Chronic Conditions Warehouse

Your source for national CMS Medicare and Medicaid research data

Chronic Conditions Warehouse

Medicare Administrative Data User Guide

SEPTEMBER 2022 | VERSION 3.9

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	3.9
April 2022	K. Schneider R. Van Gilder A. Sisco	Revised Chapter 2.D. Chronic Conditions segment to reflect the addition of the 30 CCW Chronic Conditions file. Corrected "STAYS" definition in the MBSF Cost and Use segment, Chapter 2.E	3.8
February 2022	K. Schneider D. Happe	Transferred document to the new template; added caution regarding payment reform, refreshed content to reflect current source data and CCW RIF products	3.7
June 2019	K. Schneider	Revision log created; included Sickle Cell Disease (SCD)	3.6

Table of Contents

List of Tables

Overview

One of the goals of Section 723 of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003 was to make Medicare data readily available to researchers who are studying chronic illness in the Medicare population. To support this effort, Centers for Medicare & Medicaid Services (CMS) established the Chronic Conditions Warehouse (CCW).

The CMS CCW is a research database designed to make Medicare, Medicaid, Assessments, and Part D Prescription Drug Event data more readily available to support research designed to improve the quality of care and reduce costs and utilization. The CCW provides researchers with Medicare and Medicaid beneficiary, claims, and assessment data linked by beneficiary allowing researchers to analyze information across the continuum of care. The CCW system currently contains data from Medicare, Medicaid, and all assessments regardless of payer, from 1999 forward. CMS makes the Medicaid data available to researchers as the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF) from 2014 forward; previously, Medicaid data were available as Medicaid Analytic eXtract (MAX) data files. Table 1 provides a list of CMS data available and the type of files associated with each file type.

Table 1. Types of CMS data files available to researchers

CMS data	Types of files	
Medicare	• Master Beneficiary Summary File (1999 forward; previously known as the Denominator File) • Plan characteristics file (2007 forward; initially delivered as the Part D plan characteristics files) • Institutional and non-institutional fee-for-service (FFS) claims (1999 forward) • Medicare Advantage (MA) encounter data (2015 forward) • Part D drug event (PDE) files (2006 forward) • Part D characteristics files (2006 forward)	
Medicaid	TAF files (2014 forward) • Annual demographic and eligibility (DE) • Inpatient hospital (IP) claims • Long-term care (LT) claims • Prescription drug (RX) claims • Other services (OT) claims • Annual provider (APR) • Annual managed care plan (APL)	MAX files (1999–2015) • Person summary (PS) • Medicaid Enrollee Supplemental File (MESF) • Inpatient hospital (IP) • Long-term care (LT) • Prescription drug (RX) • Other services (OT)
Assessment	• Minimum Data Set (MDS) (1999 forward) • Outcome and Assessment Information Set (OASIS) (1999 forward) • Swing bed assessments (1999 forward) • Inpatient rehabilitation facility/Patient assessment instrument (IRF/PAI) (1999 forward)	

The CCW data files are available upon request for a random 5% sample or specific chronic condition cohorts. Researchers may also request data for other cohort(s) of interest. CMS and its contractors have defined the specific chronic condition categories. The supporting documentation is available on the CCW website (e.g., chronic condition definitions, standard data dictionary files, etc.). Reference https://www.ccwdata.org.

The intended use of the CCW data is to identify areas for improving the quality of care provided to chronically ill Medicare beneficiaries, reduce program spending, and make current Medicare data more readily available to researchers studying chronic illness in the Medicare population. By predefining the chronic conditions, data extraction from the CCW is very efficient, allowing the CCW team to fulfill data requests quickly and cost-efficiently.

Academic researchers and certain government agencies may request Research Identifiable Files (RIFs) with approval under a Data Use Agreement (DUA). The CCW Medicare data contain identifiable information and are subject to the Privacy Act and other Federal government rules and regulations (reference the Research Data Assistance Center [ResDAC] website for details on requesting Medicare data http://www.resdac.org/).

CCW removes the Medicare beneficiary health insurance claim (HIC)[1] numbers from the data files delivered to researchers (unless otherwise specified/approved in the DUA). CCW adds a unique CCW beneficiary identifier (variable called the BENE_ID) in each data file delivered as part of the output package (reference Chapter 7 for details), thus allowing linkage of an individual's data across data sources/types. CCW provides a separate file for those requests requiring beneficiary identifiable data. If a researcher needs to obtain the HIC (or MBI) to link to outside data sources or extract claims not part of the CCW database, then the researcher will need to submit justification for this information in the study protocol and request identifiable variables.

The unique CCW beneficiary identifier field is specific to the CCW and does not apply to any other identification system or data source. CCW encrypts this identifier and all data files before delivering the data files to researchers (reference Encryption Information in Chapter 7 for details). CCW employs a different encryption key for each research request when encrypting the beneficiary identifier field and the data files.

This guide provides users with information that may be helpful in understanding and working with the CCW Medicare data. Appendix A — List of Acronyms, lists abbreviations used in this document. Throughout this document, when we identify a particular data variable by name, we will often identify the specific SAS name, appearing in all capitals.

---

[1] CMS began using a new Medicare beneficiary identifier (MBI) in place of the HIC, starting in 2018.

Chapter 1. CCW Medicare Population

The CCW database contains CMS Medicare administrative enrollment and claims data for all Medicare beneficiaries with coverage during a specified period. The CCW data are available for services beginning January 1, 1999, through the most current year of data available.

CCW contains enrollment data for 100% of Medicare beneficiaries and FFS claims from 1999 forward. In addition, managed care encounter data is available for people enrolled in MA plans in 2015 (the *Medicare Encounter Records Codebook* is available on the CCW website). CCW also contains all Part D events (prescription drug fill records) from the inception of the Part D benefit in 2006, regardless of whether the beneficiary enrolled in a managed care plan or a stand-alone prescription drug plan during that time.

This rich data repository allows for tremendous flexibility in defining cohorts or populations of interest. CCW offers various sampling options such as: a random 5% sample, a CCW condition cohort, or an investigator-defined sample. If approved by CMS, researchers may select a sample using finder files from populations they previously studied. The CCW team encourages a random 5% sample or a smaller population subset due to the substantial volume of Medicare data.

The 5% random sample consists of people who had a Medicare HIC number equal to the Claim Account Number (CAN) plus Beneficiary Identity Code (BIC) (HIC=CAN+BIC) where the last two digits of the CAN are in the set {05, 20, 45, 70, 95}; an "enhanced" 5% sample consists of those who were ever part of the 5% sample at any time, beginning with January 1, 1999, forward. The Social Security Administration (SSA) assigns the HIC number when a person becomes eligible for benefits. However, the number may change over time if a person's reason for entitlement changes. The CAN number is the policy number of the wage earner eligible for benefits — which means the SSA joins the spouse's CANs. A marriage may cause a change in the HIC due to entitlement for benefits through the spouse. CMS designs two variables to make it easy to identify the random 5% sample for a particular year (variable called SAMPLE_GROUP), and also to follow the 5% sample longitudinally even when a HIC change causes the person to drop out of the 5% at a later point in time (variable called ENHANCED_FIVE_PERCENT_FLAG). Note that researchers can also use the SAMPLE_GROUP variable to identify a 1% or 20% sample.

Chapter 2. Medicare Enrollment and Beneficiary-Level Data

The CCW has always disseminated files that include data regarding Medicare enrollment. Currently, the CCW enrollment data file is the Master Beneficiary Summary File (MBSF) that uses the CMS Common Medicare Environment (CME) database as its source.[2]

The MBSF contains many enrollment and other person-level variables contained in file "segments." These segments are separate components of the file researchers may request. The data dictionaries on the CCW website describe the variables contained in the 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 are 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.

The MBSF — Part A/B/C/D segment, also known as the base beneficiary summary file, consists of variables that identify monthly Medicare Part A, B, C, and D enrollment status and other key demographic and coverage variables. Table 2 through Table 5 illustrate some key demographic, enrollment, and coverage variables. The CMS CME database is the original source for all fields regarding beneficiary demographics or enrollment in the MBSF. The CCW database updates information from CME each month and allows a full year of additions and updates after the end of the calendar year before finalizing the MBSF. For example, the 2018 MBSF covers the calendar year 2018; the CCW team finalized the data in December 2019 and then extracted it to populate the 2018 MBSF.

The additional segments of MBSF are: 1) CCW Chronic Conditions, 2) CMS Other Chronic or Potentially Disabling Conditions (OTCC), 3) Cost and Use, and 4) National Death Index (NDI). Below are descriptions for these four optional segments.

A. Medicare Part A, B, C, and D Enrollment Segment

Essential information for most study denominators appears in the Base A/B/C/D segment of the MBSF. It consists largely of beneficiary demographic and Medicare Part A, B, C, and D coverage information. Table 2 below identifies some key demographic variables, and Table 3 identifies Medicare enrollment and coverage variables.

---

[2] A CCW White Paper, *Medicare Enrollment: Impact of Conversion from EDB to CME* contains a description of the rationale and impact of the data conversion from the CMS Enrollment Database (EDB) to the CME for producing the MBSF.

Table 2. Examples of demographic information in MBSF

Variable name (long)	Variable description	Brief definition
STATE_CODE	SSA state code	Beneficiary SSA state code
COUNTY_CD	SSA county code	Beneficiary SSA county code
STATE_CNTY_FIPS_CD_{MM}	Monthly state and county FIPS code (01–12)	Beneficiary FIPS state and county code (12 monthly fields)
ZIP_CD	ZIP code of residence	Beneficiary ZIP code
AGE_AT_END_REF_YR	Age at the end of the reference year	Age at the end of the reference year
BENE_BIRTH_DT	Date of birth (DOB)	DOB from the SSA
BENE_DEATH_DT	Date of death (DOD)	DOD from SSA
SEX_IDENT_CD	Sex	Beneficiary sex
BENE_RACE_CD	Beneficiary race code	Beneficiary race code
RTI_RACE_CD	Research Triangle Institute (RTI) race code	RTI race code

Beneficiary state, county, and ZIP code — the beneficiary's geographic information comes from the beneficiary's mailing address used to deliver benefits to the beneficiary (such as Social Security) or for other purposes (such as Medicare premium billing). Therefore, it may not reflect the location where the beneficiary resides.

RTI race code — this variation on the race code which the SSA has historically used, classifies an additional group of beneficiaries as Hispanics or Asians. Using this enhanced classification algorithm, Hispanics and Asians include beneficiaries who either have an SSA race code which = Hispanic or Asian, or a first name/last name which RTI has determined is likely to be Hispanic or Asian in origin (**NOTE:** the contractor who created this field is RTI International — a trade name of Research Triangle Institute).

