Claims Files

CMS groups the Medicare administrative claims data files into three high-level categories: 1) institutional claims, submitted by healthcare facilities (such as hospitals or nursing homes), 2) non-institutional claims, submitted by physicians or physician groups, as well as DME suppliers, and 3) prescription drug event records, which represent administrative records for every prescription submitted under Medicare’s Part D benefit.

Healthcare providers often submit more than one version of a claim for a particular service because they need to revise the information on the initial claim for some reason. Any revision or adjustment requires a new claim. Although the CCW database includes a copy of each transaction claim record, the RIF contains only the final action claim. The final action claim is the version of the claim where the CCW team has resolved all adjustments to earlier claims (submitted by the provider or Medicare Administrative Contractors [MAC]) and accurately recorded the final action on the claim. Since weeks or months can pass between providing a service and submitting the final action claim, CCW

² The CCW team describes the rationale and impact of the data conversion from EDB to the CME database as the source for the MBSF in a CCW White Paper, [Master Beneficiary Summary File \(MBSF\): Impact of Conversion from EDB to CME](#).

generally waits for the final action claim to appear before extracting data files for delivery to researchers. We call the time claims files mature after the run-out period is 12 months. CCW published a white paper regarding Medicare claims maturity, available on the [CCW website](#).

Medicare claims include the diagnosis code(s) for the visit. 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 all services provided on or after October 1, 2015, utilize ICD-10 codes. Regardless of when a provider submitted a claim for payment, they billed using ICD-9 codes for services that occurred before October 1, 2015. This requirement applies to all medical claims in the U.S. and includes claims other than 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, providers must use ICD-10 codes.

The institutional and non-institutional claims data files have two components — the “base” and “line item” records, which we explain in greater detail in section [C. Claim Structure](#).

The CCW team adds key variables to the CMS administrative data files to help researchers join them together as appropriate (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 a researcher will use may vary depending on which files they join. For example, some projects will need to 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 researcher will want to link the information at the person level (i.e., the BENE_ID) after aggregating the claim level files.

1. Medicare Institutional Claims

The CCW team develops its institutional FFS claims files from data submitted by institutional providers and/or provider settings covered by the Medicare Part A benefit. 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 files. 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. CMS contractors known as MAC, formerly known as Fiscal Intermediaries (FI), are responsible for processing the Part A claims. More information regarding the contents and processing of the Medicare FFS claim forms is available on the [CMS website](#), and at <http://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/837I-FormCMS-1450-ICN006926.pdf>. Institutional claims files include the following five types of claims:

- **Skilled nursing facility (SNF)** —this file includes FFS claims submitted by SNF providers. This file includes ICD-9 or ICD-10 diagnosis codes, dates of service, reimbursement amount, and SNF provider number.
- **Inpatient** —this file includes FFS inpatient [IP] claims. Inpatient claims data files have claims from healthcare facilities that treat conditions that require the patient to stay at least one night in the care-related facility, e.g., an acute-care hospital. The IP file includes ICD-9-CM or ICD-10-CM diagnosis and ICD-10-PCS procedure codes, Diagnosis Related Group (DRG) information, dates of service, reimbursement amount, hospital provider, and beneficiary demographic information.
- **Home health agency (HH)** —this file includes FFS 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.

- **Hospice** —this file includes 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. Claims for all beneficiaries receiving hospice services appear in this file — even hospice services provided to beneficiaries enrolled in managed care plans.
- **Institutional outpatient** —this file includes FFS claims 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.

2. Medicare Non-Institutional Claims

The Medicare FFS 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. MACs process all claims. 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 two types of claims:

- **Carrier** — the carrier files contain FFS 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, or the national provider identifier (NPI)).
- **DME** — the Durable Medical Equipment Regional Carrier (DMERC) files contain FFS claims submitted by DME suppliers and providers for medical devices and supplies that the beneficiary can repeatedly use, such as a hospital bed or a prosthetic limb. 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).

As mentioned earlier, on October 1, 2015, the conversion from the 9th version of the ICD-9-CM to version 10 (ICD-10-CM and ICD-10-PCS) occurred. CCW includes the diagnosis code version used on the claim as a field in the carrier and DME data file layout (i.e., variable called IDC_DGNS_VRSN_CD1).

3. Medicare Prescription Drug Event Files

For Part D prescription drug events, one record represents a unique event — 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 or 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](#). In addition, a technical guidance paper entitled [CCW Technical Guidance: Summarizing and Describing Prescription Drug Utilization](#) is on the CCW website.

CMS covers some drugs under the Medicare Part B benefit, and these drugs 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. Refer to the CCW website for [file record layouts and definitions](#). Most claim types consist of multiple related files.

1. Structure of Medicare Institutional Claims

For institutional claims, the **base file** contains the base claim record. In contrast, the **revenue center file** contains at least one and potentially multiple revenue center records for a corresponding base claim record. Users also refer to revenue center records as “revenue line” records.

- **Base 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. 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** — 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. Researchers should not use the revenue center payments fields since the total payments are on the base file. Reference the [Medicare Fee-For-Service Institutional Claim](#) file layout on the CCW website for more detail.
 - You may only need to use the base file; however, if you are interested in more detailed procedure information, you will benefit from the revenue center detail. The revenue center data includes institutional cost centers for separately billed charges. For example, there are revenue centers for the emergency department (ED), intensive care unit (ICU), 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 data contains important information to help distinguish between care settings (e.g., clinic versus dialysis care). We present additional information regarding how to use the revenue center information to distinguish between types of outpatient services in [Chapter 2.B](#). Records in the revenue center file are associated with the header claim (i.e., a particular CLM_ID). CCW sequentially numbers the revenue center lines with the claim line number (CLM_LINE_NUM).
 - Additionally, CCW disseminates four types of reference code files to link to the base claim, including condition, occurrence, span, and value code files. Many projects will never use the reference code files; however, they contain information regarding special conditions which may affect payer processing. For example, inpatient claims analysts may use the Value codes reference file to obtain the operating Indirect Medical Education (IME) Amount and the operating Disproportionate Share (DSH) Amount fields. These are important to understand if you are examining geographic variation in payment amounts or variation by facility type. Reference [Table 1](#) for more detail on reference code files.
 - Starting in 2010, CCW also delivered a demonstration/innovations code file with each institutional 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 project related to the claim (e.g., the identification number for the project).

Table 1. Reference code files

Filename	Description
Condition codes	Codes indicating 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 identifying a significant event/date relating to an institutional claim (e.g., accident, medical condition, insurance-related, service-related, etc.). The file also includes the corresponding date of the occurrence.
Span codes	Codes relating to a time period or span of dates, such as during an institutional stay (e.g., exhausted all full/coinsurance days but covered on the cost report, hospital prior stay dates, visits occurring in this billing period if different, etc.). The file also includes the span code situation from and through dates.
Value codes	Codes indicating a value of a monetary condition used by the payer (e.g., MAC) to process an institutional claim (e.g., the wage index applied to home health care due to the beneficiary location, the amount for some CMS payment reform initiatives). The file gives the amount indicated by the value code.
Demonstration/ Innovations codes	A record in this file indicates the claim was part of a demonstration or innovation project. The file also includes the demonstration ID number and descriptive text.

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](#).

2. Structure of Medicare Non-institutional Claims

A claim in the **base file** will correspond to one or more associated line-item records in the **line file**.

- **Base 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. Beginning with claims files from 2009, the claims (beginning with version J of the layout) allow for 12 diagnosis codes.
- **Line file** — contains the individual line level information from the claim. This includes the HCPCS code(s), diagnosis code(s), first and last expense (service) dates, line allowed charge amount, line submitted charge amount, line paid amount, and performing provider identifier.

It is common to find the line-item detail helpful because non-institutional providers price and beneficiaries pay at the line level. 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. Records in the line files are associated with the base claim using the claim ID (CLM_ID). The line file includes at least one (and often many more) claim lines, numbered sequentially in the claim line number field (LINE_NUM).

Starting in 2010, CCW disseminates a demonstration/innovation code file with each claim type to link with the base claims. Similar to the reference code files (refer to [Table 1](#)), it provides information regarding any innovation or demonstration projects related to the claim (e.g., the identification number for the project).

3. Structure of Medicare Part D Events

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

D. Receiving CCW Data

Researchers have two options for accessing CCW data files — they may access them directly from the CCW within the Virtual Research Data Center (VRDC), or have their data shipped to them to use in their own data center. The CCW team initiates access to the VRDC data users upon receipt of the approved DUA and CMS notification the researcher paid fees.

For researchers who request CCW to ship their data, the CCW team outputs the approved data files on either a USB external hard drive or a DVD/CD. The CCW team packages these data files as encrypted self-decrypting archive (SDA) files (reference the [CCW Medicare Administrative Data User Guide](#) on the CCW website for additional information on encryption). The CCW team sends you the decryption password 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 users may be able to decrypt the files on the shipped 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. Within each SDA, the CCW team includes the data file(s), SAS code, and a file transfer summary (.fts) file, which you can use to verify the data.

Chapter 2. Methods

In this chapter, the CCW team describes methods analysts may employ with the standard record layout (reference ccwdata.org) for Medicare FFS data files that CCW delivers to data users. Although the variable names and values are generally consistent over time, you may need to make slight modifications (e.g., with the onset of Part D data in 2006, additional data became available).

The CCW team suggests that researchers may prefer several smaller analytical files to manipulate the data for the desired results easily.

A. Sample Selection

Most researchers request the MBSF, which includes Medicare enrollment and coverage information for all beneficiaries in their study population. This file contains a single row of data for each beneficiary. It will be important to determine whether you 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, you may hypothesize that some treatment patterns may accurately assess only for people of a certain age or 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 beneficiaries “at risk” for events of interest to you. Not all Medicare beneficiaries enroll in FFS Medicare; however, with few exceptions, only FFS claims appear in the data files (one of the notable exceptions is hospice claims for managed care enrollees, which appear in the FFS hospice claims data files). CMS explains nuances of the managed care benefit in greater detail in the CMS Policy Manual.³

For Medicare Part D data, all covered prescription drug fills are present, regardless of whether the beneficiary enrolls 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 for you to exclude someone from your sample. For example, if the subject died during the surveillance period or other times when it is important to capture the death as an outcome. The death variable in the Medicare data files only indicate deaths that occurred in the data file year. The Medicare data files do not include deaths occurring after the end of the calendar year. The CCW team recommends caution when requiring a certain length of Medicare coverage because Medicare terminates coverage after death. For example, if you are studying acute myocardial infarction (AMI) and want to examine cases for 2015, you should restrict your denominator (beneficiary sample) 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 exclude beneficiaries who did not survive the AMI. Best practice is 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, refer to a Technical Guidance paper called, [CCW Technical Guidance: Calculating Medicare Population Statistics](#) on the CCW website. Later in this document, the CCW team presents a code example for selecting a sample based on Medicare coverage.

