

Chronic Conditions Warehouse

Your source for national CMS Medicare and Medicaid research data



Chronic Conditions Warehouse Virtual Research Data Center

Medicare Healthcare Effectiveness Data and Information Set (HEDIS) Patient-Level Detail Data Files User Guide

JUNE 2025 | VERSION 1.4

Revision Log

Date	Changed by	Revisions	Version
June 2025	D. Happe	Edits made to comply with Executive Order 14168	1.4
September 2024	K. Schneider	Removed table depicting number of variables in file 1	1.3
July 2023	K. Schneider	Added information regarding 2020 files and number of variables	1.2
April 2023	K. Schneider	Added information regarding 2017–18 files and number of variables, including the six new SES stratification variables for the PCR file (file 2) for 2018	1.1
July 2022	K. Schneider E. Erdem	Created initial document	1.0

Table of Contents

Overview	1
Chapter 1: Introduction	2
A. Source HEDIS Patient-Level Detail Data	2
1. File 1 — All Medicare HEDIS Measures Except for Plan All-Cause Readmissions	2
2. File 2 — Plan All-Cause Readmissions (PCR)	3
B. CMS Preparation of HEDIS Files	3
Chapter 2: CCW Medicare HEDIS Files	4
A. Assignment of a beneficiary identifier	4
B. All Medicare HEDIS Measures RIF (File 1) Except for Plan All-Cause Readmissions	5
1. Types of Measures	6
C. HEDIS Readmissions RIF (File 2)	10
Chapter 3: Linkage with Other CCW Medicare Data	12
A. The Master Beneficiary Summary File (MBSF)	12
B. Plan Characteristics File	12
C. Medicare Encounter RIFs	13
Chapter 4: Format, Content, and Encryption of CCW Output Files	14
Chapter 5. Further Assistance with CCW Data	16
Appendix A — List of Acronyms	17

List of Tables

Table 1. Example of SAS® naming convention for measure topics and associated metrics	6
Table 2. Example of SAS naming convention for Use of Service measure topics	7
Table 3. Example of SAS naming convention for File 1 utilization measure with weights	9
Table 4. Count of variables in readmissions RIF over time	10
Table 5. Example of SAS naming convention for readmissions file weighting variables	11
Table 6. Format and naming convention for the CCW files	14
Table 7. CCW resources accompanying data files	15

Overview

The Centers for Medicare & Medicaid Services (CMS) uses the Chronic Conditions Warehouse (CCW) to develop and manage CMS research data sources. With approval under a Data Use Agreement (DUA) to obtain CMS data for research purposes, academic researchers and certain government agencies may request Research Identifiable Files (RIFs).

CMS has authorized the release of the Healthcare Effectiveness Data and Information Set (HEDIS^{®1}) patient-level detail (PLD) data files to the research community. HEDIS is a set of standardized healthcare quality measures designed to provide information for a reliable comparison of health plan performance. HEDIS PLD data files contain audited medical record data for beneficiaries enrolled in Medicare Advantage (MA) organizations. CMS uses HEDIS data for determining the Categorical Adjustment Index values for the Medicare Part C Star Ratings,² as well as in analyses to assess differences in the quality of care across various beneficiary subgroups (e.g., ethnic, racial, sex, geographic).³ Public reporting as part of Part C Star Ratings includes only a subset of the HEDIS measures reported to CMS annually. The CMS website publicly displays some HEDIS measures at [Part C and D Performance Data](#), but CMS does not include them in the calculation of Star Ratings.

Medicare Advantage Organizations (MAO) and cost and demonstration plans submit the HEDIS PLD data files each June, reflecting health care delivered in the prior year. CMS posts annual Public Use Files (PUFs)⁴ that provide contract-level information on each HEDIS measure. Most users find that the PUFs meet their research needs. The PUFs do not contain protected health information; therefore, data users do not need a DUA.

This user guide describes the HEDIS PLD researcher files, how to use them in combination with other CCW data files, and how to obtain them. A list of acronyms used throughout the paper is in [Appendix A](#).

¹ HEDIS[®] is a registered trademark of the National Committee for Quality Assurance. The National Committee for Quality Assurance holds the copyright for HEDIS 2016, 1100 13th Street, NW, Suite 1000, Washington, DC, 20005. All rights reserved.