Table 3. Examples of Medicare enrollment information in MBSF

Variable name (long)	Variable description	Brief definition
COVSTART	Medicare coverage start date	The first historical date for Medicare coverage
ENTLMT_RSN_ORIG	Original reason for entitlement code	Original reason for entitlement to Medicare
ENTLMT_RSN_CURR	Current reason for entitlement code	Current (year) reason for entitlement to Medicare
ESRD_IND	End-stage renal disease (ESRD) indicator	Beneficiary entitled to ESRD benefits
MDCR_STATUS_CODE_{MM}	Monthly Medicare status code (01–12)	Current (monthly) reason for entitlement to Medicare (12 monthly fields)
MDCR_ENTLMT_BUYIN_IND_{MM}	Monthly Medicare entitlement/buy-in indicator (01–12)	Monthly indicator of entitlement to Medicare A and B, as well as whether a state-paid Medicare premiums (12 monthly fields)
DUAL_STUS_CD_{MM}	Monthly State reported dual-eligible status code (01–12)	Monthly indicator of dual eligibility status, where both Medicaid and Medicare (12 monthly occurrences) enrolls the beneficiary
BENE_STATE_BUYIN_TOT_MONS	State buy-in coverage months count	Number of months of state buy-in for the beneficiary
DUAL_ELGBL_MONS	Months of Dual Eligibility	Number of months where the beneficiary had dual eligibility (DUAL_STUC_CD_MM not equal to '00')

B. Medicare Part C

Beneficiaries may elect to receive original fee-for-service (FFS) Medicare or, as an alternative, enroll in Medicare Part C MA. Medicare Advantage Organizations (MAOs) are private managed care plans, such as health maintenance organizations (HMOs), preferred provider organizations (PPOs), private fee-for-service plans (PFFS), and special needs plans (SNPs), which provide Medicare Part A and Part B services. MA-PD plans are MA plans that include the Medicare Part D prescription drug benefit. Table 4 and Table 5 display Part C enrollment variables.

Table 4. Part C enrollment variable descriptions

Variable name (long)	Variable description	Brief definition
HMO_IND_{MM}	Monthly HMO indicator (01–12)	Monthly indicator of whether the beneficiary enrolled in a managed care plan, currently referred to as MA premiums (12 monthly fields)
BENE_HMO_CVRAGE_TOT_MONS	HMO coverage months count	Number of months where the beneficiary had MA (HMO) coverage
PTC_CNTRCT_ID_{MM}	Monthly Contract ID (01–12)	CMS assigns the unique number to each contract that a Part C plan has with CMS (12 monthly occurrences). The first character of the contract ID is a letter representing the type of plan, e.g., managed care organizations, regional PPO, regional PPO, prescription drug plan (PDP), not Part D enrolled, employer direct plan (beginning in 2007) (12 monthly occurrences).
PTC_PBP_ID_{MM}	Monthly Plan benefit package ID (01–12)	CMS assigns the unique number to identify a specific Part C plan benefit package within a contract (12 monthly occurrences).
PTC_PLAN_TYPE_CD_{MM}	Monthly Plan type code	Monthly Part C plan type code (12 monthly occurrences).

Starting with the 2015 benefit year, the CCW team has made a plan characteristics suite of six files per year that contains detailed information regarding the Part C and Part D plans selected by beneficiaries. Additional details regarding plan characteristics are in Chapter 5. For 2015, Medicare encounter records are also available on the CCW website. Further details are in the *CCW Medicare Encounter Data User Guide*.

C. Medicare Part D

The Medicare prescription drug benefit, a voluntary benefit offered through the Medicare Part D program, is an optional drug benefit beneficiaries may purchase through private plans. Coverage of prescription drugs through Medicare Part D began in 2006. The Part D enrollment data are available in the MBSF. Table 5 displays Part D variables.

Table 5. Part D enrollment variable descriptions

Variable name (long)	Variable description	Brief definition
PTD_CNTRCT_ID_{MM}	Monthly Contract ID* (01–12)	CMS assigns the unique number to each contract that a Part D plan has with CMS (12 monthly occurrences). The first character of the contract ID is a letter representing the type of plan, e.g., Managed Care Organizations, Regional PPO, PDP, not Part D enrolled, employer direct plan (beginning in 2007)
PTD_PBP _ID_{MM}	Monthly Plan benefit package ID* (01–12)	CMS assigns the unique number to identify a specific Part D plan benefit package within a contract (12 monthly occurrences)
PTD_SGMT_ID_{MM}	Monthly Segment ID* (01–12)	CMS assigns the segment number to identify a segment or subdivision of a Part D plan benefit package within a contract (12 monthly occurrences)
CST_SHR_GRP_CD_{MM}	Monthly Cost-share group (01–12)	Monthly indicator of beneficiary liability of cost-sharing. Includes values to indicate whether CMS deemed the beneficiary eligible or whether there was a subsidy (12 monthly occurrences)
RDS_IND_{MM}	Monthly Retiree drug subsidy indicators (01–12)	Monthly indicator of whether CMS should subsidize the employer for a retired beneficiary (12 monthly occurrences)
PTD_PLAN_CVRG_MONS	Part D plan coverage months	Total number of months of Part D plan coverage
RDS_CVRG_MONS	Retiree drug subsidy months	Total number of months CMS entitles the employer to retiree drug subsidy for the beneficiary

* Before the 2013 data release, the CMS privacy rules required the CCW team to encrypt these sensitive data fields. Currently, CMS does not require the CCW team to encrypt the plan identifiers.

D. Conditions Segments

There are two types of conditions segments: 1) the CCW Chronic Conditions, and 2) the CMS OTCC.

The algorithms examine service patterns in claims data, which serve as a proxy indicating that a beneficiary is likely receiving treatment for the condition.

The MBSF Conditions segments contain variables that indicate the presence of treatment for common or chronic conditions using claims-based algorithms (as a proxy for evidence of the presence of a condition). This information is present for all beneficiaries included in the requested sample, regardless of whether the person has any of the conditions.

These files include the yearly indicator variables and first "ever" dates for each chronic condition described in Chapter 3; the MBSF file for the 27 CCW Chronic Conditions also includes a mid-year variable for each condition. These three variables appear for each of the 27 CCW Chronic Conditions, as illustrated with the acute myocardial infarction (AMI) example in Table 6 below. Table 7 presents the two types of variables available for each of the 30 CCW Chronic Conditions.

Table 6. Examples of variables for the 27 CCW Chronic Conditions

Variable name (long)	Brief description
AMI	Acute myocardial infarction end-of-year indicator
AMI_MID	Acute myocardial infarction mid-year indicator
AMI_EVER	First occurrence of acute myocardial infarction

Table 7. Examples of variables for the 30 CCW Chronic Conditions

Variable name (long)	Brief description
PNEUMO	All cause pneumonia end-of-year indicator
PNEUMO_EVER	First occurrence of all cause pneumonia

Additional details regarding the condition variables appear in Chapter 3 of this document.

1. CCW Chronic Conditions

The CCW contains two versions of the Chronic Conditions: 30 CCW Chronic Conditions (2017 forward) and 27 CCW Chronic Conditions (1999–2020). CMS developed the 27 CCW Chronic Condition variables using a multi-stage process. Initially, CMS used national data sources to identify candidate conditions which they could define using claims-based algorithms. Next, CMS conducted extensive literature reviews to gather diagnosis code sets for each candidate condition. Finally, CMS engaged other federal agencies in a series of conversations to vet the proposed category definitions. In 2020, CMS contracted an expert panel to validate the algorithms following the change from ICD-9 to ICD-10-CM. CMS also asked the expert panel to refine these algorithms and identify additional conditions to add to the CCW, resulting in the 30 CCW Chronic Condition algorithms.

2. CMS OTCC

The CMS OTCC segment of the MBSF contains 15 mental health and substance use conditions, 15 developmental disorder and disability-related conditions, and 10 other chronic physical and behavioral health conditions — all developed by CMS specifically to enhance the research of the Medicare-Medicaid dually enrolled population. These variables are similar in structure to the variables in the CCW Chronic Conditions segment; there is a yearly indicator for the variables and a first "ever" date. The variable naming convention includes *_MEDICARE to distinguish these variables from conditions in the CCW Conditions segment (Table 8). Chapter 3 presents additional details.

Table 8. Examples of other condition variables

Variable name (long)	Brief description
AUTISM_MEDICARE	Autism spectrum disorders end-of-year indicator
AUTISM_MEDICARE_EVER	Autism spectrum disorders first-ever occurrence date

E. Cost and Use Segment

This segment of the MBSF contains summarized patient-level utilization information by care setting for the calendar year of the data file. CMS uses the last date on the claim, referred to as the CLM_THRU_DT, to partition the claims into calendar year files. This MBSF segment also includes Medicare and beneficiary payment information overall and by setting. These cost and use summaries use Medicare Part A and Part B fee-for-service claims; therefore, there is no opportunity to determine whether the managed care enrollees have received services, nor can we calculate the

annual costs of these services. The exception is for Part D events — where we have both the cost and use information regardless of whether the data shows the beneficiary enrolled in a stand-alone PDP or a MA plan with prescription drug coverage (MA-PD).

Variables are null (missing) if the beneficiary did not use a particular type of service during the year. For example, those without an acute hospitalization will have missing values for all variables associated with the acute care setting.

NOTE: Starting in 2014, some payments may not reflect the amount paid to the provider since Medicare uses value codes and other applied indicator codes to indicate adjustments that CMS made to base payment amounts for various CMS payment incentive programs. CMS may have applied reductions to the base Medicare payment to the provider, but CMS then included them in separate lump-sum payments to that provider's Accountable Care Organization (ACO) or other population-based payment (PBP) program. This means that CMS did not pay a portion of the actual Medicare payment amount to the provider; rather they distributed it to the ACO or PBP program. CCW provides additional details, and technical guidance to identify these payment adjustments using the FFS claims data. Reference the *Technical Guidance: Getting Started with CMS Medicare Administrative Research Files* CCW website.

Settings. To better illustrate the cost and use summaries, Table 9, Table 12, Table 13, and Table 14 describe the methodology for dividing claims into settings. We classify Medicare services using four major setting categories:

- **Part A institutional claims** — claims from institutions or facilities which the Medicare Part A benefit generally covers. This includes claims from the inpatient, SNF, HH, and hospice data files;
- **Part B institutional claims** — claims from institutions such as hospital outpatient facilities, which the Medicare Part B benefit generally covers. This includes claims from the hospital outpatient data file;
- **Part B non-institutional claims** — claims from non-institutional providers such as providers/practitioners and durable medical equipment or prosthetic/orthotics providers. This includes claims from the carrier and DME data files; and
- **Part D event data** — a final transactional record for all Medicare Part D prescription drug events.