³ CMS. [Medicare Managed Care Manual. Chapter 4 – Benefits and Beneficiary Protections](#). (accessed 11/10/2021)

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. The reason you should not select beneficiaries enrolled in Medicare Part C is because they have managed care coverage, and the FFS claims data files do not include transactional data regarding services received through Medicare Advantage plans.⁴ Variables within the MBSF indicate the type of Medicare coverage obtained each month.

The Medicare state buy-in variable appears 12 times in the MBSF to represent each month of coverage (MDCR_ENTLMT_BUYIN_IND_01–12). The values for 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. State Medicaid programs can pay Medicare premiums for certain dual eligibles (i.e., for beneficiaries also enrolled in a state Medicaid program); CMS calls this action “buying in” and so this variable is the “buy-in code.” One limitation of the state buy-in variable is that it does not provide information regarding whether Medicaid rules entitle the beneficiary to full or only partial benefits. The CCW team includes a summary variable in the MBSF which counts the months of state buy-in. Beginning with 2006 data, the CCW files also include the state-reported dual status. The monthly dual-status variable (DUAL_STUS_CD_01–12) describes the level of dual benefits (e.g., none, partial, full). The [CCW Technical Guidance: Options for Determining Which CMS Medicare Beneficiaries are Dually Eligible for Medicare and Medicaid Benefits](#) paper provides additional details regarding options for identifying populations dually enrolled in Medicare and Medicaid.

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 (HMO_IND_01–12). CCW includes a summary variable that counts the months of managed care coverage (BENE_HMO_CVRAGE_TOT_MONS) in the MBSF.

The MBSF contains additional information regarding persons dually eligible for both Medicare and Medicaid services. As stated previously, beginning with 2006 data, the monthly state-reported dual-status code is available (DUAL_STUS_CD_01–12). In addition, CCW added a monthly cost share group field (CST_SHR_GRP_CD_01–12) when Part D enrollment information became available with the 2006 benefit year. The Part D benefit allows for premium and/or coinsurance subsidies for low-income Medicare enrollees who do not qualify for Medicaid. The level of the subsidy varies by income level; therefore, the cost-share group variable contains some interesting information regarding poverty, which is more granular than just knowing someone was Medicaid-eligible. It contains information regarding who Medicaid “deems” eligible for the low-income subsidy. Refer to the CMS Medicare website⁵ for additional details regarding the low-income subsidy (LIS) provisions within the Part D benefit.

Refer to [Table 2](#) for the variable names of interest for the monthly Medicare enrollment data.

⁴ Managed care encounter data is available for people enrolled in Medicare Advantage (MA) plans, starting in 2015. The CCW provides the [CCW Medicare Encounter Data User Guide](#) and the [record layouts](#) for these files on the CCW website.

⁵ <https://www.cms.gov/Medicare/Eligibility-and-Enrollment/LowIncSubMedicarePresCov/EligibilityforLowIncomeSubsidy> (accessed 11/10/2021)

Table 2. Variable names for Medicare enrollment variables with monthly values

Variables with monthly values	SAS variable name <i>The last two 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 aged 65 or older. The majority of older adults in the U.S. enroll in Medicare; therefore, researchers often use the Medicare administrative data for studies regarding medical care for older adults nationwide. The Medicare program also provides insurance coverage for people under age 65 with certain disabilities, and people of all ages with ESRD or ALS. The beneficiary population enrolled in Medicare under age 65 differs from the general Medicare population and the general U.S. population in important ways to consider when selecting a sample for the study.

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

Researchers will want to consider a variety of study design questions when using Medicare data for population statistics. For example, there are a variety of approaches to calculate Medicare enrollment counts or duration of coverage, such as person-time of enrollment in FFS Medicare. The CCW team discusses some of these considerations in a separate paper titled, [CCW Technical Guidance: Calculating Medicare Population Statistics](#) on the CCW website.

4. Aggregating Data to Summarize Coverage Variables

This document describes useful coverage type variables included within the MBSF 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. 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, we advise that you truncate coverage at the month of death (or month before the month of death). The CCW team developed a code example below.

Code Example 1. Define a 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 uses the Medicare Part A, B, C, D (or base) segment of the MBSF. We read in a subset of variables (since files are large) and keep the monthly coverage variables to require at least 11 months of FFS coverage; the next portion of the code demonstrates one use of the summary variables, which count the number of covered months. The input data source is the MBSF ABCD segment from 2018 (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 Enrollment Database (EDB) to using the CME as the source, it resulted in updated SAS names for some of the MBSF fields).

NOTE: For all of the coding examples, CCW data presents both the filenames and variable names in all capital letters if the CCW team pulls directly from the source (CCW) data file; for all other derived files and variable names, we may use a combination of upper- and lower-case letters.

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

```

data work.coverage (keep=BENE_ID SEX_IDENT_CD RTI_RACE_CD AGE_AT_END_REF_YR
BENE_DEATH_DT STATE_CODE Member_Mos Cov18 Ffs18 Ptd18);

set MBSF.MBSF_ABCD_2018 (keep=BENE_ID MDCR_ENTLMT_BUYIN_IND_01 -
MDCR_ENTLMT_BUYIN_IND_12 HMO_IND_01 - HMO_IND_12 BENE_HI_CVRAGE_TOT_MONS
BENE_HMO_CVRAGE_TOT_MONS BENE_SMI_CVRAGE_TOT_MONS PTD_PLAN_CVRG_MONS
BENE_DEATH_DT SEX_IDENT_CD RTI_RACE_CD AGE_AT_END_REF_YR STATE_CODE);

/*If you are using the legacy MBSF AB file and the MBSF D segment
(which CCW disseminated before 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)
thenCov18=1;
else Cov18=0;

  /*coverage using summary variables*/
if BENE_HMO_CVRAGE_TOT_MONS=0 and ((year(BENE_DEATH_DT)~=2018 and
BENE_HI_CVRAGE_TOT_MONS=12 and BENE_SMI_CVRAGE_TOT_MONS=12)
  or (year(BENE_DEATH_DT)=2018 and
  BENE_HI_CVRAGE_TOT_MONS=month(BENE_DEATH_DT) and
  BENE_SMI_CVRAGE_TOT_MONS=month(BENE_DEATH_DT))) then ffs18='Y';
else ffs18='N';

  /*Part D coverage */
if ((year(BENE_DEATH_DT)~=2018 and PTD_PLAN_CVRG_MONS=12 ) or

```



```

        (year(BENE_DEATH_DT)=2018 and
        PTD_PLAN_CVRG_MONS=month(BENE_DEATH_DT)) then ptd18='Y';
    else ptd18='N';

label
    Cov18= '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'
    ffs18 = '12 months of Part A&B FFS-no HMO, or until death month'
    ptd18 = '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 tabulate data= work.coverage;
class RTI_RACE_CD ffs18;
table RTI_RACE_CD * (N), ffs18;
title 'Demographic description of full compared to non-full FFS
population';
run;

```

We present the tabulation for the demographic description of full compared to non-full FFS population in [Table 3](#).

Table 3. Distribution of race code — full versus non-full FFS

Research Triangle Institute (RTI) race code	No – less than 12 months of Part A and B FFS	Yes 12 months of Part A and B FFS, no HMO, or until death month
0 (Unknown)	554,073	511,713
1 (Non-Hispanic White)	20,997,019	25,405,860
2 (Black (Or African American))	3,759,260	2,841,495
3 (Other)	289,462	247,273
4 (Asian/Pacific Islander)	1,297,025	833,838
5 (Hispanic)	4,073,621	1,837,055
6 (American Indian / Alaska Native)	97,350	185,740

```

/*alternatively – you can subset the population to only tabulate results for FFS
benes*/

```

```

proc tabulate data= work.coverage;
class RTI_RACE_CD ffs18;
where ffs18 = 'Y';
table RTI_RACE_CD * (N), ffs18;
title 'Demographic description of full FFS population';
run;

```

With 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.

You may wish to require full or partial Part D coverage if you plan to examine Part D event data (e.g., PTD_PLAN_CVRG_MONS ='11' or PTD_PLAN_CVRG_MONS ='12'). Researchers may find that monthly coverage detail

is important for 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, a sample based on having coverage for the full year (or until the time of death) may suffice. Analysts may select variables from the MBSF as key study variables of interest, such as whether both Medicare and Medicaid enrolled the beneficiary. For other projects, you may find that a more summarized version of this information will suffice (e.g., knowing that the beneficiary was dually eligible for at least one month of the year).

CCW describes options for summarizing dual enrollee beneficiary coverage information 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.

The CCW includes a few variables in the MBSF, which summarize the number of months of particular types of coverage for each beneficiary during the year. Refer to [Table 4](#) below for the relevant variable names.

Table 4. Variable names for variables related to months of eligibility

Variables that count the number of eligibility* months	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 the variables in this file as some may not be of interest for your study. **NOTE:** the data file can become very wide if you intend to merge multiple years of data, or information regarding utilization from Medicare claims data files. In the code examples, we illustrate how to read in only the variables you need as a way of limiting your file size.

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 beneficiaries do not use any medical services, while others may use services paid for 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

CMS partitions Medicare administrative FFS claims into claim types. Claim types may encompass a variety of different payment rules and processes which apply to different care settings. Since each of the claim's files contain a large number of variables, we recommend extracting only the required variables and observations for your analysis.

There are some types of services that providers may bill using either an institutional or non-institutional claim, or both claim types. In general, provider claims for the professional component of a service (e.g., the physician or therapist care, etc.) appear in the non-institutional file (i.e., Part B carrier), whereas facility or institutional provider claims for the service appear in the institutional file (e.g., Part A inpatient or skilled nursing facility [SNF] files).