² The CMS website has additional information regarding the Star Ratings and the quality measures.

<https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/PerformanceData> (Accessed 09/27/2024)

³ The CMS Office of Minority Health shares stratified reports of MA performance, including information from HEDIS PLD data files. <https://www.cms.gov/About-CMS/Agency-Information/OMH/research-and-data/statistics-and-data/stratified-reporting> (Accessed 09/27/2024)

⁴ The CMS website has the MA HEDIS PUFs and associated documentation available for download (within annual ZIP files). <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDENrolData/MA-HEDIS-Public-Use-Files> (Accessed 09/27/2024)

Chapter 1: Introduction

A. Source HEDIS Patient-Level Detail Data

The National Committee for Quality Assurance (NCQA) developed and implemented HEDIS to assess the clinical quality of health plans. CMS requires that MAOs annually submit HEDIS data files for Medicare beneficiaries. Each year, NCQA publishes the technical specifications, or instructions, for all health plans to use in identifying and calculating each measure. CMS requires all MAOs to report HEDIS measures specified in its annually issued Reporting Requirements Memorandum.⁵ The required measures are in several categories, such as the effectiveness of care, access/availability of care, and utilization measures.

Plans submit HEDIS PLD data files to CMS in May through June of each year, following rigorous guidelines for the layout, format, and submission of the raw HEDIS files to meet CMS requirements.⁶ Due to the COVID-19 public health emergency, CMS did not require MAOs to submit HEDIS data for the 2019 measurement year (MY).⁷ The HEDIS metrics and the layout of the HEDIS PLD data files change each year because CMS may:

- Add or retire measures
- Update the measure specifications
- Stratify the measures to collect the metrics for subgroups

Annual CMS documentation regarding the HEDIS PUF⁸ includes information pertinent to the PLD data files and describes all the HEDIS measures submitted to Medicare. The HEDIS measures in the PUF match the PLD files, except when there are exceptions for submitting PLD files for a particular measure.

Two data files comprise the annual HEDIS PLD submission — File 1 includes all the HEDIS measures submitted except for the plan all-cause readmissions (PCR) reported in File 2.

1. File 1 — All Medicare HEDIS Measures Except for Plan All-Cause Readmissions

This person-level file includes information about all HEDIS measures, except for the PCR reported in File 2. It represents beneficiaries enrolled in each MA plan/contract during the year. The file contains indicator variables for whether the beneficiary met the denominator and numerator requirements for the HEDIS measures and weights for some measures. Some measure variables contain frequencies (i.e., the number of visits or medications) rather than binary indicators.

⁵ From 2019 forward, reference the CMS website for Part C reporting requirements. <https://www.cms.gov/Medicare/Health-Plans/HealthPlansGenInfo/ReportingRequirements> (Accessed 09/27/2024)

⁶ The CMS website has instructions for MAOs to follow for HEDIS PLD file submission. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDENrolData> (Accessed 09/27/2024)

⁷ CMS disseminated a memo to health plans on April 18, 2020. <https://www.cms.gov/files/document/covid-grs-and-marketplace-quality-initiatives-memo-final.pdf> (Accessed 09/27/2024)

⁸ The CMS website has the MA HEDIS Public Use Files and associated documentation available for download (within annual ZIP files). <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDENrolData/MA-HEDIS-Public-Use-Files> (Accessed 09/27/2024)

The annual CMS instructions for plans to complete and submit the PLD includes a full list of required HEDIS measures.⁹ In those instructions, CMS refers to the data file that pertains to MY2015 as the 2016 file (since MAOs submitted the data in 2016). The MY2015 and 2016 files use the same layout. Files change annually after that. The CMS documentation for MAOs to submit the HEDIS PLD data includes a crosswalk document illustrating which variables are the same/different from one year to the next.¹⁰

2. File 2 — Plan All-Cause Readmissions (PCR)

The HEDIS MA PCR file (referred to as File 2 in the file name) is a stay-level file that contains all variables pertinent to the PCR measures. There is a separate record for each acute inpatient hospital discharge during the year. The patients in this file are a subset of those included in File 1; however, there may be more than one record for a person if there was more than one inpatient discharge during the year. There is only a patient record in this file if the patient had inpatient services during the measurement year.