Table 9. Algorithms used in categorizing Part A institutional claims into settings

Part A values	Label	Algorithm	Medicare payment variable name
ACUTE	Inpatient acute care hospital (and CAH)	NCH_CLM_TYPE_CD= 60,61 and PRVDR_NUM has 3rd digit =0 or 3rd and 4th digits = 13	ACUTE_MDCR_PMT*, **
OIP	Other inpatient hospital (children's, cancer, IPF, IRF, LTC hospital)	all other NCH_CLM_TYPE_CD= 60,61 (where 3rd digit of PRVDR_NUM is not 0 and 3rd and 4th digits are not 13)	OIP_MDCR_PMT*
SNF	Skilled nursing facility	NCH_CLM_TYPE_CD= 20, 30	SNF_MDCR_PMT
HH	Home health	NCH_CLM_TYPE_CD= 10	HH_MDCR_PMT
HOS	Hospice	NCH_CLM_TYPE_CD= 50	HOS_MDCR_PMT

  * Note that for these hospital settings, this variable may not completely represent the total amount paid by Medicare.
 ** For inpatient stays — the algorithms consider all claims (in the ACUTE_STAYS variable). However, for the count of covered days (ACUTE_COV_DAYS) and payments (ACUTE_MDCR_PMT), the algorithms include only claims that ended during the calendar year where the payment amount was >= $0.

To calculate total Medicare payments, you must add the pass-through-per-diem payments to this variable; CCW provides guidance for constructing this algorithm in the *CCW Technical Guidance: Getting Started with CMS Medicare Administrative Research Files* document on the CCW website.

Cost or Payments. Three types of payment variables are present in the MBSF Cost and Utilization (MBSF-CU) segment:

1. **Medicare payments** —the annual amount Medicare paid for services on behalf of the beneficiary. CMS aggregates the payments for each beneficiary for the year (i.e., the sum of all Part A CLM_PMT_AMT[3] and Part B LINE_NCH_PMT_AMT and Part D CPP_AMT and LIS_AMT). For hospital settings, you must also add in the pass-through-per-diem payments to obtain the total Medicare payments (i.e., total Medicare payments for acute inpatient hospitalizations = ACUTE_MDCR_PMT + ACUTE_PERDIEM_PMT). Additional information regarding Medicare payments is available in the *CCW Technical Guidance: Getting Started with CMS Medicare Administrative Research Files* document on the CCW website.

2. **Beneficiary payments** — represent the aggregated beneficiary liability for cost-sharing, including coinsurance and deductible payments for the year. This includes the annual sum of all claims for one of the following (depending on the setting):
   - Part A — NCH_BENE_IP_DDCTBL_AMT and NCH_BENE_PTA_COINSRNC_LBLTY_AMT, all
   - Part B institutional — REV_CNTR_CASH_DDCTBLE_AMT and REV_CNTR_WAGE_ADJSTD_COINS_AMT, and all
   - Part B non-institutional — B_DED, B_COINS, and
   - Part D — PTPAYAMT, OTHTROOP, LICS_AMT, and PLRO_AMT

3. **Primary payer other than Medicare** — represents the amount a primary payer (e.g., the VA or TRICARE) paid for services on behalf of the beneficiary. CMS aggregates the payments for each beneficiary for the year (i.e., the sum of all Part A NCH_PRMRY_CLM_PD_AMT and all Part B CARR_CLM_PRMRY_PYR_PD_AMT).

These three types of payments are present for almost every service type. For example, researchers will find the total annual Medicare payments for a beneficiary for SNF care in the variable called SNF_MDCR_PMT, find the corresponding beneficiary payments in the SNF_BENE_PMT variable, and find the other primary payer amounts in the SNF_PRMRY_PMT variable. Two service types, the HOS and HH, do not have a beneficiary payment variable since the coinsurance and deductible amounts for these two settings were $0. Table 10 illustrates the three types of payment variables that appear for nearly every service setting.

Table 10. Part A service settings and corresponding payment variables

Service type	Medicare payment	Beneficiary payment	Primary payer amount
Inpatient acute care hospital (and CAH)	ACUTE_MDCR_PMT + ACUTE_PERDEIEM_PMT	ACUTE_BENE_PMT	ACUTE_PRMRY_PMT
Other inpatient hospital (IPF, cancer hospital, children's hospital)	OIP_MDCR_PMT + OIP_PERDIEM_PMT	OIP_BENE_PMT	OIP_PRMRY_PMT
Skilled nursing facility	SNF_MDCR_PMT	SNF_BENE_PMT	SNF_PRMRY_PMT
Home health	HH_MDCR_PMT		HH_PRMRY_PMT
Hospice	HOS_MDCR_PMT		HOS_PRMRY_PMT

The same three types of payment variables are present for the Part B institutional outpatient (variables called HOP_MDCR_PMT, HOP_BENE_PMT, and HOP_PRMRY_PMT), and for each of the 11 Part B non-institutional outpatient settings that appear in Table 13 (all such variables follow the naming convention *_MDCR_PMT, *_BENE_PMT, and

---

[3] Remember that starting in 2015, the CLM_PMT_AMT does not necessarily reflect the amount paid to the provider, rather it may include adjustments due to ACO or PBP programs.

*_PMRY_P). For Part D drugs, in addition to the Medicare and beneficiary payment variables, there is a gross drug cost variable that is the annual sum of the total drug costs (i.e., TOTALCST accrued on behalf of the beneficiary for the year).

Utilization. Additional summary variables within the cost and use segment include many variables that identify service use in a granular fashion:

- **Stays variables** — count of hospital stays (unique admissions, which may span more than one facility) in the inpatient setting for a given year. CMS defines a hospital stay as a set of one or more consecutive inpatient claims where the provider discharges the beneficiary only on the most recent claim in the set. If a different provider accepts a beneficiary, CMS considers the stay continued even if there is a discharge date on the claim from which a provider transferred the beneficiary. The CLM_THRU_DT for the last claim associated with the stay must have been in the year of the data file. Stays that cross-over into another calendar year would only appear in the year when the stay ended (e.g., a stay that began in 2017 but ended in 2018, CMS counts this as a stay in the 2018 file).
  - Acute stays — count of acute hospital stays during the year; variable called ACUTE_STAYS
  - Stays — in the OIP, SNF, and hospice settings; variable called STAYS is the count of each type of stay during the year (OIP_STAYS, SNF_STAYS, HOS_STAYS)

- **Covered days** — for acute, OIP, SNF, and hospice settings; Medicare-covered days (Medicare will not cover all the days for an institutional stay). CMS calculates the covered days (variables in a data file called ACUTE_COV_DAYS,[4] OIP_COV_DAYS, SNF_COV_DAYS, and HOS_COV_DAYS) by summing the CLM_UTLZTN_DAY_CNT for the particular type of services for the year. The CCW team considers all claims that end during the calendar year (i.e., the CLM_THRU_DT must have been during the year).

- **Readmissions** — count of hospital readmissions in the acute inpatient setting for a given year (a variable called READMISSIONS). The CLM_THRU_DT for the original admission must have been in the year of the data file; however, the algorithm allows for the readmission claim to have occurred in January of the following year.

  For example, analysts should consider a beneficiary readmitted when she has an acute inpatient stay with a discharge status that indicates she has not expired (DSCHRG_STUS≠20) or left against medical advice (DSCHRG_STUS≠07) within 30 days of a previous acute inpatient stay that also has a discharge status that she has not expired or left against medical advice. All beneficiaries without an ACUTE stay will have a missing value; beneficiaries with an ACUTE stay who do not have a subsequent readmission will have a zero value.

- **Emergency room visits** — the MBSF-CU also identifies emergency room (ER) visit use. We capture two setting scenarios — ER visits where the patient became hospitalized at the same facility, or outpatient ER visits and the hospital did not admit the patient in the same facility. Researchers can obtain information for ER utilization by examining the revenue center records for the claims. ER revenue center codes were any of the following: 0450, 0451, 0452, 0456, or 0459. The inpatient ER visits (variable called IP_ER_VISITS) are a subset of ACUTE services because CMS does not pay the ER visit separately from the hospitalization. The hospital outpatient ER visits (variable called HOP_ER_VISITS) are a subset of HOP services. The sum of these two variables is the total ER use for the beneficiary for the year.

---

[4] For inpatient stays — the algorithms consider all claims (in the ACUTE_STAYS variable). However, for the count of covered days (ACUTE_COV_DAYS) and payments (ACUTE_MDCR_PMT), the algorithms include only claims that ended during the calendar year where the payment amount was >= $0.

A listing of the variables that summarize utilization for Part A claims appears in .

Table 11. Part A service settings and corresponding utilization variables

Service type	Medicare-covered days	Stays	ER use
Inpatient acute care hospital (and CAH)	ACUTE_COV_DAYS	ACUTE_STAYS	IP_ER_VISITS
Other inpatient hospital*	OIP_COV_DAYS	OIP_STAYS	
Skilled nursing facility	SNF_COV_DAYS	SNF_STAYS	
Home health		HH_VISITS	
Hospice	HOS_COV_DAYS	HOS_STAYS	

* Other hospitals include inpatient psychiatric facilities (IPF), cancer hospitals, and children's hospitals.

The variables that summarize utilization for the Part B institutional claims are in .

Table 12. Algorithm used in categorizing Part B institutional claims into a setting

Part A values	Label	Algorithm	Medicare payment variable name
HOP	Hospital outpatient	NCH_CLM_TYPE_CD= 40	HOP_MDCR_PMT

- **Visits**
  - HH — the annual sum across all Part A claims for the number of home health visits on each claim (i.e., sum of CLM_HHA_TOT_VISIT_CNT); variable called HH_VISITS
  - HOP — this variable is the count of unique revenue center dates (as a proxy for visits) in the hospital outpatient setting for a given year; variable called HOP_VISITS
  - Emergency room — two different variables 1) ER visits where the patient became hospitalized at the same facility (variable called IP_ER_VISITS), or 2) outpatient ER visits and the same hospital did not admit the patient (variable called HOP_ER_VISITS), which are a subset of HOP services

Table 13. Algorithms used in categorizing Part B non-institutional claims into settings

Non-institutional Part B values	Label	Algorithm: NCH_CLM_TYPE_CD= 71, 72, 81, 82 (BETOS codes)*	Medicare payment variable name
ASC**	Ambulatory surgical center	LINE_CMS_TYPE_SRVC_CD="F" and NCH_CLM_TYPE_CD= 71 or 72	ASC_MDCR_PMT
ANES	Anesthesia	Anesthesia (P0) where CARR_LINE_MTUS_CD='2' and NCH_CLM_TYPE_CD= 71 or 72	ANES_MDCR_PMT
PTB_DRUG	Part B drug	Chemotherapy (O1D), other Part B drug (O1E), immunization (O1G), DME drug (D1G), imaging drugs (I1E and I1F)	PTB_DRUG_MDCR_PMT
PHYS	Physician office	Physician office (M1A or M1B)	PHYS_MDCR_PMT
EM	Evaluation and management	Hospital (M2), emergency room (M3), home or nursing home visit (M4), specialist (M5), and consultation (M6)	EM_MDCR_PMT
DIALYS	Dialysis services	Dialysis services (P9) and NCH_CLM_TYPE_CD= 71 or 72	DIALYS_MDCR_PMT
OPROC	Other procedures (not ANES or DIALYS)	Other — major procedure (P1), major cardiac (P2), major orthopedic (P3), eye (P4), ambulatory procedure (P5), minor procedure (P6), oncology	OPROC_MDCR_PMT