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

Another common example of a service that may appear on an institutional or non-institutional claim is a surgical procedure, which a provider may perform 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. The CCW team recommends you 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. The CCW team recommends several steps to organize utilization accurately:

1. Verify you are examining utilization for the appropriate claim type category (institutional, non-institutional, or both);
2. Determine the claim type or care setting(s) with a claim type, 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 (e.g., 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 the previous section: [1.C. Claim Structure](#). For some research questions, you could use more than one methodology to address the question. For example, there are numerous ways you could classify Medicare claims, and the CCW team does not

intend to be prescriptive, but rather we wish to demonstrate one method which may be helpful to analysts when beginning to examine Medicare administrative data files.

- **Utilization from institutional claims** — institutional claims cover a variety of setting types, some of which are inpatient, while others are outpatient.

Often, providers include the entire inpatient stay on a single claim; however, this is not always the case, particularly for lengthy stays. To be certain you have captured all 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). There are different ways to define a care setting using administrative institutional claims data; below are descriptions of several commonly used variables.
- Investigators may determine an individual stay by sorting all claims for a beneficiary within a type of service by descending first service date, last service date, and claim ID. Acute inpatient services, 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 5](#) for an illustration of how to identify the different inpatient settings.
- You might identify a new stay as a record with a missing discharge date that is not a transfer and is different from the previous discharge date in the ordered claims. Many hospitals do not discharge beneficiaries to home after an inpatient stay — rather, they continue to use institutional or community-based post-acute care. A discharge status code (variable called PTNT_DSCHRG_STUS_CD) of 30 indicates they are “Still a Patient.”

The CCW team has identified important variables to consider for the classification of services for institutional claims below.

- **Claim type** — on the base claim record describes the type of claim that the MAC processed (NCH_CLM_TYPE_CD). The data dictionary describes the values (e.g., inpatient hospitals are claim types 60 and 61, SNF are claim types 20 and 30, hospices are claim type 50, and HH agencies are claim type 10).
- **Facility type (CLM_FAC_TYPE_CD)** — indicates the type of facility that cared for the beneficiary. Categories include hospital, SNF, HH, Intermediate Care, dialysis facilities, etc.
- **Type of service (CLM_SRVC_CLSFCTN_TYPE_CD)** — indicates the type of service provided to the beneficiary, and researchers can use it in combination with facility type. This code helps identify rural and FQHCs as well as **Community Mental Health Centers (CMHC)**.
- **Bill type** — derive this variable to classify the type of services that appear on the claim. We classify bill type using a combination of two variables on the original claim,⁶ 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

⁶ The standard definition of bill type code uses three variables rather than two; the final (rightmost) digits are from the variable called CLM_FREQ_CD.

=13 (hospital outpatient department). An example of using the bill type variable appears in [Table 6](#), and a sample SAS code for creating this field is in [Code Example 3](#).

- **Provider number (PRVDR_NUM)** — you may use the first two digits of a provider number to identify the provider state. The third and fourth digits identify the type of facility. This field is the facility identifier that pre-dates the National Provider Identifier (NPI). Analysts may wish to use this information with a specific claim type to identify payment settings.
- **RHCs and FQHCs** — you identify these facilities 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 uses the last four digits of the Provider Number (PRVDR_NUM).
- **Revenue center (REV_CNTR)** — identifies the cost center within a hospital or facility that provides the service. The revenue center value identifies services delivered in the ED, intensive care unit (ICU), and intermediate care unit (IMC), to name a few commonly studied revenue centers. This information is from the institutional revenue center file, not the institutional base claim file.

The CCW team identifies some care types which appear in institutional claims, and providers can bill in more than one way below.

- **Emergency department care utilization** — use the revenue center data associated with an institutional claim to determine whether the beneficiary received ED services. Medicare has two methods for billing for these services, depending on whether the hospital admitted the beneficiary. Researchers should use the inpatient revenue center code (REV_CNTR) to identify ED services that preceded a hospital admission; otherwise, the outpatient revenue center will identify beneficiaries discharged from the ED. Often a single ED visit will have bills for the facility and the provider caring for the patient.
 - Researchers can identify the hospital outpatient and acute inpatient ED claims 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. Researchers can identify the professional component of the ED care by revenue center code 0981. There may also be a physician claim for the professional component of an ED visit, which appears in the Part B carrier claims with a Berenson-Eggers Type of Service (BETOS) code 'M3'. CMS derives BETOS codes from the individual HCPCS code on the claim (reference the following section on "Utilization from non-institutional Claims" for more details regarding BETOS codes).

A SAS code example for identifying 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. The CCW team highlights some important variables to consider for classifying services as ambulatory or physician office claims 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 provider's location of the service (e.g., office or clinic).
 - **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.).
 - **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 categories researchers can use for the analysis of

patient care. These codes allow data users to distinguish between care provided by physicians and other types of services.

- For example: Researchers can identify claims with BETOS codes of D1A, D1B, D1C, D1D, D1E, and D1F as DME claims. BETOS code D1G indicates DME drug, and O1C is parenteral nutrition utilization.
- **Utilization from both institutional and non-institutional Claims** —some types of medical utilization may appear in both institutional and non-institutional claims files. The CCW team depicts two scenarios below.
 - **Outpatient physician care** —visits with health care professionals in the ambulatory setting may occur in an outpatient facility or a provider office setting.
 - Researchers can identify outpatient care from the institutional outpatient claims files. Researchers can use revenue center codes (found on the revenue center file) to identify the nature of care (e.g., clinic dialysis center, telemedicine). CCW includes the revenue center code table in the [FFS claims codebook](#) on the CCW website. Alternatively, you may identify the CPT® codes used to bill for the types of visits of interest (e.g., providers use 99201–99205, 99211–99215 for office visits for new or established patients). Researchers may reference BETOS codes described below to obtain the HCPCS codes of interest.
 - Physician office care is a small portion of the claims found in the Part B carrier files. Suppose 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). In that case, you must employ different procedures for two different claim types (i.e., the hospital outpatient/clinic and the provider office). Researchers should use both the outpatient revenue center file and the Part B carrier line file. [Table 5](#) shows some classification examples below.
 - For the Part B carrier claims, identify the relevant BETOS codes, which the CCW team has crosswalked to the appropriate HCPCS codes. The BETOS_CD data field appears in the carrier line-item files (not the Base files).
 - **The Medicare Part B benefit covers some drugs** — such as intravenous chemotherapy, other infused drugs, and some vaccines. You will find claims for these services using the HCPCS on the revenue center file for institutional claims, or on the line file for non-institutional claims (e.g., find many chemotherapy drugs in HCPCS J9000–J9999; pneumococcal vaccine may use HCPCS 90669, 90670 or 90732, depending on the age of the patient and the particular vaccine).

The CCW team shares algorithms that may help categorize services from institutional claims in [Table 5](#) below. CCW uses four categories for the type of facility (refer to the leftmost column of [Table 5](#)). If you would like to be more granular in identifying institutional care settings, the setting-description column of the table divides the Part A claims into nine settings. CCW shares additional algorithms for institutional outpatient and non-institutional claims in [Table 6](#) and [Table 7](#); algorithms to identify particular cost centers within a setting are in [Table 8](#).

Table 5. 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 and 4th digits = 13
Other inpatient hospital	60,61	Inpatient psychiatric facility (IPF)	3rd and 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; specific provider numbers for cancer centers
Post-acute care	60,61	Long-term care hospital (LTCH)	3rd and 4th digits = 20, 21, 22
	60,61	Inpatient rehabilitation facility (IRF)	Last 4 digits = 3025–3099 or 3rd digit = R or T
	20,30	SNF	n/a
	10	Home health (HH)	n/a
Hospice	50	Hospice (HOS)	n/a

In the following SAS code examples, the CCW team illustrates how to identify the claim records for the care settings of interest. For [Code Example 2](#), we use only claim types 60 and 61 (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 files into specific types of inpatient settings

The CCW team designs the following SAS code to subset the claims in the Part A inpatient base claims file into smaller files specific to the different hospital care settings. For this DATA step, we end up with six inpatient setting-specific working 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. This example uses a one-month sample of data.

```

data work.acute work.cah work.ipf work.ltch work.irf work.oth_ip;
  set RIF2018.INPATIENT_CLAIMS_01 /*Version K of the RIF claims layout
  became available in 2017 */ (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_CD25 ICD_PRCDR_CD1 - ICD_PRCDR_CD25);
  if substr(PRVDR_NUM,3,1) in ('0') then
    output work.acute;
  else if substr(PRVDR_NUM,3,2) in ('13') then
    output work.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 work.ipf;
else if substr(PRVDR_NUM,3,2) in ('20','21','22') then
      output work.ltch;
else if substr(PRVDR_NUM,3,2) in ('30') or substr(PRVDR_NUM,3,1)
      in ('R','T') then output work.irf;
/*The irf file in this example uses 2-digit PRVDR_NUM for
      brevity. To be more accurate and align with guidance in Table
      4, you would list all provider numbers in the range of 3025-3099.
      For example - else if substr(PRVDR_NUM,3,4) in
      ('3025','3026','3027'...'3099') or substr(PRVDR_NUM,3,1) in
      ('R','T') then output irf;
else output work.oth_ip;
run;

```

We consider the hospital or institutional outpatient (OP) claims to be institutional data files due to the type of claim used to bill Medicare, even though the Medicare Part B benefit generally pays these claims. All OP claims, also called institutional outpatient claims, consist of claim type = 40; various institutional providers use this. The CCW team recommends categorizing OP claims according to the type of provider or specific settings of care. To accomplish this objective, we create a two-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). In [Table 6](#) below, we display algorithms that may help you categorize services from institutional outpatient claims. We share a coding example following the table.

Table 6. Algorithms for use in categorizing Medicare institutional outpatient claims

Facility type	Bill type	Setting description*	Bill type
Hospital outpatient department (OP)	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
		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; the code examples also use these.