Non-institutional Part B values	Label	Algorithm: NCH_CLM_TYPE_CD= 71, 72, 81, 82 (BETOS codes)*	Medicare payment variable name
		procedure (P7), and endoscopy (P8) — and NCH_CLM_TYPE_CD= 71 or 72	
IMG	Imaging	Standard imaging (I1), advanced imaging (I2), echography (I3), and imaging procedure (I4)	IMG_MDCR_PMT
TEST	Laboratory or test	Laboratory test (T1) and other test (T2) — and NCH_CLM_TYPE_CD= 71 or 72	TEST_MDCR_PMT
DME	Durable medical equipment	DME supplies (D1A–D1E) and orthotic devices (D1F)	DME_MDCR_PMT
OTHC	Other Part B carrier services	Ambulance (O1A), chiropractic (O1B), parenteral nutrition (O1C), vision, hearing or speech services (O1F), and other/unclassified Part B service (Y1, Y2, Z2, and missing)	OTHC_MDCR_PMT

 * The first two or three digits of the identified BETOS codes appear in parentheses.
 ** The algorithms within this table are hierarchical — Researchers must identify ASC first, and OTHC must be last.

Additional summary variables within the Cost and Use segment related to the Part B non-institutional claims include:

- **Events** — an event is each claim line item that contains the relevant service. The variables which summarize events count all relevant line items (i.e., line items corresponding with each type of service) for the beneficiary for the year. One such variable is PHYS_EVENTS (i.e., the number of occurrences [line items on the claim] for a physician face-to-face visit); there is an *_EVENTS variable corresponding to each Part B non-institutional setting in Table 13.

There is a single variable that summarizes utilization for the Part D events, depicted in Table 14.

Table 14. Algorithm used in categorizing Part D prescription drug events

Part D values	Label	Algorithm	Total payment variable name
PTD	Part D	PDE file	PTD_TOTAL_RX_CST

The Cost and Use segment also contains Part D summary cost and use; this information is available for all beneficiaries who have Part D coverage, even those in managed care plans (i.e., MA-PD plans):

- Part D Medicare (Part D plan) payment for filled prescriptions for covered drugs (variable called PTD_MDCR_PMT) — calculated as the sum of two CCW variables: the amount paid by the plan for Part D covered drugs (CVRD_D_PLAN_PD_AMT) and any low-income subsidy (LIS) amount (LICS_AMT). **NOTE:** This variable does not include all costs to Medicare for the Part D benefit (also does not consider any applicable rebate amounts or other discounts).

- Part D beneficiary payment (variable called PTD_BENE_PMT; cost-sharing for filled prescriptions) — calculated as the sum of the CCW variables: patient pay amount (PTPAYAMT), other true out-of-pocket (TrOOP) amount (OTHTROOP), and patient liability reduction due to other payer (PLRO) amount (PLRO_AMT) for Part D drugs for a given year.

- Part D prescription drug events (PDE; variable called PTD_EVENTS) — count of all PDE IDs (i.e., unique prescription fill events) for the year.

- Part D fills (variable called PTD_FILL_CNT) — PDEs consist of highly variable days' supply of the medication. This derived variable creates a standard 30-day supply of a filled Part D prescription and counts this as a "fill." The Part D fill count does not indicate the number of different drugs the person is using, only the total months covered by medication (e.g., if a patient is receiving a full year supply of medication, whether this occurs in one transaction or 12 monthly transactions, the fill count = 12; if the patient is taking three such medications, the fill count=36).

- Part D total prescription cost (variable called PTD_TOTAL_RX_CST) — the gross drug cost (TOTALCST) of all Part D drugs for a given year. Note that the sum of the plan/Medicare and beneficiary share of the payments will not equal the total drug cost if there is a Part D LIS or third-party payer (e.g., the VA or TRICARE).

F. National Death Index (NDI) Segment

The Centers for Disease Control and Prevention (CDC) provides data for this segment of the MBSF to CCW. The original source of this information is state vital statistics offices, which record information from death certificates (reference CDC documentation regarding the NDI). These data are available for decedents, however since the source data originate outside of CMS, the update cycle is not the same as the other MBSF segments. For details, reference the CCW Data Availability web page. Researchers may only use the NDI files within the CCW Virtual Research Data Center (VRDC). Researchers wishing to obtain this NDI segment of the MBSF must obtain additional approval beyond the CMS DUA. Table 15 lists the data available in the NDI segment.

Table 15. NDI segment variables

Variable name (long)	Variable description	Brief definition
NDI_DEATH_DT	NDI death date	NDI date of death
NDI_STATE_DEATH_CD	NDI state death code	NDI state of death (SSA numeric code)
DEATH_CERT_NUM	NDI death certificate number	NDI death certificate number
ICD_CODE	ICD-10 code	ICD-10 cause of death code
ICD_TITLE	Label for ICD-10	ICD-10 cause of death title (label)
ICD_CODE_358	358 ICD-10 recodes	358 selected causes of death, ICD-10 recodes
ICD_CODE_113	113 ICD-10 recodes	113 selected ICD-10 cause of death and enterocolitis due to clostridium difficile
ICD_CODE_130	130 ICD-10 recodes	130 selected ICD-10 causes of infant death
ENTITY_COND_1 (through 20)	Entity axis conditions	NDI entity axis cause of death — condition (for 1999–2006 there were up to eight variables, for 2007–2016 there are 20 variables, sequentially numbered)
RECORD_COND_1 (through 20)	Record axis conditions	NDI record axis cause of death — condition (for 1999–2006 there were up to eight variables, for 2007+ there are 20 variables, sequentially numbered)

Researchers can find additional information regarding the cause of death recodes in *ICD-10 Cause-of-Death Lists for Tabulating Mortality Statistics* instruction manual on the CDC website.

Chapter 3. Condition Segments

The CCW makes it easy to study chronic diseases by incorporating variables for common chronic conditions and other chronic or potentially disabling conditions (OTCC), which identify additional chronic health, mental health, and substance abuse conditions. These condition variables are available through MBSF segments for two different sets of condition algorithm specifications: Chronic Conditions and OTCC.

For the CCW Chronic Conditions, there are two versions of the algorithms, depending on the years of data requested. Due to updates in the Chronic Conditions algorithms, there are a total of three MBSF conditions files:

1. the 27 CCW Chronic Conditions (available 1999-2020[5]; delivered as the MBSF_CC_YYYY file), or
2. the 30 CCW Chronic Conditions (2017 forward; delivered as the MBSF_CHRONIC_YYYY file).
3. The OTCC conditions (delivered as the MBSF_OTCC_YYYY file).

The predefined conditions use FFS claims-based algorithms to indicate that treatment for a condition appears to have taken place; therefore, researchers cannot determine whether providers treated the managed care enrollees for the condition(s) of interest. This limitation also applies, perhaps to a lesser extent, to newly eligible Medicare beneficiaries who may have only a partial year of FFS coverage. The CCW team designed the chronic condition variables to examine service patterns — which serve as a proxy for indicating the person likely is receiving treatment for the condition. Medicare claims use the International Classification of Diseases (ICD) to classify all diagnoses, identifying the condition(s) for which a patient is receiving care. CMS used the ninth version of the ICD codes (ICD-9) until September 2015. The switch to using ICD-10-CM on claims for discharges and services occurred on October 1, 2015. All CCW condition algorithms use ICD-10 codes for October 2015 forward.

It is important to note that the major objective of the chronic condition indicator variables is to allow easy extraction of relevant clinical cohorts from a very large database. The CMS and CCW teams intended for the chronic conditions definitions to be somewhat broad, so that more researchers could request data extractions based on these definitions — then refine the specifications as needed to fit their own data needs. Investigators should use caution in employing the chronic condition definitions for calculating population statistics.

The CCW team encourages investigators to determine whether they should make restrictions to the CCW Chronic Conditions segment and/or the CMS OTCC 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.

Researchers may request CCW data for any of the predefined conditions as defined by CMS.

A. CCW Chronic Conditions

Below is a list of the common chronic disease classifications. The CCW website has more information about these conditions and their algorithms; reference the Condition Categories tab on the CCW website.

---

[5] The CCW team will produce the 2021 data file after the data matures for a full calendar year; we expect the final run of the MBSF_CC_2021 in the first quarter of 2023.

1. List of CCW Chronic Conditions classifications

- Acute Myocardial Infarction
- Alzheimer's Disease
- Anemia
- Asthma
- Atrial Fibrillation and Flutter
- Benign Prostatic Hyperplasia
- Cancer, Breast
- Cancer, Colorectal
- Cancer, Endometria
- Cancer, Lung
- Cancer, Prostate
- Cancer, Urologic (Kidney, Renal Pelvis, and Ureter) ‡
- Cataract
- Chronic Kidney Disease
- Chronic Obstructive Pulmonary Disease
- Depression, Bipolar, or Other Depressive Mood Disorders

- Diabetes
- Glaucoma
- Heart Failure and Non-Ischemic Heart Disease
- Hip/Pelvic Fracture
- Hyperlipidemia
- Hypertension
- Hypothyroidism*
- Ischemic Heart Disease
- Non-Alzheimer's Dementia**
- Osteoporosis With or Without Pathological Fracture
- Parkinson's Disease and Secondary Parkinsonism†
- Pneumonia, All-cause†
- Rheumatoid Arthritis/Osteoarthritis

 * Within the 27 CCW Chronic Conditions, this condition is "Acquired Hypothyroidism."
** Within the 27 CCW Chronic Conditions, this condition is "Alzheimer's Disease, Related Disorders, or Senile Dementia."
 † These conditions are not present within the 27 CCW Chronic Conditions.

B. CMS OTCC

Below is a list of the CMS OTCC classifications. The CCW website has more information about these conditions and their algorithms; reference the [Other Chronic Health, Mental Health, and Potentially Disabling Condition Categories](#) tab on the CCW website.

1. List of OTCC

- ADHD, Conduct Disorders, and Hyperkinetic Syndrome
- Alcohol Use Disorders
- Anxiety Disorders
- Autism Spectrum Disorders
- Bipolar Disorder
- Cerebral Palsy
- Cystic Fibrosis and Other Metabolic Developmental Disorders
- Depressive Disorders
- Drug Use Disorders
- Epilepsy
- Fibromyalgia, Chronic Pain, and Fatigue
- Human Immunodeficiency Virus and/or Acquired Immunodeficiency Syndrome (HIV/AIDS)

- Intellectual Disabilities and Related Conditions
- Learning Disabilities
- Leukemias and Lymphomas
- Liver Disease, Cirrhosis and Other Liver Conditions
- Migraine and Chronic Headache
- Mobility Impairments
- Multiple Sclerosis and Transverse Myelitis
- Muscular Dystrophy
- Obesity
- Opioid Use Disorder
- Other Developmental Delays
- Peripheral Vascular Disease (PVD)
- Personality Disorders
- Post-Traumatic Stress Disorder (PTSD)
- Pressure and Chronic Ulcers

- Schizophrenia
- Schizophrenia and Other Psychotic Disorders
- Sensory — Blindness and Visual Impairment
- Sensory — Deafness and Hearing Impairment
- Sickle Cell Disease
- Spina Bifida and Other Congenital Anomalies of the Nervous System
- Spinal Cord Injury
- Tobacco Use
- Traumatic Brain Injury and Nonpsychotic Mental Disorders due to Brain Damage

- Viral Hepatitis (General), including:
  - Hepatitis A
  - Hepatitis B (acute or unspecified)
  - Hepatitis B (chronic)
  - Hepatitis C (acute)
  - Hepatitis C (chronic)
  - Hepatitis C (unspecified)
  - Hepatitis D
  - Hepatitis E

C. Variables and Values in the MBSF Conditions Segments

Within each of the CCW Chronic Conditions variables and 40 Other Condition variables, the values indicate whether the beneficiary had claims for services during the time frame for the condition (i.e., based on the FFS administrative claims pattern, providers are likely treating for the condition — or not). The reference period is the look-back period during which the other criteria occur. It is possible for a beneficiary to meet the claims criteria for a given year and not the next year. For example:
- A researcher submits Request A for the 2018 claims cohort of beneficiaries with chronic kidney disease. The CCW team identifies the cohort by applying the chronic kidney disease criteria to the universe of applicable claims for service provided on or before December 31, 2018, back through January 1, 2017 (a two-year reference period). A beneficiary meets the cohort inclusion criteria with one qualifying claim occurring in 2017 and has no subsequent claim meeting the specified criteria.

- An investigator submits Request B for 2019 claims cohort for chronic kidney disease, with a look-back period of December 31, 2019, back through January 1, 2018. Since the beneficiary's only qualifying claim occurred in 2017, the beneficiary does not meet the inclusion criteria.

- The CCW team extracts data for the beneficiary in the cohort for Request A, but not Request B.

Custom definitions allow researchers to request a cohort based on unique criteria provided by the researcher (e.g., all claims for a particular procedure, diagnosis, or specified population). A researcher can also use this approach if the researcher used a different definition for one of the conditions already defined by the CCW classifications. This type of request may also include data requested based on a finder file using identifiers from a previous study.

The condition variables consider clinical criteria (from administrative claims), coverage criteria (from enrollment data), and specified periods. The clinical criteria consider variations of the following:
- ICD-9-CM/ICD-10-CM, CPT4, or HCPCS codes
- Claim type(s) and count(s)
- Date(s) of service (e.g., claim thru dates at least one day apart)

The coverage criteria consider variations of Medicare Part A, B, and no HMO coverage. The specified periods, or reference periods, consider the length of time during which the clinical and coverage criteria occur.

The Chronic Conditions segment and the CMS OTCC segment of the MBSF include the following types of chronic condition variables:

1. **Yearly Indicator (or end-of-year indicator)**
   Algorithm criteria applied, using December 31 as the end of the reference year (e.g., 2018 yearly variable for the algorithm with one-year reference period includes services between 01/01/18–12/31/18). The following are valid values for the yearly variable:

    0 = Beneficiary did not meet claims criteria or have sufficient fee-for-service (FFS) coverage

    1 = Beneficiary met claims criteria but did not have sufficient FFS coverage (i.e., one or more months [but less than 12 months for a one-year condition, 24 months for a two-year condition, or 36 months for a three-year condition] Part A and Part B without HMO)

    2 = Beneficiary did not meet claims criteria but had sufficient FFS coverage

    3 = Beneficiary met claims criteria and had sufficient FFS coverage (i.e., all 12/24/36 months [or all months before the date of death] Part A and Part B and none of these month's HMO)

2. **Mid-year Indicator (available only in the MBSF_CC_YYYY file)**
   Algorithm criteria applied, using July 1 as the end of the reference year (e.g., 2018 mid-year variable for the algorithm with one-year reference period includes services between 07/01/17–06/30/18). The mid-year version of the conditions is only available for the 27 CCW chronic conditions ((**the MBSF_CC_YYYY file**). Researchers can use the MBSF to determine whether the beneficiary was alive and enrolled on July 1 to produce statistics. The following are valid values for the mid-year variable, and are the same as for the yearly variables:
   0 = Neither claims nor coverage met
   1 = Claims coverage not met
   2 = Claims not met; coverage met
   3 = Claims and coverage met

3. **First occurrence date (or "ever" date)**
   Date the beneficiary first met the clinical claims criteria of the algorithm (no coverage criteria applied). For the MBSF_CC and MBSF_OTCC, the earliest possible date is 01/01/1999. However, for the 30 CCW conditions, which are available starting with the MBSF_CHRONIC_2017, the earliest possible date is 01/01/2016.

   Values are null (missing) if the person never had a pattern of Medicare FFS claims indicating treatment for the condition. Some beneficiaries obviously became eligible for Medicare before the earliest possible "ever" date. For beneficiaries who joined Medicare after that date, their ever dates will not precede the start of their Medicare coverage (i.e., the COVSTART variable in the MBSF).

The algorithm for claims criteria includes a seven-day grace period for the claim through dates occurring within:
- seven days before the first date of coverage
- seven days after the date of death

NOTE: Unless otherwise specified by the researcher, standard data requests for a specified condition will include (by default) all beneficiaries with yearly (or mid-year, if requested) variable = 1 or 3.

D. Control Populations

Researchers should request control populations with the initial data request, using the data request form to specify inclusion/exclusion criteria for the control population. Researchers may request control populations consisting of beneficiaries with particular conditions; alternatively, a population may lack in any chronic conditions, if desired. Researchers can request a 1% or 5% sample file or customize the control population as needed. When requesting a control population, researchers should include type(s) of data files, applicable diagnosis, procedure codes, or DRGs, periods, and any related demographic selection criteria.

Chapter 4. Medicare Claims Data Available through the CCW

The CCW system includes Medicare enrollment and eligibility information, Medicare institutional and non-institutional claims, Medicare Part D enrollment and prescription drug fill events. The FFS claims data files delivered from the CCW contain a subset of the source files. The CCW team removes the variables used infrequently or not applicable to a particular setting. The CCW team adds key variables in the 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]). The CCW team uses the last date on the claim, referred to as the CLM_THRU_DT, to partition the claims into calendar year files. The Data Dictionaries tab on the CCW website. contains a description of the variables.

A. Medicare Part A and B Claims

The Medicare claims found in the CCW are generally fee-for-service (FFS) Part A and B claims only (i.e., encounter information for services provided by MA plans is available starting with 2015; encounter records are in a separate CCW data product[6]). However, there are a few situations where the claims data does include services for MA enrollees. The two most notable instances are hospice care, which MA plans do not cover (FFS Medicare pays), and inpatient and skilled nursing facility services for beneficiaries enrolled in certain MA plans reimbursed based on costs and have the option of getting CMS to process those claims. A technical publication from ResDAC explains the nuances of when services for managed care enrollees appear in Medicare's claims data; reference *Identifying Medicare Managed Care Beneficiaries from the Master Beneficiary Summary or Denominator Files* on the ResDAC website.

1. Structure of Claims

Institutional providers (such as hospitals, skilled nursing facilities, clinics, home health agencies, hospices, and outpatient dialysis facilities) bill for services using the 837I standard electronic format (previously known as the UB-04 claim form, or more recently as Form CMS-1450). Note that institutional providers use this form to bill for all services that they provide, regardless of whether Part A or Part B covers the service. Each claim has "base" and "revenue center" records. More information regarding the contents and processing of these claim forms is available on the CMS website; please reference *Medicare Billing: Form CMS-1450 and the 837 Institutional* on the CMS website.

Non-institutional providers (such as physicians, other health care practitioners, and durable medical equipment [DME] providers) bill for services using the 837P electronic claim form (also known as the CMS-1500 claim form). The Part B benefit covers these services and consists largely of professional services and DME. Similar to institutional claims, each non-institutional claim has "base" and "line item" records. More information regarding the contents and processing of these claim forms is available on the CMS website.

The CMS contractors known as Medicare Administrative Contractors (MACs) submit and process both kinds of claims. As a historical note, the MACs have replaced separate entities that once processed institutional claims (known as Fiscal Intermediaries or FIs) and non-institutional claims (known as carriers).

Part D is not the only part of Medicare that covers outpatient prescription drugs; Part B covers certain drugs generally administered in a medical setting, such as chemotherapy, infused drugs, and some vaccines. Records for those drugs will appear as Part B claims rather than Part D events.

---

[6] The CCW team provides an Encounter data dictionary and user guide on the CCW website.

For institutional and non-institutional claims, the **base record** contains the base or core portion of the claim. Each claim, uniquely identified by the CLM_ID, will also have at least one associated detailed record with more information regarding the particular services rendered. For institutional claims, detailed records are known as **revenue center records**; for non-institutional claims, they are known as **line-item records.** Investigators can identify the detailed line/revenue records within a claim using the sequential CLM_LINE_NUM. Both types of claims data have their structure and reference the source file record layouts and definitions on the CCW website.

Additional details regarding the use of the base claim or detailed revenue or line records are available in the *CCW Technical Guidance: Getting Started with CMS Medicare Administrative Research Files* document on the CCW website.

The types of Medicare enrollment and claims files, as well as the linkage key(s) to use are in Table 16.