The CCW team prepared the following SAS code examples to depict how to identify the claim records for various care settings within the OP setting. For [Code Example 3](#), we use both the outpatient base and outpatient revenue center files. **CCW pre-sorts the claims files and indexes them by BENE_ID and CLM_ID so that researchers can easily merge the files without using a separate sorting step.** We will 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 CCW team designed the following SAS code to associate the claims in outpatient data files with particular care settings. For this DATA step we end up with 11 outpatient setting-specific working files. This example uses a one-month sample of data.

```

data work.opps work.cah_opd work.fqhc work.rhc work.npta work.snf
work.thrpy work.cmhc work.bt34 work.bt14 work.bt83;
  merge RIF2018.OUTPUTPATIENT_REVENUE_01 /*Version K of the RIF claims
  layout became available in 2017 */ (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 REV_CNTR_PRVDR_PMT_AMT
  REV_CNTR_PTNT_RSPNSBLTY_PMT RNDRNG_PHYSN_NPI RNDRNG_PHYSN_UPIN)

  RIF2018.OUTPUTPATIENT_CLAIMS_01 (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 work.OPPS;
  else if bill_type in ('72') then
    output work.cah_OPD;
  else if bill_type in ('73','77') then
    output work.FQHC;
  else if bill_type in ('71') then
    output work.RHC;
  else if bill_type in ('12') then
    output work.NPTA;
  else if bill_type in ('22','23') then
    output work.SNF;
  else if bill_type in ('74','75') then
    output work.THRPY;
  else if bill_type in ('76') then
    output work.CMHC;
  else if bill_type in ('34') then
    output work.BT34;
  else if bill_type in ('14') then
    output work.BT14;

```

```

else if bill_type in ('83') then
  output work.BT83;
/*BT83 does not have any records for 2014 and beyond*/

run;

```

For the Part B non-institutional claims, either “carrier” claims (claim types 71 and 72) or DME claims (types 81 and 82), the individual claim lines have detailed information on the nature of services provided and the corresponding payment. For these non-institutional claims, this information is not available on the base claim. For Part B non-institutional services, CCW bases our service categories largely on the BETOS classification scheme that groups HCPCS codes into distinct categories. We present an algorithm for identifying types of service in [Table 7](#); a code example follows. For the non-institutional claims, we identify ASC services first since the BETOS codes overlap with other non-ASC categories.

Table 7. Algorithms for use in categorizing Medicare non-institutional claims

Type of service	Description*	Values (BETOS codes; BETOS_CD)
ASC†	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	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	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	Standard imaging (STD)	I1A–I1F
	Advanced imaging (ADV)	I2A–I2D
	Echography (ECHO)	I3A–I3F
	Imaging procedure (IMGPRC)	I4A, I4B
LABTST	Laboratory test (LAB)	T1A–T1H

Type of service	Description*	Values (BETOS codes; BETOS_CD)
	Other test (TEST)	T2A–T2D
DME	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 services appear in parentheses, also used in the code examples.

† To classify ASC services, we first identify claims where the CLM_TYPE_CD=71 or 72 and LINE_CMS_TYPE_SRVC_CD='F'. The algorithms in this table use a hierarchy.

The CCW team presents a code example below to demonstrate how to partition the Part B line-item records into service settings (note: this code does not use the carrier or DME Base claims). We combine information from the CCW carrier and DME line files to classify the service types. Investigators may modify this code to meet their particular research needs. You may further segment these service types using BETOS Codes, which CMS derives from HCPCS codes.

Code Example 4. Use Part B carrier and DME line files to identify claim settings

The CCW team designed the following SAS code to identify service types in the Part B carrier and DME line files. First, the CCW team combined the claims from both files. Then, we identified the service types; we also illustrated how more granular coding for settings can be employed. This example uses a one-month sample of data.

```

data work.PartB;
set RIF2018.BCARRIER_LINE_01 */ (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)

RIF2018.DME_LINE_01 (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, we use only NPI, 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 mmddy10.;

run;

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

proc print data= work.partb_freq noobs;
where count > 10;
run;

```

We present the service categories created in this code example in [Table 8](#).

Table 8. Part B carrier and DME service event categories

srvc_1	srvc_2	betos2	Count	Percent
—	—	—	17054	0.0103
ASC	—	—	35	0.0000
ASC	—	D1	13479	0.0082
ASC	—	I1	12838	0.0078
ASC	—	I3	3524	0.0021
ASC	—	I4	4288	0.0026
ASC	—	M2	16	0.0000
ASC	—	M5	800192	0.4839
ASC	—	O1	15619	0.0094
ASC	—	P0	1751	0.0011
ASC	—	P1	37142	0.0225
ASC	—	P2	895	0.0005
ASC	—	P3	11289	0.0068
ASC	—	P4	172949	0.1046
ASC	—	P5	60568	0.0366
ASC	—	P6	138509	0.0838
ASC	—	P7	238	0.0001
ASC	—	P8	205843	0.1245
ASC	—	P9	57	0.0000
ASC	—	T1	1511	0.0009
ASC	—	T2	506	0.0003
ASC	—	Y1	178	0.0001
ASC	—	Z2	5485	0.0033
DME	—	D1	7311066	4.4213
DRUG	—	O1	5979402	3.6160
EM	—	M1	23527323	14.2278
EM	—	M2	11168873	6.7542
EM	—	M3	1988924	1.2028
EM	—	M4	3223032	1.9491
EM	—	M5	15369064	9.2942
EM	—	M6	17168	0.0104
IMG	—	I1	8546618	5.1684
IMG	—	I2	3127900	1.8916
IMG	—	I3	2850966	1.7241
IMG	—	I4	446366	0.2699
LABTST	—	T1	35927550	21.7267
LABTST	—	T2	7006183	4.2369
OTHER	—	O1	4583272	2.7717
OTHER	—	Y1	1062758	0.6427
OTHER	—	Y2	606229	0.3666
OTHER	—	Z2	7536408	4.5575
PROC	AMBLTY	P5	3689569	2.2312
PROC	ANES	P0	1778084	1.0753

srvc_1	srvc_2	betos2	Count	Percent
PROC	DIALYS	P9	503505	0.3045
PROC	ENDO	P8	925250	0.5595
PROC	EYE	P4	797143	0.4821
PROC	MAJCAR	P2	384396	0.2325
PROC	MAJORT	P3	319389	0.1931
PROC	MAJOTH	P1	516096	0.3121
PROC	MINOR	P6	13728399	8.3020
PROC	ONCOL	P7	936705	0.5665

You may find it helpful to understand the types of service the beneficiary received in terms of particular cost centers or areas of the hospital. The institutional base claims do not contain this level of detail; researchers must use the revenue center data files. Revenue center codes identify particular cost centers, such as the emergency department, intensive care unit, operating room, etc. Reference the list of revenue center codes in the [CCW Medicare Fee-For-Service Claims Codebook](#). We share some examples of algorithms you may find helpful in understanding the types of services the patient received ([Table 9](#), and the code example that follows).

Table 9. Algorithms for use in categorizing cost centers/areas within settings

Claim type	Institutional service setting (or cost center)	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 (OP) 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; Part B claims contain professional services)

* Researchers can link the base claim records with the associated revenue center using the (CLM_ID and CLM_LINE_NUM).

** Swing beds are rural hospital beds that providers use for either acute or SNF care, as needed. (reference, for example: <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/SNFPPS/SwingBed>).

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., the example does not include SNF or OP revenue center lines). You can modify the code example to read 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 the inpatient facility

The CCW team designed the following SAS code to subset the claims in the Part A inpatient revenue center file into smaller files specific to care settings. For this DATA step, we end up with the following inpatient setting-specific working files: 1) emergency department, 2) intensive care, 3) coronary care unit, 4) intermediate care, and 5) all other inpatient revenue centers. This example uses a one-month sample of data.

```

data    work.ed work.icu work.ccu work.imc work.oth_ip_revs;
         set RIF2018.INPATIENT_REVENUE_01;
         rename CLM_THRU_DT=REV_CNTR_DT;

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

run;

```

Once you have partitioned the data into settings, the claims files are more compact. It is simpler to query the relevant files to determine the medical services the study population has received.

3. Constructing a Longitudinal Data File

Before merging the utilization file with the denominator/sample file, you need to sort both data files by BENE_ID. CCW recommends that you 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 utilization information exists. 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, we advise against joining the person-level information (e.g., the denominator) with the claims information until you have processed the files 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 handle this sort and merging process more efficiently. This type of processing negates the need for a separate sorting step.

You may use an alternative method of managing a large data file. One method 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; or monthly files based on service date). 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 examine utilization patterns more efficiently.

You may determine that you require multiple years of claims data to determine whether an event occurred or not. It will be especially important to be efficient regarding the number of variables to retain if you work with more than one year of data.

4. Tabulating or Summarizing Utilization

You have many options for exploring and describing utilization. The Medicare FFS claims data files have numerous options for variables 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, the CCW team shares some common counting and summarizing utilization methods.

a. Tabulating Utilization of Care

As discussed above, to examine patterns of medical utilization, you may need to combine information from different claim types to ascertain a complete assessment of the caretaking place in a particular setting.

For this example, we seek to identify all ED care — both the claims from inpatient hospitals and care billed as OP (i.e., the latter occurs when the acute care facility does not immediately admit the patient). CCW performs various tabulations.

Code Example 6. Use base and revenue center records to identify care received in specific settings and calculate population-level utilization statistics

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

```

data work.op_ed;
set RIF2018.OUTPATIENT_REVENUE_01 (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 two input files which contain ED revenue lines, then merge*/

proc sort data = work.op_ed;
by BENE_ID REV_CNTR_DT;
run;

/* Code Example 5 created the next file, ed*/;

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

data work.join_ed;
set work.op_ed (in=o) work.ed (in=i);
if o then op_ed=1;
if i then ip_ed=1;
run;

```



```

/*create a small data file identifying ED use by IP and Outpatient
settings*/

proc sql;
create table work.all_ed as
select BENE_ID, max (op_ed) as op_ed_user, max (ip_ed) as ip_ed_user
from work.join_ed
group by BENE_ID;
quit;

proc freq data= work.all_ed;
tables op_ed_user ip_ed_user;
title 'Number of OP and IP Ed visits';
run;

```

We present the results of this code example in [Table 10](#).

Table 10. Number of OP and IP ED visits

SAS variable	Frequency	Percent	Cumulative frequency	Cumulative frequency	Frequency missing
op_ed_user	1,273,266	100	1,273,266	100	559,852
ip_ed_user	662,619	100	662,619	100	1,170,499

Next, we demonstrate calculating per capita use statistics (i.e., per enrolled beneficiary) and other utilization statistics — both in aggregate and for population subsets (e.g., by race).

Code Example 7. Calculate population-level utilization statistics

The CCW team designed the following SAS code to illustrate several types of population (and subpopulation) utilization statistics, requiring bringing in data from the *coverage* data file (a working file created in code example 1) and the *all_ed* working file created in the previous code example.

```

data work.bene_ed_cov;

merge work.coverage (in=a) work.all_ed;
by BENE_ID;
if a;
if ip_ed_user = 1 or op_ed_user = 1 then ed_user = 1;
bene_cnt = 1;
run;

/*the CCW team designed the macro to do various tabulations - first
overall, then by population subgroup, using the race code - from the
coverage file, which the investigator can specify */

%macro summary (demo);
proc sort data= work.bene_ed_cov; by &demo.;
run;

proc means data= work.bene_ed_cov noprint;

```

```

by &demo.;
where ffs18='Y';
output out= work.&demo._FFS (drop=_freq_ _type_) sum(op_ed_user
ip_ed_user ed_user bene_cnt)=;
run;

data work.&demo._FFS_ED_CALCULATIONS;
set work.&demo._FFS;
op_1000bene=round (op_ed_user/bene_cnt * 1000, .02);
ip_1000bene=round (ip_ed_user/bene_cnt * 1000, .02);
all_1000bene=round (ed_user/bene_cnt * 1000, .02);
op_user_pct=round (op_ed_user/bene_cnt * 100, .02);
ip_user_pct=round (ip_ed_user/bene_cnt * 100, .02);
ed_user_pct=round (ed_user/bene_cnt * 100, .02);

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= work.&demo._FFS_ED_CALCULATIONS noobs;
title "ED Visits";
run;
%mend;

/*summary statistics overall, then for specified demographic variables
such as race code*/
%summary (RTI_RACE_CD)

```

We present the results of this code example in [Table 11](#).

Table 11. ED user per beneficiary by race code

RTI_RACE_CD	op_ed_us er	ip_ed_us er	ed_user	bene_cnt	op_1000 bene	ip_1000be ne	all_1000b ene	op_user_p ct	ip_user_p ct	ed_user_p ct
0	10,898	4,363	14,574	511,713	21.30	8.52	28.48	2.12	0.86	2.84
1	889,732	465,766	1,282,732	25,405,860	35.02	18.34	50.48	3.50	1.84	5.04
2	161,672	75,510	222,981	2,841,495	56.90	26.58	78.48	5.68	2.66	7.84
3	6,875	3,950	10,260	247,273	27.80	15.98	41.50	2.78	1.60	4.14
4	19,552	12,762	30,755	833,838	23.44	15.30	36.88	2.34	1.54	3.68
5	79,186	38,309	111,032	1,837,055	43.10	20.86	60.44	4.32	2.08	6.04
6	9,807	3,380	12,443	185,740	52.80	18.20	67.00	5.28	1.82	6.70

C. Describing Utilization

1. Identifying Care for a Particular Diagnosis

Medicare requires providers to submit various kinds of information about a patient’s diagnoses or procedures obtained during a healthcare encounter. [Table 12](#) summarizes the types of information available in the CCW files:

Table 12. Diagnosis, procedure, and service codes used on Medicare claims

Type of codes	Part A	Part B institutional	Part B non-institutional
ICD-9 or ICD-10-CM diagnosis code	X	X	X
DRG	X		
Revenue center code	X	X	
ICD-9 or ICD-10-PCS procedure code	X		
Current Procedural Terminology (CPT) code	X	X	
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 ICD-9-CM; or after October 1, 2015, ICD-10-CM⁷ to classify all diagnoses, which identify the condition(s) for which a patient received 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 CMS implemented the version J data file layout; 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 DRGs, a classification system that groups similar clinical conditions and procedures. CMS uses the beneficiary’s principal diagnosis and secondary diagnoses, and any procedures furnished during the stay, 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. CMS classifies any claims that CMS received on or after that date 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 case-mix groups to pay for SNF care (resource utilization groups, or RUGs) and home health (home health resource groups, or HHRGs). The RUG for SNF claims appears in the HCPCS_CD field (when the REV_CNTR is 0022, then the first three digits of the HCPCS_CD are the RUG). The data field called the APC or Health Insurance Prospective Payment System (HIPPS) code holds the HHRG for a particular revenue center (claims variable called REV_CNTR_APC_HIPPS_CD in the revenue center file).

Revenue centers are distinct cost centers within an institutional provider that can each submit separate charges. Most hospitals have distinct revenue centers for the emergency department, intensive care unit, physical therapy,

⁷ A CCW white paper identifies some important issues related to this code conversion. Refer to [CCW Condition Categories – Impact of Conversion from ICD-9-CM to ICD-10-CM](#).

laboratory, pharmacy, imaging, and so on. Revenue center codes help identify different areas of the facility where the patient received care — and other types of care which may affect payment (such as blood transfusions or laboratory tests). For the institutional outpatient setting, more than 90% of the records have a CMS HCPCS code; in general, RHCs and FQHCs are less likely to have HCPCS codes than other records (54.6% and 43.7% of revenue center lines, respectively).

For inpatient hospital claims, providers use ICD-9 (or after October 1, 2015, ICD-10)-PCS procedure codes to describe the specific procedures; these appear on the base claim file. For Part A claims that do not involve inpatient care (e.g., for HH or SNF), outpatient claims, and non-institutional claims, providers use HCPCS codes to describe the services rendered; these codes appear in the revenue or line-item files (not the base files). An HCPCS_CD typically does not appear for revenue center records for hospitalization claims; however, CMS nearly always populates the HCPCS_CD 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).

HCPCS codes have two levels. The first level consists of codes from version 4 of the CPT-4, a numeric coding system maintained by the American Medical Association (AMA). The CPT consists of descriptive terms and codes that researchers can use 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 not included in the CPT codes; researchers can use these for non-physician services, such as ambulance services or durable medical equipment.

As stated previously (in [Chapter 2.B. Extracting Utilization Information from Claims](#)), the 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. The [CCW Medicare FFS Claims Codebook](#) includes a list of BETOS codes. These codes appear in the line item files (not the base claims; variable called BETOS_CD).

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 called REV_CNTR_APC_HIPPS_CD). For more information about APCs, refer to:

<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. You may identify services related to these conditions using the DRGs. Since the DRGs group conditions for payment, the categories may be too broad for some study objectives.

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

The claim includes additional diagnoses if there are complicating factors (e.g., diagnoses that may not be the major reason for the stay/visit but have a bearing on the patient's condition) called secondary diagnosis codes (i.e., ICD_DGNS_CD2–ICD_DGNS_CD25). Claims commonly have a condition such as diabetes listed as a secondary diagnosis — a co-occurring condition provider must manage, rather than the major cause of the hospitalization.

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

Your study objectives may determine whether the claims indicated that a beneficiary received medical care for particular conditions (diagnoses) or some routine types of care (procedures). To reduce the risk of false positives (i.e., erroneously stating that the patient received treatment for a particular condition, if the condition was not present), there may be times when we recommend requiring the presence of more than one claim with the diagnosis of interest to make this determination. In the CCW, for example, a wide variety of common and chronic conditions in the Medicare population are pre-coded and appear for each beneficiary in the MBSF — either the CCW Chronic Conditions segment or the CMS Other Chronic or Potentially Disabling Conditions (OTCC) segment. Much of the supporting documentation is available on the CCW website (e.g., chronic condition definitions, data dictionary, etc.). Reference <https://www.ccwdata.org>.

CMS designed the condition variables to examine patterns of services — which serve as a proxy for indicating the person likely is receiving treatment for the condition. The CCW algorithms for these conditions are very precise regarding the number of FFS claims, the specific types of services, and the number of years of data that a researcher must examine to make an inference regarding whether a person was likely receiving care for a particular condition. The use of this information assumes that if a provider processes a claim with a particular diagnosis code, the patient was receiving care for that particular condition. It is important that your analyses include only those diagnosis codes of interest and excludes (if desired) any screening or preventive care codes that do not represent a definitive diagnosis. Refer to the [CCW condition algorithm documentation](#) on the CCW website. These definitions are not prescriptive; you may use any claims-based algorithms appropriate for your analysis. The CCW team encourages investigators to determine whether they should make restrictions to the CCW Chronic Conditions segment or Other Chronic or Potentially Disabling Conditions segment of the MBSF for their analyses. More information is available in the [CCW Technical Guidance: Calculating Medicare Population Statistics](#) document on the CCW website.

Medicare administrative FFS claims data will only represent beneficiaries receiving FFS care. Although Medicare encounter data have recently become available — starting with 2015, CMS does not yet consider encounter service records for CCW Chronic Conditions. As a reminder, in [Code Example 1](#), we demonstrate a method for identifying beneficiaries with Medicare FFS coverage.

If your study objective is to identify patterns of utilization for a population identified as having a condition of interest, CCW urges caution since a single claim may include many diagnoses. 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; providers may hospitalize a person with diabetes for a cardiac event rather than the diabetes). It is not unusual for Medicare beneficiaries to have multiple chronic conditions, making attribution difficult (reference paper describing multiple chronic conditions in the Medicare population: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2748070/>).

3. Identifying Particular Procedures

Researchers may be interested in determining whether patients received certain procedures or interventions (i.e., the institutional revenue center records and non-institutional line items have HCPCS codes). As stated previously, CMS divides the HCPCS codes into two major categories — the Level 1 (i.e., CPT codes) and Level II HCPCS codes. The CMS website provides documentation regarding HCPCS codes (<http://www.cms.gov/MedHCPCSGenInfo/>), as do 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). In the code example that follows, the CCW team uses the MBSF_CC file segment; researchers could use the MBSF_CHRONIC or MBSF_OTCC file segments, as desired.

Code Example 8. Identify procedures for a cohort with particular condition of interest

The CCW team designed the following SAS code to identify a sample of beneficiaries with heart failure (we use the variable called CHF, from the input file for the 27 CCW Chronic Conditions segment, i.e., the MBSF_CC file). Using this BENE_ID finder file, we identify inpatient claims (for a one-month service period) and determine all ICD-PCS procedures performed.