Table 16. Available CCW Medicare data files (enrollment/FFS claims)

Type of file	Files	Years	Linking key or stand-alone file
Medicare enrollment	Master Beneficiary Summary File (MBSF with CME as source)[7]	2006–current	CCW BENE_ID
	Master Beneficiary Summary File (MBSF with EDB as source)	1999–2005	CCW BENE_ID
Institutional claims	Inpatient (IP) base claim files	1999–current	CCW BENE_ID
	Inpatient revenue center files	1999–current	CCW CLM_ID*
	Outpatient (OP) base claim files	1999–current	CCW BENE_ID
	Outpatient revenue center files	1999–current	CCW CLM_ID*
	Skilled nursing facility (SNF) base claim files	1999–current	CCW BENE_ID
	Skilled nursing facility revenue center files	1999–current	CCW CLM_ID*
	Home Health Agency (HHA) base claim files	1999–current	CCW BENE_ID
	Home Health Agency revenue center files	1999–current	CCW CLM_ID*
	Hospice (HOS) base claim files	1999–current	CCW BENE_ID
	Hospice revenue center files	1999–current	CCW CLM_ID*
Reference code — delivered with institutional claim files (**NOTE:** separate file for each institutional setting–IP, OP, SNF, HHA, and HOS)	Institutional condition code files	1999–current	CCW CLM_ID
	Institutional occurrence code files	1999–current	CCW CLM_ID
	Institutional span code files	1999–current	CCW CLM_ID
	Institutional value code files	1999–current	CCW CLM_ID
	Demonstration/Innovation code files	2010–current	CCW CLM_ID
Non-institutional claims	Carrier claim files (physician/supplier)	1999–current	CCW BENE_ID
	Carrier Line Files (physician/supplier)	1999–current	CCW CLM_ID*
	Demonstration/Innovation code files	2010–current	CCW CLM_ID
	Durable medical equipment (DMERC) claim files	1999–current	CCW BENE_ID

[7] CCW delivered the Master Beneficiary Summary File (MBSF with EDB as source) for enrollment years 1999–2015.

Type of file	Files	Years	Linking key or stand-alone file
	Durable medical equipment (DMERC) line files	1999–current	CCW CLM_ID*
	Demonstration/Innovation code files	2010–current	CCW CLM_ID
Other CCW Medicare	Medicare Part D event data**	2006–current	CCW BENE_ID
	Other Research Identifiable Files (CCW BENE_ID to HIC/SSN/other crosswalk)	Current	CCW BENE_ID

*The CCW_CLM_ID is the unique key to link revenue center information (for institutional claims) or line-item information (for non-institutional claims) to a specific claim.

**The *CCW Medicare Part D Data User Guide* on the CCW website explains Medicare Part D events and Part D characteristics files in detail.

2. Final Action Status of Claims

Health care 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. For example, a hospital might need to revise the dates of service or diagnosis codes on an inpatient claim, or a physician might need to submit additional modifiers to specify the type of surgery that they performed. Any revision or adjustment requires a new claim (technically, most changes require two new claims, because the provider must submit one claim to cancel the initial claim and then submit another claim with the updated information), and providers may revise some claims more than once.

The final action claim is the version of the claim where providers have resolved all adjustments to earlier claims. CMS accurately records the final action on the claim. Since weeks or months can pass between providing a service and submitting the final action claim, CCW generally waits for the final action claim to appear before extracting data files for delivery to researchers. The CCW and CMS teams call this time the run-out period.

Researchers should not consider claims final or complete until one year after the claim through date (CLM_THRU_DT). CCW allows a full 12 months for claims to "mature" and before considering the data files final (i.e., a 12-month run-out period). From 2005 forward, the CCW flags late-arriving records and does not deliver claims processed more than 365 days after the date of service. Providers process over 99% of claims within nine months of service (**NOTE:** this processing time varies slightly by care setting). For additional information, reference the *CCW Technical Guidance: Getting Started with CMS Medicare Administrative Research Files* document on the CCW website that describes the level of claims maturity by processing month and setting.

To be consistent with the National Claims History (NCH), in May 2017, the CCW reprocessed the 2005–2017 Medicare institutional claims using the revised CMS NCH final action (FA) algorithm.[8] The RIFs for all FFS claims thereafter use the updated FA algorithm.

3. Payment Fields on Claims

Numerous variables in the claims files make it possible to determine various perspectives on payments. The first is to determine the responsibility for payments: 1) the amount the provider charged, 2) the amount paid by Medicare, 3)

---

[8] The updated FA logic did not have a significant impact on overall claim counts or Medicare payment amounts. Total institutional claim counts and Medicare payments for 2015 changed by less than 0.03%. There was no impact from the FA update on the carrier, durable medical equipment, Part D, or enrollment data.

the beneficiary cost-sharing amount, and 4) the total amount paid to the provider. CCW describes these fields in the codebook, and we provide some guidance on how to use and interpret some of the payment fields for the various claim types — reference the *CCW Technical Guidance: Getting Started with CMS Medicare Administrative Research Files* document on the CCW website.

CMS routinely updates fee schedules, and periodically, there are significant changes in Medicare payment policy. It is beyond the scope of this document for CCW to summarize these changes; however, we wish to alert analysts to a payment reform initiative that impacts the dollar amounts in the CLM_PMT_AMT field starting in 2014 (with significant impact observed 2017 forward). Medicare may have included separate lump-sum payments to an ACO or other PBP program to an organization that is different from the provider organization that billed for the claim. 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 does not pay a portion of the actual Medicare payment amount to the provider; rather they distribute it to the ACO or PBP program.[9]

4. Claims Coding Systems

Health care claims vary in the kinds of information they require providers to submit about a patient's diagnoses and/or procedures obtained during a health care encounter.

Table 17 summarizes the types of information available in the various types of claims files.

Table 17. Diagnosis, procedure, and service codes used on Medicare claims

Type of code	Part A	Part B institutional	Part B non-institutional
ICD-9/ICD-10-CM diagnosis code	X	X	X
Diagnosis-related group (DRG)	X		
Revenue center code	X	X	
ICD-9/ICD-10-PCS procedure code	X		
Current Procedural Terminology (CPT) code	X	X	
Healthcare Common Procedure Coding System (HCPCS) code		X	X
Berenson-Eggers Type of Service (BETOS) code			X
Ambulatory payment classification (APC) code		X	

Medicare claims used the ninth version of the International Classification of Diseases (ICD-9) to classify all diagnoses, which identify the condition(s) for which a patient is receiving care until September 2015. CMS switched to the next version of diagnosis codes, ICD-10-CM, for discharges and services on October 1, 2015. 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),[10] with one diagnosis identified on the claim as the principal or primary diagnosis.

---

[9] Additional information and details regarding the identification of these claims and dollars are in the *CCW Technical Guidance: Getting Started with CMS Medicare Administrative Research Files* document on the CCW website.

[10] For services in 2010 and earlier, the CCW used a different version of the claim record, and there were up to 10 diagnosis codes for Part A and eight for Part B. Starting with the NCH version "J" of the claim record in 2011, more diagnoses fields are available.

Medicare pays for inpatient hospital care using case-mix groups known as diagnosis-related groups (DRGs), a classification system that groups similar clinical conditions and procedures. To determine the appropriate DRG, Medicare uses the beneficiary's principal diagnosis and secondary diagnoses, as well as any procedures furnished during the stay. 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 claims. CMS maps the ICD-10 diagnosis codes to the appropriate MS-DRG.

Medicare uses other forms of case-mix groups to pay for skilled nursing facility care (resource utilization groups, or RUGs) and home health (home health resource groups, or HHRGs). The RUG for SNF claims appear in the HCPCS_CD field (when the REV_CNTR code is 0022, then the first three digits of the HCPC_CD are the RUG). The HHRG for a particular revenue center (when the REV_CNTR code is 0023) is located in the data field called the ambulatory payment classification (APC) or Health Insurance Prospective Payment System (HIPPS) code (REV_CNTR_APC_HIPPS_CD). For more information about APCs, reference Medicare Payment Systems on the CMS website.

Revenue centers are distinct cost centers within an institutional provider that can each submit separate charges. For example, most hospitals have distinct revenue centers for the emergency department, intensive care unit, physical therapy, laboratory, pharmacy, imaging, and so on. Revenue center codes (variable called REV_CNTR in the revenue center file for all Part A claims) help identify different areas of the facility where the patient received care — and other types of care that may affect payment (such as blood transfusions or laboratory tests).

For Part A inpatient hospital claims, providers used ICD-9 procedure codes to describe the specific procedures they performed. Starting October 2015, CMS used the ICD-10 procedure coding system (ICD-10-PCS) in place of ICD-9 for procedure coding. For Part A claims that do not involve inpatient care, and for Part B claims, providers use CMS Healthcare Common Procedure Coding System (HCPCS) codes to describe the services rendered (variable called HCPCS_CD). There are two levels of HCPCS codes. The first level are codes from version 4 of the Current Procedural Terminology® (CPT-4), which is a numeric coding system maintained by the American Medical Association (AMA). The CPT consists of descriptive terms and identifying codes that CMS uses primarily to identify medical services and procedures furnished by physicians and other health care professionals. The AMA makes decisions about adding, deleting, or revising CPT codes. The second level consists of codes for procedures CPT does not include in the CPT codes; CMS primarily uses these for non-physician services, such as ambulance services or durable medical equipment.

The Berenson-Eggers Type of Service (BETOS) classification scheme maps HCPCS codes into seven major categories (physician evaluation and management, procedures, imaging, tests, durable medical equipment, other services, and exceptions/unclassified services), with additional sub-categories within each. Additional details regarding the use of BETOS codes (variable called BETOS_CD) in analyses are available in *CCW Technical Guidance: Getting Started with CMS Administrative Research Files* document on the CCW website.

B. Medicare Part D Prescription Drug Events

The CCW contains all Part D prescription drug events (PDEs), regardless of whether a managed care plan that includes coverage for prescription drugs or a stand-alone prescription drug plan enrolls the beneficiary. A detailed description of the Part D data is available in the *CCW Part D Data User Guide* on the CCW website.

Chapter 5. Medicare Plan Characteristics

Starting with the 2015 benefit year, the CCW has prepared a plan characteristics set of six files per year with detailed information on those plan characteristics. Researchers can join these plan characteristics to the MBSF to better understand the benefits available to enrollees. Note that these plan characteristics files enhance the original CCW Part D plan characteristics files, which are available for benefit years 2006–2014. The CCW team updated the plan base, premium, plan crosswalk, and service area files to include Part C only plans (i.e., plans that do not offer Part D prescription drug coverage — referred to as MA-PD plans) along with an indicator that allows users to subset by Medicare program type. In addition, the CCW team has added a file that identifies the conditions that apply to each special needs plan (SNP). The SNP contracts file and the current cost-sharing tier file only include plans offering a Part D benefit.

The six plan characteristics files are:
1. **Plan benefit base file or "base" plan file** — contains key information about the managed care and/or drug benefit offered by the plan sponsor. Many of the variables in this file apply only to the Part D benefit and will be blank for Part C only plans.
2. **Plan premium file** — has information on the premiums that each plan charged its enrollees. Most of the variables in this file only apply to plans that offer a Part D benefit.
3. **Plan Part D cost-sharing tier file** — describes the features of the Part D plan benefit package, such as the tiers of the formulary, and has detailed information on how the cost of drug products will vary by benefit phase, the quantity of the drug dispensed, and the type of pharmacy used (e.g., in- or out-of-network). These are always MA-PD plans.
4. **Plan service area file** — provides the regions included in the plan service area and has at least one row for every distinct plan ID and segment ID within a contract.
5. **Plan crosswalk file** — will be useful to analysts interested in examining changes over time to the plans available to enrollees. Investigators can identify new plans, terminated in the prior year, renewed, or consolidated with other plans. The file includes information for all plans in the Plan Characteristics files for the current year or the prior year.
6. Starting with the 2015 file, the **special needs plans (SNP) contracts file** — contains indicators to show covered condition categories (e.g., heart failure, diabetes) in the SNP. SNPs are always MA-PD plans.