To tabulate the frequencies for procedures, we use only the primary procedure and output the top 15 procedure codes.

```

/*Restrict to a Heart Failure sample - make finder file of these
BENE_IDS*/;
data work.hf;
    set MBSF.MBSF_CC_2018 (keep=BENE_ID CHF);
    if CHF in (1,3) then output;
proc sort; by BENE_ID;
run;

*Identify inpatient claims for benes in finder file (with HF)
this code uses one month of IP claims*/

proc sort data=RIF2018.INPATIENT_CLAIMS_01 (keep=CLM_ID BENE_ID
ICD_PRCDR_CD1-ICD_PRCDR_CD25) out=work.ip_all_01;
by BENE_ID CLM_ID;
run;

data work.hf_ip_clms;
    merge work.hf (in=s) work.ip_all_01 (in=t);
    by BENE_ID;
    if s and t;

run;

/*create SAS data sets with claim counts for non-missing Procedure codes.
We look only at the primary procedure code, although 25 Procedure Codes are
Available beginning with the Version J layout*/
proc freq data= work.hf_ip_clms noprint;
    tables ICD_PRCDR_CD1 / out=PRCDR_CODE nocol norow nopercnt;
run;

proc sort data=work.PRCDR_CODE;
    by descending count;
run;

proc print data=work.PRCDR_CODE (obs=15) noobs;
    where count > 10;
    Title 'Top 15 ICD 10 procedure codes for people identified with CHF';
run;

```

We present the results of this code example in [Table 13](#).

Table 13. ICD-10_PCS codes for CHF population

ICD_PRCDR_CD1	Count	Percent
—	244,241	—
5A1D70Z	20,779	7.97980
5A09357	11,362	4.36337
30233N1	10,268	3.94324
02HV33Z	10,194	3.91482
4A023N7	7,388	2.83723
5A1955Z	5,614	2.15596
5A1945Z	5,226	2.00695
0BH17EZ	4,392	1.68667
0DJ08ZZ	4,169	1.60103
5A09457	3,870	1.48620
027034Z	3,548	1.36255
5A2204Z	3,070	1.17898
0W993ZZ	2,750	1.05609
0DB68ZX	2,630	1.01000

4. Payments for Services

You have many options for exploring and describing payments on the claim. The Medicare claims files have several payment variables, which make it possible to determine various perspectives on payments. One approach you may use 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. In addition to these variables, CMS updated the RIF layout in 2017 to add several variables related to CMS payment reform initiatives and demonstrations. Refer to the [Medicare FFS claims data dictionary](#) on the CCW website.

Many of the payment variables are available on both the base claim and the detailed revenue or line-item records for the claims. The base claim total payment is the sum of the revenue center/line record payments.⁸ 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). In [Code Example 3](#), [Code Example 4](#), and [Code Example 5](#), we demonstrate how to identify particular services — and whether researchers can find these in the base or revenue/line files.

The Part A institutional claim include variables that indicate payment components, such as per diem amounts, or amounts providers 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). CCW describes some common methods for counting and summarizing payments below.

⁸ 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.

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 include:

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, CMS pays inpatient hospitals largely based on 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_CHRG_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. 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, researchers must add the per diem amount multiplied by the utilization days, to the CLM_PMT_AMT (refer to [Table 10](#)). The CCW team also shares some important information regarding payments; refer to section [d. Medicare payment reform and claim payment amounts](#).
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 coinsurance.⁹ 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>). The CCW uses a variety of different variables in the files to identify these components of payment (refer to and [Table 10](#)).
4. **Total amount paid to provider** — the CCW team also refers to this 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, CMS calls this amount 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 14](#) and [Table 15](#). CCW also shares some important information regarding payments not in this file; refer to section [d. Medicare payment reform and claim payment amounts](#).

[Table 14](#) and [Table 15](#) present the variable names for certain types of payment found on non-institutional and institutional Medicare claims (both the base file and line/revenue center files), respectively.

⁹ There are some services that do not require beneficiary cost-sharing, such as hospice and home health care.

Table 14. Variable names for payment fields in Medicare non-institutional claims

Payment type	Non-institutional claims carrier and DME (base file)	Non-institutional claims 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_CHRG_AMT	LINE_SBMTD_CHRG_AMT

* In Version J of the claim's 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 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, CMS pays residential care (i.e., for inpatient, SNF and hospice) a per diem amount that CMS adjusts for beneficiary case mix (severity/acuity of services needed). For inpatient services, DRG is the basis for payment; whereas for SNF, CMS pays each day of care 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 CMS pays 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, CMS already includes the amount in the total Medicare payment amount field (CLM_PMT_AMT); researcher can find the portion of the payment that is an outlier 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" (reference: <http://www.medpac.gov/-documents-/payment-basics>), and also in the Medicare Learning Network (MLN) "Payment System Fact Sheet Series" (reference the list of MLN publications at: <https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/MLN-Publications.html>).

Table 15. Variable names for payment fields in Medicare institutional claims

Payment type	Inpatient* base file	SNF* base file	HOS* base file	HH* base file	HH* revenue center	OP* base file	OP* revenue center
Medicare payment	CLM_PMT_AMT + (CLM_PASS_THRU_ PER_DIEM_AMT x CLM_UTLZTN_DAY_ CNT)	CLM_PMT_AM T	CLM_PMT_AMT	CLM_PMT_AMT	REV_CNTR_PMT_AM T_AMT	CLM_PMT_AMT	REV_CNTR_PMT_AMT_ AMT
Beneficiary payment	NCH_IP_TOT_DDCT N_AMT (or could sum NCH_BENE_IP_DDC TBL_AMT + NCH_BENE_PTA_CO INSRNC_LBLTY_AM + NCH_BENE_BLOOD_ DDCTBL_LBLTY_AM)	NCH_IP_TOT_D DCTN_AMT (or could sum NCH_BENE_IP_ DDCTBL_AMT + NCH_BENE_PTA _COINSRNC_LB LTY_AM + NCH_BENE_BL OOD_DDCTBL_ LBLTY_AM)	none	none	REV_CNT_PTNT_RSP NSBLTY_PMT	NCH_BENE_PTBL_DD CTBL_AMT + NCH_BENE_PTBL_CO INSRNC_AMT + NCH_BENE_BLOOD_ DDCTBL_LBLTY_AM	REV_CNTR_PTNT_RSPN SBLTY_PMT (or could sum REV_CNTR_CASH_DDC TBL_AMT + REV_CNTR_COINSRNC_ WGE_ADJSTD_C)
Primary payer payment	NCH_PRMRY_PYR_C LM_PD_AMT	NCH_PRMRY_P YR_CLM_PD_A MT	NCH_PRMRY_PY R_CLM_PD_AMT	NCH_PRMRY_PY R_CLM_PD_AMT	none	NCH_PRMRY_PYR_C LM_PD_AMT	REV_CNTR_1ST_MSP_P D_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_A MT	CLM_TOT_CHR G_AMT	CLM_TOT_CHRG _AMT	CLM_TOT_CHRG _AMT	REV_CNTR_TOT_CHR G_AMT	CLM_TOT_CHRG_A MT	REV_CNTR_TOT_CHRG _AMT

*Institutional claim type (on CMS 1450 forms)

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 CMS has resolved all adjustments to earlier claims and accurately recorded the final action on the claim. 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 some final action claims in the non-institutional settings (i.e., carrier or DME claims) files, may have particular lines denied. Although CMS already factored this into 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_PRCSG_IND_CD] does not indicate CMS accepted the line item; that is, keep only those where LINE_PRCSG_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. Researchers can identify these 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. 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 (reference CLM_MDCR_NON_PMT_RSN_CD in the institutional claims, or the CARR_CLM_PMT_DNL_CD in the non-institutional claims).

d. Medicare payment reform and claim payment amounts

Medicare claims use value codes and Other Applied Indicator Codes to indicate adjustments that CMS made to base payment amounts. The codes cover a wide variety of adjustments, including sequestration, the Physician Quality Reporting System (PQRS), and electronic health record incentive program. The codes (and corresponding adjustment amounts) are available in the Medicare claim value code files for Part A (inpatient, SNF, hospice, and HHA) and Part B institutional outpatient (OP) claims (refer to [Table 1](#)), and CMS captures in the Other Applied Codes/amounts variables in the Part B non-institutional claims line files (carrier and DME). These source data variables reflect amounts that CMS applied to the base Medicare payment amount which resulted in a net reduction (or increase) in Medicare spending for a given service.

Some codes reflect reductions that CMS applied to the base Medicare payment to the provider, but then included in separate lump-sum payments to that provider’s Accountable Care Organization (ACO) or other population-based payment (PBP) program. In other words, these types of “split payment” arrangements reflect a change in payment to a given provider for a specific service, but not a change in total Medicare spending. This means that CMS did not pay a portion of the actual Medicare payment amount to the provider, rather distributed it to the ACO or PBP program.

The impact of these payment adjustments means that, starting with the 2017¹⁰ claims RIFs, researchers must use the dollar amounts found in the Value Code/Line Other Applied Amount to adjust the total Medicare payment amount so

¹⁰ Medicare began making Q1 payment reductions (and applied indicator code “L” adjustments) in 2014, but these were very small in aggregate until 2017.

that it continues to reflect the total amount Medicare paid for the service. To be able to more accurately determine the amount CMS paid the provider for the service, we suggest that, starting with 2017, we distinguish between:

- the provider payment amount (i.e., what CMS paid the provider for the service), and
- the payment reduction amount (i.e., the payment reform amount that CMS did not pay to the provider, but rather paid to the ACO or PBP program) for the service.

It is important to note that the sum of these two variables equals the total Medicare payment amount. In cases where there is no value code/other applied amount, the provider payment amount equals the Medicare payment amount.

Identifying claims with provider payment adjustments. The provider payment adjustment amount is in the value code payment amount field (CLM_VAL_AMT) for Part A and Part B institutional value code files (designed for researchers to use with the Part A claim or Part B institutional outpatient claim files). When the value code (CLM_VAL_CD) = "Q1" (ACO Payment Reduction Amount [Pioneer reduction] — the actual amount of the Pioneer reduction, [effective 1/2014]) then this amount found in the CLM_VAL_AMT field reflects the ACO payment reduction amount. CMS makes all these adjustments at the claim-level (i.e., they are not related to any particular revenue center record payment amounts). The total amount Medicare paid the provider on the claim is in the CLM_PRVDR_PMT field; it is this Medicare payment amount that researchers should adjust to reflect the amount paid to an ACO or PBP program.

For Part B non-institutional claims (carrier or DME), researchers can find the payment reduction amount for each line when any of the seven Line Other Applied Indicator Code fields (LINE_OTHR_APLD_IND_CD1–LINE_OTHR_APLD_IND_CD7) = "L" (ACO Payment Adjustment Amount [Pioneer reduction] — the actual amount of the Pioneer reduction — effective 1/2014). Since the line payment amounts sum to the total claim payment amounts — you may choose to adjust either the line or claim-level payments to reflect the amount paid to an ACO or PBP program.

Adjusting claims for ACO or PBP payments. For Part A and Part B institutional services, start with Value code files, and identify all records with a value of Q1 (keep all the fields in the value code file so that we can link back to the source claims). Use these CLM_VAL_AMT and add them to the base claim CLM_PMT_AMT field; there are times there is a zero dollar or negative dollar amount in the CLM_VAL_AMT field.

For Part B carrier, use the Line file (not the base claim) — identify all "L" in the Line Other Applied Amount fields (keep the BENE_ID, CLM_ID, LINE_NUM so that we can identify the base claim). Use the LINE_OTHR_APLD_AMT{#} — whichever number is associated with the L (1–7). Since there may be multiple lines for the claim, researchers must examine each LINE_NUM for the presence of the Line Other Applied Amounts = L.

To distinguish between various payment fields, we create a new claim-level variable called actual_PMT.

actual_PMT = the total Medicare payment amount
 = provider payment + ACO payment
 = CLM_PMT_AMT + VAL_AMT (for the Q1 value code records)
 = CLM_PMT_AMT + LINE_OTHR_APLD_AMT (for L, for every line on the claim)

Code Example 9. Identify inpatient claims with the Q1 value code

Examine the inpatient value code files to identify the claim IDs that have a Q1 value code. Then ascertain the dollar amount associated with this value code and use it to adjust the inpatient claim amounts on the base claim.

The following SAS code creates a data file to capture all of the Q1 value code payments (for the month of January); then the code merges the claim value code amounts into a file with the base claim payment amounts.

```

data IP_val_q_Jan2018;
set rif2018.inpatient_value_codes_01;
if CLM_VAL_CD="Q1";
run;

proc sql;
create table IP_PMT_ADJ_Jan2018 as
select a.*, b.BENE_ID,b.CLM_ID,b.NCH_CLM_TYPE_CD, b.CLM_FROM_DT,
b.CLM_PASS_THRU_PER_DIEM_AMT,
b.CLM_UTLZTN_DAY_CNT, b.CLM_TOT_CHRG_AMT,b.CLM_PMT_AMT,

  b.CLM_PMT_AMT + a.CLM_VAL_AMT as actual_PMT

from IP_val_q_Jan2018 a left join RIF2018.inpatient_claims_01 b
  on (a.CLM_ID=b.CLM_ID) and (a.BENE_ID=b.BENE_ID);

quit;

/*an example of the payment fields is output
Remember this is for one month of IP claims*/

proc print data=ip_PMT_ADJ_Jan2018 (obs=6);
var BENE_ID CLM_ID CLM_PMT_AMT CLM_VAL_AMT actual_PMT;
run;

```

We present the results of this code example in [Table 16](#).

Table 16. Example — IP value code amount Q1 added to the claim payment amount

Obs	BENE_ID	CLM_ID	CLM_PMT_AMT	CLM_VAL_AMT	actual_PMT
1	Jones	12345	186.70	7414.69	7601.39
2	Smith	78910	0.00	4401.48	4401.48
3	Adams	55772	13469.40	46827.65	60297.05
4	Moore	345789	87.19	5184.96	5272.15
5	Edwards	852147	554.69	24319.52	24874.21
6	Parks	357159	1665.84	5194.63	6860.47

We use a slightly different approach to identifying all payment adjustments in the carrier file. Since CMS uses the Line Other Applied Amounts to adjust the line payment amount rather than the base claim amount, we must identify all line adjustment amounts first, and then aggregate this information and apply the adjustment to the base claim.

Code Example 10. Identify carrier lines with Line Other Applied Amount equal to L

Examine the carrier line numbers where any of the seven- Line Other Applied Amounts have a value of “L.” Then, ascertain the dollar amount associated with this L code. Researchers must aggregate all of the L lines for each claim to the claim-level, so that we may use this amount to adjust the carrier payment amounts on the base claim.

The following SAS code creates a data file to capture all of the L Line Other Applied Amounts (for the month of January) on all lines of the claim. The code aggregates the Line Other Applied Amounts to the claim level, and then merged into a file with the base claim payment amounts.

```

data Carr_lineOth_L_Jan2018;
set rif2018.BCARRIER_LINE_01 (keep=BENE_ID CLM_ID LINE_NUM CLM_THRU_DT
    LINE_OTHR_APLD_IND_CD1-LINE_OTHR_APLD_IND_CD7 LINE_OTHR_APLD_AMT1-
    LINE_OTHR_APLD_AMT7);

    array Lcode (7) LINE_OTHR_APLD_IND_CD1-LINE_OTHR_APLD_IND_CD7;
    array Ldollar (7) LINE_OTHR_APLD_AMT1-LINE_OTHR_APLD_AMT7;
    LAmount=0;

do i=1 to 7;
if Lcode(i) = 'L' then LAmount=Ldollar(i);
end;
drop i;

run;

/*at this stage you have the line level payment adjustment amount*/

proc print data=Carr_lineOth_L_Jan2018 (obs=50);
where LAmount NE 0;
run;

```

Table 17. Sample output, carrier line adjustment amounts

BENE_ID	CLM_ID	LINE_NUM	LINE_OTHR_APLD_IND_CD1	LINE_OTHR_APLD_IND_CD2	LINE_OTHR_APLD_AMT1	LINE_OTHR_APLD_AMT2
Jones	12345	1	H	L	0.35	1.73
	12345	2	N	L	2.94	0.29
Smith	78910	1	H	L	0.26	0.64
	78910	2	J	L	10.23	3.07
Adams	55772	1	J	L	49.94	14.98

```

/*there may be more than one line on the claim that has an LAmount
Unless you intend to analyze the payments at the line-level
We need to aggregate the line LAmounts to the base claim level
CMS call the variable Tot_L*/

```

```

proc sql;
create table Laggregate as
select a.*, sum (Lamount) as Tot_L
from Carr_lineOth_L_Jan2018 a
where Lamount NE 0
group by CLM_ID;
quit;

```

```
/*then deduplicate by claim ID so that the total L amount appears once per
claim*/
```

```
proc sql;
create table Ldedup as
select distinct a.CLM_ID, a.BENE_ID, a.Tot_L
from Laggregate a;
quit;
```

```
/*after taking the sum of all the lines, add these line other applied amounts to
the base claim payment amount*/
```

```
proc sql;
create table Carr_PMT_ADJ_Jan2018 as
select a.BENE_ID, a.CLM_ID, a.Tot_L, b.BENE_ID, b.CLM_ID, b.CLM_FROM_DT,
b.CLM_PMT_AMT,
  b.NCH_CARR_CLM_SBMTD_CHRG_AMT, b.NCH_CARR_CLM_ALOWD_AMT,

  b.CLM_PMT_AMT + a.Tot_L as actual_PMT

from Ldedup a left join RIF2018.BCARRIER_CLAIMS_01 b
on (a.CLM_ID=b.CLM_ID);

quit;
```

```
proc print data=Carr_PMT_ADJ_Jan2018 (obs=3);
  where Tot_L NE .;
  var BENE_ID CLM_ID CLM_PMT_AMT Tot_L actual_PMT;
run;
```

Table 18. Example adjustment to CLM_PMT_AMT in carrier data

BENE_ID	CLM_ID	CLM_PMT_AMT	Tot_L	Actual_PMT
Jones	12345	38.24	2.02	40.26
Smith	78910	70.41	3.71	74.12
Adams	55772	34.96	14.98	49.94

If you used the CLM_PMT_AMT without making the payment reform adjustment, a portion of the amount Medicare paid for the service would be missing (reference Tot_L column).

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

- Per capita and per user rates

Code Example 11. Medicare FFS beneficiaries who used acute inpatient care, and the associated claim payments

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 inpatient utilization and payments only for beneficiaries with full FFS coverage.

The CCW team designed the following SAS code to identify all beneficiaries with an acute inpatient claim (for a sample month, from the working file called *acute*). We merge in the information from 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). Researchers can identify members of our population with and without acute IP care (in the working file called *acute_use*).

We use PROC MEANS to summarize payments for each record, and then we include another data step to calculate metrics for the entire cohort. Finally, we output beneficiary and service user counts and associated payments.

```

/*start with the data file you created in code ex 2 for acute
hospitalization claims and sort by bene_id and clm_ID, remember this is for
one month of IP claims*/
proc sort data= work.acute;
by BENE_ID CLM_ID;
run;

/*use the coverage data file you sorted by BENE_ID in the previous code
example and merge with the sorted acute claim file */

data work.acute_cov;
merge work.coverage work.acute;
by BENE_ID ;
run;

/* Identify full FFS beneficiaries and acute inpatient care users*/
data work.acute_use;
set work.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 ffs18='Y';
run;

/*create a file consisting of a summarized dataset*/
proc means data=work.acute_use noprint;
output out=work.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 work.acute_summ;
set work.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_utlzttn_day_cnt='IP days'
    clm_pmt_amt='Total IP Payments'
    nch_bene_ip_ddctbl_amt='Total IP bene deductible amounts';
run;

/*output the results*/
proc print data=work.acute_summ noobs;
run;

```

We present the results of this code example in [Table 19](#).

Table 19. IP acute payment amounts

CLM_UTLZTN_ DAY_CNT	CLM_PMT_AMT	NCH_BENE_IP_ DDCTBL_AMT	bene_cnt	user_cnt	acute_pmt_per_user	acute_pmt_per_cap
4144598	9641574248.7	783673778.27	31862974	768936	12538.85	302.59

- Rates of events — such as acute inpatient hospital re-admissions

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. Researchers can use the code example below to determine whether one facility transferred a patient with more than one IP claim to another as part of the same continuous hospitalization, or whether the claim constitutes a separate admission or re-admission.