Chapter 6. Other CMS Data Available through the CCW

The CCW system contains various types of CMS data from multiple care settings. In addition to Medicare enrollment and FFS claims data files, Medicare Advantage encounter data files are available from 2015 forward. CCW delivers Medicaid eligibility and claims data (delivered as MAX files) from 1999 through 2013, and the new T-MSIS data files starting in 2014. In addition, CCW disseminates assessment data — including the skilled nursing/nursing facility Minimum Data Set (MDS) and home health Outcome and Assessment Information Set (OASIS) assessment files. Table 17 displays a variety of CCW data files.

For each Medicare beneficiary in the data file, the unique CCW identifier provides a common link across all applicable types of data available (variable called the CCW BENE_ID). Based on the approved research request, the CCW data delivered may not include patient identifying information. Regardless of whether the delivered data has patient identifying information included, the unique beneficiary identifier provides researchers with the ability to analyze information across the continuum of care for a particular beneficiary or chronic condition cohort.

The CCW team creates the unique CCW BENE_ID from the CMS CME database, using the CME beneficiary link key, MBI or HIC number, and other beneficiary identifiers (i.e., gender, Social Security number [SSN]) for each beneficiary in the CME data. The CCW team performs analyses to ensure that they do not represent the beneficiary multiple times in the CCW BENE_ID history table. This unique CCW identifier follows an enrollee across years and other CCW research data sources. For example, the CCW also contains Medicaid enrollment and claims data, as well as assessment data (e.g., Minimum Data Set [MDS] and Outcome and Assessment Information Set [OASIS]). The BENE_ID facilitates analysis across all CMS data sources in the CCW.

A. Medicare Encounter Data

MA (Part C) encounter data RIFs are available to researchers starting with 2015. MAOs who provide services to beneficiaries under the Medicare Part C benefit, submit data to CMS, which then the CCW team uses to create the RIFs. Reference the Medicare Encounter Records data dictionaries along with a *CCW Medicare Encounter Data User Guide* on the CCW website. Table 18 lists the Medicare encounter data files available for request.

Table 18. Available encounter files

Type of encounter file	Files	Years	Linking key or stand-alone file
Institutional encounter records	Inpatient (IP) base claim files	2015 forward	CCW BENE_ID
	Inpatient revenue center files	2015 forward	ENC_JOIN_KEY*
	Skilled nursing facility (SNF) base encounter files	2015 forward	CCW BENE_ID
	Skilled nursing facility revenue center files	2015 forward	ENC_JOIN_KEY*
	Home health agency (HHA) base encounter files	2015 forward	CCW BENE_ID
	Home health agency revenue center files	2015 forward	ENC_JOIN_KEY*
	Outpatient (OP) base encounter files	2015 forward	CCW BENE_ID
	Outpatient revenue center files	2015 forward	ENC_JOIN_KEY*
Reference code — delivered with institutional encounter files (**NOTE:** separate file for each institutional setting–IP, SNF, HHA, and OP)	Institutional condition code files	2015 forward	ENC_JOIN_KEY*

Type of encounter file	Files	Years	Linking key or stand-alone file
	Institutional occurrence code files	2015 forward	ENC_JOIN_KEY*
	Institutional span code files	2015 forward	ENC_JOIN_KEY*
	Institutional value code files	2015 forward	ENC_JOIN_KEY*
Non-institutional encounter records	Carrier encounter files (physician/supplier)	2015 forward	CCW BENE_ID
	Carrier line files (physician/supplier)	2015 forward	ENC_JOIN_KEY*
	Durable medical equipment (DME) encounter Files	2015 forward	CCW BENE_ID
	DME line files	2015 forward	ENC_JOIN_KEY*

B. Assessment Data

Table 20 shows the assessment files available in the CCW. Data dictionaries for all assessment data files are on the CCW website.

C. Medicaid Data Files

1. MAX Files

The MAX data dictionaries, as well as details regarding the construction of the MAX files, are available on the Medicaid Data Sources section of CMS website. Researchers will find useful information about file content by state and year by reviewing the MAX Data Validation Tables and Data Anomalies Reports under the MAX General Information on Data section of the CMS website. Information regarding some state-specific information is also available, such as managed care penetration. Researchers can reference MAX file layouts and "Knowledgebase" articles regarding use of MAX files on the ResDAC website. A separate CCW technical guidance document provides many details regarding the use of the MAX data. Reference, *CCW Technical Guidance: Getting Started with MAX Data Files* on the CCW website.

The MAX files for all states for 1999–2012, 28 states in 2013, and 17 states in 2014 are available. For the more recent years of MAX data, fewer states have data because starting in 2013, states began to transition from using the Medical Statistical Information System (MSIS) to the T-MSIS. As a result, many states are choosing to submit data to CMS using only the T-MSIS format. Using the T-MSIS data files from states, CMS creates the TAFs.

Note that this transition to T-MSIS is a major change for states that can take many months (or years) to implement fully. Additional details are available in the *CCW T-MSIS Analytic Files (TAF) User Guide* on the CCW website. Table 19 summarizes and displays the count of states with MAX versus TAF by year.

Table 19. Count of states with MAX, MAX-T, or TAF files in CCW, by year

Count of states	2011*	2012	2013	2014	2015	2016	2017 forward
with MAX files	50	48	28	17	n/a	n/a	n/a
with MAX-T files**	1	3	23	15	21	n/a	n/a
With TAF	n/a	n/a	n/a	19	31**	52	53†

   * MAX includes all states and the District of Columbia.
  **MAX-T is a MAX file produced using T-MSIS, at least partly.
   † Starting in 2015, the TAF RIFs include Puerto Rico, and starting in 2017, the Virgin Islands.

2. TAF RIF Files

Each state compiles information regarding their Medicaid and CHIP enrollment, service utilization, and payments in the recently implemented T-MSIS format and provides T-MSIS data files to CMS. Using the T-MSIS data files from states, CMS creates the TAFs. The CCW obtains the TAF files, loads them to a database and creates claims and enrollment RIFs.

T-MSIS represents the next generation of national data for Medicaid and CHIP beneficiaries and the services they use. T-MSIS differs from MSIS in several important ways, including the timing of submissions (monthly vs. quarterly) and the amount of content states report (nearly four times as many data elements, including several new segments).

There are seven types of data files available for TAF RIF starting with 2014, including the following:
1. Annual demographic and eligibility (DE)
2. Inpatient (IP) claims
3. Long-term care (LT) claims
4. Pharmacy (RX) claims
5. Other services (OT) claims
6. Annual managed care plan (APL)
7. Annual provider (APR)

Researchers will find data dictionaries and a *CCW T-MSIS Analytic Files (TAF) User Guide* on the CCW website. Additional details regarding the contents of the TAF claims RIFs are in the TAF user guide.

3. MMLEADS Files

CCW collaborated with CMS to offer a data product designed for studying the Medicare and Medicaid dually enrolled population and calls it the Medicare-Medicaid Linked Enrollee Analytic Data Source (MMLEADS).

The CCW team produced the original MMLEADS (called V2) from manipulated CCW Medicare and MAX source data and includes enrollment/eligibility and summary claim/utilization and payment information. These files are available from 2006–2012. A data dictionary is available on the CCW website.

The CCW team made significant changes in the analytic code and key algorithms used for the MMLEADS data product due to the transition of Medicaid source data from MAX files to TAF RIFs. CCW developed an updated version of the

MMLEADS files using the TAF RIFs,[11] along with Medicare data, as the source files starting with 2016. In addition to updated algorithms, MMLEADS has a new file format, in which all files are person-level summary files. Reference the *CCW MMLEADS User Guide* and a MMLEADS data dictionary on the CCW website.

Table 20. Available assessment and MMLEADS data files

Type of file	Files	Years	Linking key or stand-alone file
Assessment	Minimum Data Set (MDS)	1999–current	CCW BENE_ID
	Outcome and Assessment Information Set (OASIS)	1999–current	CCW BENE_ID
	Swing bed	1999–current	CCW BENE_ID
	Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI)	1999–current	CCW BENE_ID
Other Medicaid analytic files	MMLEADS	2006–2012	CCW BENE_ID*
	MMLEADS (using TAF)	2016	CCW BENE_ID*

* Investigators should use the BENE_ID and STATE_CD, and if BENE_ID is not available, then use the encrypted MSIS_ID and STATE_CD as the unique key.

---

[11] TAF RIF 2016 (release 2).

Chapter 7. Format, Content, and Encryption of CCW Output Files

This section describes the content and format of the CCW output package (the CCW data physically delivered to researchers). Researchers can reference all user guides from the Analytic Guidance tab on the CCW website.
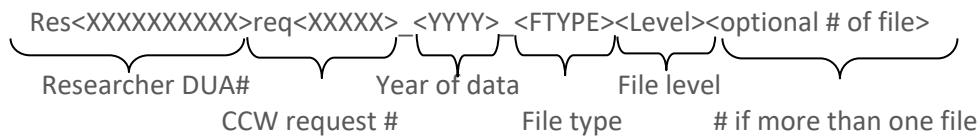
A. Format

The CCW team delivers files to the researcher in the following format. There will be several folders, each of which contains multiple files. The folders are:

📁 XXXX (folder with your CCW data request number)
📁 Extract file documentation
There may be additional folders if you have requested additional types of data.

The CCW team inserts all your data files within the folder with your CCW data request # (reference Table 22 and Table 23). There is a separate sub-folder for each year of data you requested.

The naming convention for data files is as follows:

Res<XXXXXXXXXX>req<XXXXX>_<YYYY>_<FTYPE><Level><optional # of file>

Researcher DUA#   Year of data   File level
          CCW request #   File type   # if more than one file

For example, if your DUA # was 0000077777, your CCW request number was 12345, and you obtained 2014 Part B carrier data, your folders and data files would look like this:

📁 12345
   📁 2014
     READ_ME_FIRST_REQ12345_2014.txt
     res0000077777req12345_2014_BCARRB
     res0000077777req12345_2014_BCARRL

The naming convention uses abbreviations to indicate the types of claims (or other data files).