Code Example 12. Identify people who experience a particular event — a hospital re-admission

Several data processing steps are necessary to accomplish this task. We use two years of IP claims data – the year of interest and six months before and after the year to ensure we capture all IP stays (start and end dates). This period of time allows for the hospital stay to finish, and to determine whether the hospital re-admitted the beneficiary

The CCW team designed the following SAS code to combine all acute and CAH inpatient hospitalizations for a beneficiary into distinct “stays.”

We identify all beneficiaries with an acute inpatient or CAH claim, and then sort the claims in date of service order to identify and group the claims into inpatient stays (this is the working file called *IP_claims*). We identify transfers to different facilities, so that they are not confused with re-admissions (i.e., where there is a break in the acute stay, during which time a provider may provide a lower level of care, followed by another inpatient admission).

We sort the claims by date, and then group the claims into stays (in the working file called *Stay_ID*).

```

/*example starts with 2 years of claims, all of 2018 as well as 6 months before
and after the calendar year*/

```

```

data IP_claims;
    set RIF2017.INPATIENT_CLAIMS_07-RIF2017.INPATIENT_CLAIMS_12
        RIF2018.INPATIENT_CLAIMS_01-RIF2018.INPATIENT_CLAIMS_12

```

```

RIF2019.INPATIENT_CLAIMS_01-RIF2019.INPATIENT_CLAIMS_06;
by BENE_ID;

if (substr (PRVDR_NUM,3,1.) in ("0") or substr (PRVDR_NUM,3,2.) in ("13"));

if PTNT_DSCHRG_STUS_CD = '02' then
    transfer=1;
else transfer=0;

if PTNT_DSCHRG_STUS_CD = '07' then
    AMA=1;
else AMA=0;

if PTNT_DSCHRG_STUS_CD in ("20","40","41","42") then
    died=1;
else died=0;
run;

/*sorting claims and grouping them into stays*/

proc sort data=IP_claims; BY BENE_ID CLM_FROM_DT descending transfer clm_id;
run;

data work.Stay_ID (drop = temp_dt);
set work.IP_claims;
BY BENE_ID CLM_FROM_DT descending transfer CLM_ID;
retain stay_grp_new stay_beg stay_end temp_dt;
format stay_beg stay_end mmddyy10.;

if first.BENE_ID then
    do;
        stay_grp_new = 1;
        stay_beg = CLM_FROM_DT;
        stay_end = CLM_THRU_DT;
        temp_dt = stay_end;
    end;
else
    do;
        if CLM_FROM_DT <= temp_dt then
            do;
                stay_beg = min(stay_beg,CLM_FROM_DT);
                stay_end = max(CLM_THRU_DT,stay_end);
                temp_dt = stay_end;
            end;
        else
            do;
                stay_grp_new+1;
                stay_beg=CLM_FROM_DT;
                stay_end=CLM_THRU_DT;
                temp_dt=stay_end;
            end;
        end;
    end;
end;

```

run;

We merge the stays data (the working fall called **Stay_ID**) with beneficiary coverage information from the MBSF_ABCD files. Then we use a 30-day interval between IP stays so that we are calculating 30-day re-admission rates.

Finally, we calculate re-admission rates.

```

DATA IP_STAYS_EXCL;
    MERGE work.Stay_ID (IN=a WHERE=(YEAR(STAY_END)=2018) ) MBSF.MBSF_ABCD_2018
    (KEEP=BENE_ID BENE_DEATH_DT MDCR_ENTLMT_BUYIN_IND_ : HMO_IND_ :)
    MBSF.MBSF_ABCD_2019(KEEP=BENE_ID BENE_DEATH_DT
MDCR_ENTLMT_BUYIN_IND_01 HMO_IND_01
    RENAME=(MDCR_ENTLMT_BUYIN_IND_01=MDCR_ENTLMT_BUYIN_IND_13 HMO_IND_01=
HMO_IND_13 ));
    BY BENE_ID;

    IF a;
    ARRAY ENTITLE [*] MDCR_ENTLMT_BUYIN_IND_01 - MDCR_ENTLMT_BUYIN_IND_13;
    ARRAY HMO [*] HMO_IND_01 - HMO_IND_13;
    ARRAY FFS [*] FFS_01 - FFS_13;

    DO I = 1 TO DIM(FFS);
        IF ENTITLE (I) IN ('3','C') AND HMO (I) IN ('0','4') THEN
            FFS (I) = 1;
        ELSE FFS(I) = 0;
    END;

    IF FFS(MONTH(stay_end))=1 AND (FFS(MONTH(stay_end)+1) =1 OR stay_end<
BENE_DEATH_DT< stay_end+30) THEN
        OUTPUT;
RUN;

/*calculating 30-day re-admission interval and associated stay dates*/
proc sql;
    Create table readmits as
    Select a.BENE_ID,
          a.stay_beg, a.stay_end, a.CLM_ID,
          b.stay_beg as readmit_beg_dt,
          b.stay_end as readmit_end_dt,
          b.PRVDR_NUM as readmit_prvdr_num,
          b.CLM_ID as readmit_clm_id,
          1 as readmit
    From ip_stays_excl as a, IP_stays as b
    where a.BENE_ID=b.BENE_ID and
a.stay_end<b.stay_beg<=a.stay_end+30;
Quit;

proc sort data=readmits;
    by BENE_ID stay_beg stay_end CLM_ID readmit_beg_dt readmit_end_dt
readmit_clm_id;

```

```

run;

Data readmits_nodup;
  set readmits;
  By BENE_ID stay_beg stay_end CLM_ID;
  If first.CLM_ID;
Run;

/* sort file so you can merge*/
proc sort data=ip_stays_excl;
by BENE_ID stay_beg stay_end CLM_ID;
run;

/*Calculating re-admissions*/

Data re-admissions;
  Merge ip_stays_excl(in=a) readmits_nodup(keep=BENE_ID stay_beg stay_end
CLM_ID readmit:);
  By BENE_ID stay_beg stay_end CLM_ID;
  IF a;
  Admit=1;
  If readmit =. then
    readmit=0;
Run;

/* summarizing overall rates */

Proc sql;
  Create table rates as
    select  Count(distinct BENE_ID) as unique_benes,
    Sum(admit) as total_admissions,
    Divide(sum(readmit),sum(admit)) as readmit_rate format=percent8.2
  From re-admissions
  Quit;

proc means data=rates ;
var readmit_Rate;
output out=rates mean=readmit_rate / NOINHERIT;
run;

```

Overall re-admission rates example.

Table 20. All cause re-admission rate example, using readmit_rate

N	Mean	Std Dev	Minimum	Maximum
1	0.1740178	-	0.1740178	0.1740178

This analysis determined the percentage of IP stays with a 30-day all-cause re-admission. With minor modifications to the code example, you may output provider-specific re-admission rates or rates using various beneficiary demographic characteristics.

Chapter 3. Further Assistance with CCW Data

Researchers interested in working with CCW data should contact the Research Data Assistance Center (ResDAC). They offer free assistance to researchers using Medicare data for research. The ResDAC website provides links to descriptions of the CMS data available, request procedures, supporting documentation, such as record layouts and SAS input statements, workshops on how to use Medicare data, and other helpful resources. Visit the ResDAC website at <http://www.resdac.org> for additional information.

ResDAC is a CMS contractor and researchers should first submit requests to ResDAC for assistance in the application, obtaining, or using the CCW data. Researchers can reach ResDAC by phone at 1-888-973-7322, email at resdac@umn.edu, or online at <http://www.resdac.org>.

If a ResDAC technical advisor cannot answer your question, the technical advisor will direct the researcher to the appropriate person. If you require additional CMS data (data not available from the CCW) to meet research objectives, or the researcher has any questions about other data sources, the researcher should first visit the ResDAC website.

The CCW Help Desk provides assistance between 8:00 am to 5:00 pm ET, Monday through Friday. Contact the CCW Help Desk at ccwhelp@ccwdata.org or 1-866-766-1915.

Appendix A — List of Acronyms and Abbreviations

Acronym	Definition
ACO	Accountable Care Organization
AMA	American Medical Association
APC	Ambulatory Payment Classification
ASC	Ambulatory Surgery Center
BETOS	Berenson-Eggers Type of Service
CAH	Critical Access Hospital
CCW	Chronic Conditions Warehouse
CHF	(Congestive) Heart Failure
CME	CMS Common Medicare Environment
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
ED	Emergency Department
ESRD	End-Stage Renal Disease
FFS	Fee-for-Service
FQHC	Federally Qualified Health Center
HCPCS	Healthcare Common Procedure Coding System
HH	Home Health Agency
HHRG	Home Health Resource Group
HMO	Health Maintenance Organization
ICD-9-CM	International Statistical Classification of Diseases and Related Health Problems (9 th Revision)
ICD-10-CM	International Statistical Classification of Diseases (10 th Revision; in effect starting 10/1/2015)
ICD-10-PCS	International Statistical Classification of Diseases – Procedure Coding System (10 th Revision; in effect starting 10/1/2015)
ICU	Intensive Care Unit
ID	Identification Number
IMC	Intermediate Care Unit
IME	Indirect Medical Education
IP	Inpatient Hospital
IPF	Inpatient Psychiatric Facility
LIS	Low-Income Subsidy
LTCH	Long-Term Care Hospital
MA	Medicare Advantage
MAC	Medicare Administration Contractor
MA-PD	Medicare Advantage – Prescription Drug Plan
MBSF	Master Beneficiary Summary File
MLN	Medicare Learning Network
NDC	National Drug Code
NPI	National Provider Identifier

Acronym	Definition
OP	Hospital or Institutional Outpatient
PBP	Population-Based Payment
PCS	Procedure coding system
PDE	Prescription Drug Event
PDP	Prescription Drug Plan
ResDAC	Research Data Assistance Center
RHC	Rural Health Clinic
RIF	Research Identifiable File
RUG	Resource Utilization Group
SDA	Self-Decrypting Archive
SNF	Skilled Nursing Facility
VRDC	Virtual Research Data Center