- Medicare Beneficiary Summary File
  - Part A, B, C, and D segment — MBSF_ABCD_SUMMARY
  - Condition segment — MBSF_CC_SUMMARY
- Assessments
  - MDS
  - OASIS
  - IRF
- Claims (reference Table 21)

Table 21. Claim file names

Claim type	Claim/Revenue/ Line files	File level	Reference code files*	Demonstration code file**
Inpatient	IP	B (base) or R (revenue)	INPT	INPT_DEMO
Skilled nursing facility	SNF	B R	SNF	SNF_DEMO
Hospice	HOSPC	B R	HSPC	HSPC_DEMO
Home health administration	HHA	B R	HHA	HHA_DEMO
Hospital outpatient	OP	B R	OTPT	OTPT_DEMO
Part B carrier	BCARR	B (base) L (line item)	n/a	BCAR_DEMO
Durable medical equipment	DME	B L	n/a	DME_DEMO
Part D event data	PDE	n/a	n/a	n/a

* The reference code files include: _COND (condition code file), _OCCR (occurrence code file), _SPAN (span code file), and _VAL (value code file).

** The demonstration code file populates after 2010.

The "B" at the end of the institutional claim file names indicates a base claim file, whereas the "R" identifies the corresponding revenue center. Similarly, for the non-institutional claim files, the "B" at the end of the file name identifies the base claim file, and file and the "L" indicates a line-item level file.

If the files are large, the CCW team may divide them into two or more files, in which case there would be a sequential number at the end of the file name — such as "001," "002," to enumerate how many files of this type you receive (e.g., res0000077777req12345_2014_BCARRL001).

There are three items in the data file folder:

1. **READ_ME_FIRST_REQXXXX_YYYY.txt** — this is a text file that describes the files contained in the output package. Filename example: READ_ME_FIRST_REQ012345_2021.txt

2. **Claims files** —
   Table 22. Examples of the file names, description, and unit of analysis

Claims files filename	Claims files file description	Claims files unit of analysis
res<DUA number>_req<XXX>_<YYYY>_IPB	Inpatient base claim file	Claim
res<DUA number>_req<XXX>_<YYYY>_IPR	Inpatient revenue center file	Revenue center detail

3. **Reference code files** —are a set of five files routinely delivered for each of the institutional claim types (IP, SNF, hospice, HH, and OP). The demonstration code file exists for carrier and DME files, in addition to all of the institutional claims files (2010 forward).

Table 23. Reference code files

Reference code files filename	Reference code file description	Unit of analysis
res<DUA number>_req<XXX>_<YYYY>_ IP_COND	Institutional condition code file (our example is for an inpatient file)	Code detail
res<DUA number>_req<XXX>_<YYYY>_ IP _OCCR	Institutional occurrence code file	Code detail
res<DUA number>_req<XXX>_<YYYY>_ IP_SPAN	Institutional span code file	Code detail
res<DUA number>_req<XXX>_<YYYY>_ IP_VAL	Institutional value code file	Code detail
res<DUA number>_req<XXX>_<YYYY>_ INPT_DEMO	Inpatient demonstration code file	Demonstration number detail

B. Content

Within each yearly data folder is a README file, which you will want to read first. It is a text file that describes the files contained in the output package.

Executable files (self-decrypting archive [SDA]) contain all of the data files. **You will need to enter a password to extract each file.** Additional details regarding the data encryption and extraction process are in section 6C below.

After extracting the data files, you should compare your record count to the control counts that CCW obtained in producing your data file. These control counts are in the *.fts (file transfer summary) file. There is a separate .fts for each data file. The data files are in fixed column flat files. You can use whatever analytic software you choose. For convenience, we have included SAS read-in files. In addition to the raw data files, each executable in the output package, generates the following files as shown in <u>Table 24</u>.

Table 24. Files contained within SDAs

Filename	File description
<file name>.fts	For each extracted data file there will be a corresponding transfer summary file. The names of these files will correspond with the data file name (e.g., res<DUA number>_req<XXX>_<YYYY>_ IPB. Fts). This file transfer summary files contain: • File name • File source • File transfer mode • Row length • File transfer format • # Columns • # Rows • File size
<file name>_v6.sas	Program to read data into a SAS version 6.x environment. For example, the file inpatient_base_claims_read_v6.sas reads the inpatient base claims data into a SAS version 6.x environment.
<file name._v8.sas	Program to read data into a SAS version 8.x environment.

📁 Extract file documentation
    📊 Code reference sets.xls — describes the ICD-9 or -10 diagnosis and procedure codes, HCPC codes, revenue center, and other codes in the data files
    📄 Decryption Instructions.pdf — instructions for decrypting/uncompressing the data files
    📄 Tips on Getting Started with Data

C. Encryption Information

The encryption technique for files extracted from the CCW uses the Pretty Good Privacy (PGP) Command Line 9.0 with the SDA method. This method builds a compressed, encrypted, password-protected file using a FIPS 140-1/140-2 approved AES256 cipher algorithm. The SDA builds on the CCW production server, downloads to a desktop PC, and burns to a CD, DVD, or USB hard drive depending on the size of the files.

After the CCW team ships the data media to the researcher, they send the password to decrypt the archive to the researcher by electronic mail. Each researcher request will have unique encryption. The CCW team never packages the password and data media together. To decrypt the data files, the researcher needs to access the email containing the decryption password. The shipped data package contains detailed instructions for using this password.

Each SDA contains the data file(s), SAS® code, and a file transfer summary (.fts) file, which analysts can use to verify the data is correct.

The CCW beneficiary identifier field (BENE_ID) is specific to the CCW (not applicable to any other identification system or data source). All the requested data links use this field. CCW encrypts the data files using a cipher before delivery to the researchers. The CCW team also encrypts the claim ID (CLM_ID) and assessment ID (ASMT_ID) using the same cipher since these identifiers are also unique to a beneficiary. The CCW team designed the encrypted BENE_ID for researchers to link the data and the encrypted CLM_ID and ASMT_ID to identify records from the same claim/assessment). The cipher is unique for each DUA, and the system produces it when the researcher requests the data. The CCW team keeps this key on file for future use if requested. A researcher may stipulate in a new DUA that previous data requested, that the CCW team link to that data. CMS will then evaluate and approve or disapprove the request. If approved, the CCW team encrypts the data obtained using the same cipher as the previous DUA requests linking data.

Chapter 8. Limitations of the CCW Data

Researchers should expect anomalies in working with large, national, administrative datasets. Minimal data cleansing has occurred during the processing of CCW data. However, we describe some of the known limitations of CMS or CCW data below.

A. CCW Medicare Claims Data

Since claims (or encounter records) for most services provided to Medicare beneficiaries in managed care do not reach the FFS Medicare claim data files, researchers should view the CCW Medicare claims as providing utilization information primarily for the FFS population.

Every month, the CCW database receives and loads CMS data files. The CCW team does not consider claims final or complete until one year after the claim thru date. CCW allows a full 12 months for claims to "mature" and to consider the data files final, as explained earlier in section 4A. Since researchers request data files based on calendar years, researchers should evaluate the claims maturity or "completeness" of claims processing when requesting CCW data.

B. Assessment Data

The Quality Improvement and Evaluation System (QIES) sends the updated assessment records, 12 months after the completion date for the CCW team to load. CMS estimates that providers process over 99% of assessments within nine months of service (this processing time may vary by assessment type). QIES may update the assessments until one year after the assessment date.

The CCW team applies beneficiary matching logic, and populates the assessments with the person's BENE_ID. The presence of the BENE_ID enables the assessment records to link to other CCW Medicare and/or Medicaid data, if the DUA allows for this.

C. Invalid Values

Some of the CCW data files may contain invalid values or values not conforming to the valid values provided in the CCW supporting documentation. This is because the CCW data files contain data as received and processed from the original CMS processing source. The CCW receives data — possibly containing invalid values — processes it, stores, and delivers the data as received. The CCW team makes no modifications or conversions to "correct" invalid variable values.

One exception is the removal of spaces or decimals to the left of diagnosis or procedure codes. The CCW system removes any periods or blank spaces occurring to the left of the first valid numeric value within a diagnosis or procedure code field. The CCW team stores the diagnosis and procedure codes without periods and does remove any blank spaces occurring within a diagnosis or procedure code (between valid numeric values).

Chapter 9. Further Assistance with CCW Data

Researchers interested in working with CCW data should contact 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 is not able to 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

Acronym	Definition
ACO	Accountable Care Organization
AMI	Acute myocardial infarction
APL	Annual managed care plan file
APR	Annual provider file
BENE_ID	unique CCW beneficiary identifier
BETOS	Berenson-Eggers Type of Service
BIC	Beneficiary identification number
BSF	Beneficiary summary file
CAN	Claim account number
CC	Chronic conditions
CCW	Chronic Conditions Warehouse
CDC	Centers for Disease Control and Prevention
CME	CMS Common Medicare Environment database
CMS	Centers for Medicare & Medicaid Services
CPT4	Current Procedural Terminology
DE	Demographic and eligibility file
DME	Durable medical equipment
DMERC	Durable medical equipment regional carrier
DOB	Date of birth
DOD	Date of death
DRG	Diagnosis-related group
DUA	Data Use Agreement
EDB	CMS Enrollment Database
ER	Emergency Room/Department setting
ESRD	End- stage renal disease
FFS	Fee-for-service
HHA	Home health agency
HIC	Health insurance claim
HMO	Health maintenance organization
HOP	Hospital outpatient
HOS	Hospice
ICD-9 (or -10)	International Classification of Diseases, Ninth Revision (or Tenth)
IP	Inpatient hospital
IPF	Inpatient psychiatric facility
IRF	Inpatient rehabilitation facility
IRF/PAI	Inpatient Rehabilitation Facility/Patient Assessment Instrument
LIS	Low-income subsidy
LT	Long-term care
LTC	Long-term care hospital
MA	Medicare Advantage
MAO	Medicare Advantage Organizations
MAX	Medicaid Analytic eXtract
MBI	Medicare beneficiary identifier
MBSF	Master Beneficiary Summary File
MCBS	Medicare Current Beneficiary Survey
MDS	Minimum Data Set
MESF	Medicaid Enrollee Supplemental File
MMLEADS	Medicare-Medicaid Linked Enrollee Analytic Data Source

Acronym	Definition
NDI	National Death Index
OASIS	Outcome and Assessment Information Set
OP	Hospital or other institutional outpatient setting
OT	Other services
OTCC	Other Chronic or Potentially Disabling Conditions
PAI	Patient Assessment Instrument
PBP	Planned Benefit Package
PDE	Part D prescription events file
PDP	Prescription drug plan
PGP	Pretty Good Privacy
PLRO	Patient liability reduction due to other payer
PPO	Preferred provider organization
PS	Person summary
QIES	Quality Improvement and Evaluation System
ResDAC	Research Data Assistance Center
RIF	Research Identifiable Files
RTI	Research Triangle Institute
RX	Prescription drug
SDA	Self-decrypting archive
SNF	Skilled nursing facility
SSA	Social Security Administration
SSN	Social Security number
TAF	T-MSIS Analytic File
T-MSIS	Transformed Medicaid Statistical Information System
TrOOP	True Out-of-Pocket