- CMS requires all MAOs to submit this file
- Medicare 1876 Cost contracts can voluntarily submit PCR data

CMS designed the PCR file, File 2, to measure unplanned acute readmissions for any diagnosis within 30 days of an inpatient discharge.¹¹ The file includes the predicted probability of acute readmission that CMS uses to calculate an actual to expected readmission ratio. CMS uses this file for the Star Ratings PCR measure.

B. CMS Preparation of HEDIS Files

CMS adds a small number of variables to the submitted files. For example, CMS assigns a record ID field (called HEDIS_REC_ID in the CCW file) to each row of data in the files. Additionally, CMS appends fields from NCQA's list of MA plans with submitted data for the CMS contract number (called HEDIS_CNTRCT_ID in the CCW file), plan type (HEDIS_PLAN_TYPE), organization type (HEDIS_ORG_TYPE), and the name of the CMS region where the organization is located (CMS_RGN_NAME).

Medicare health plans submit both summary-level HEDIS data at the plan/contract level for the annual PUF files and HEDIS PLD data to CMS for additional analyses. It is these PLD files that CCW releases to approved DUA holders, starting with MY2015.

The CCW HEDIS RIFs described in this document are the PLD files available through the CCW (i.e., the numerators, denominators, and other information used to calculate rates for each measure). The CCW team describes the two files in the next chapter.

⁹ The CMS website has instructions for MAOs to use for HEDIS PLD file submission. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData> (Accessed 09/27/2024)

¹⁰ The CMS website has instructions for MAOs to use for HEDIS PLD file submission. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData>. The PLD crosswalk is available via downloading the annual file. (Accessed 09/27/2024)

¹¹ The CMS website has instructions for MAOs to use for HEDIS PLD file submission. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData> (Accessed 09/27/2024)

Chapter 2: CCW Medicare HEDIS Files

Each year of HEDIS data stands alone; that is, the Medicare health contracts that submit data could change from year to year. In addition, the beneficiary population for which the contract submits data changes yearly, as beneficiaries enroll in different plans or choose fee-for-service (FFS) Medicare coverage.

The two annual HEDIS PLD files are available each year and delivered as a package, so users obtain both files. Since the data files contain protected health information, the CCW team considers them RIFs. CMS limits the records in the file to the approved population for the researcher's CMS DUA.

CCW obtains the final HEDIS PLD files from the CMS source data maintainer and adds two variables: the unique CCW beneficiary identifier, the BENE_ID, and the reference year, called the RFRNC_YR_HEDIS. The reference year is the HEDIS measurement year, which also corresponds with the year of the CCW HEDIS data file (not the calendar year when MAOs submitted the data).

The full record layouts are on the CCW [Data Dictionaries](#) website. When the same variable appears in more than one year of data, the CCW team includes it in the file using the same SAS name. If there is any difference in how CMS specifies the measure or defines the variable, it has a new variable name and label.

Researchers can identify each row of data in the file using the RECORD_ID. There is no intelligence included in the RECORD_ID, and a unique person may have more than one RECORD_ID if they enroll in more than one contract during the year; therefore, the CCW team includes a unique person identifier in the data files (the BENE_ID).

A. Assignment of a beneficiary identifier

The CCW team removes the Medicare beneficiary health insurance claim (HIC)¹² numbers from the data files delivered to researchers. Again, the CCW team adds a unique person identifier (the CCW beneficiary identifier; variable called the BENE_ID) to the HEDIS research files. The BENE_ID allows researchers to analyze information across the entire continuum of care, both within and across years, without using standard person identifiers such as a Social Security number, Medicare HIC, or Medicare Beneficiary Identifier (MBI) for Medicare beneficiaries. This unique CCW identifier follows an enrollee across states and other CCW research data sources. If the same person enrolls in Medicare and Medicaid, researchers may combine those records using the BENE_ID. For example, the CCW contains Medicare enrollment data, FFS claims data, MA encounter data submitted by plans to CMS, and 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.

For some years, it is possible that the CCW team did not assign BENE_IDs to a small number of HEDIS PLD records due to insufficient identifiers on the PLD files, and CCW does not release these records.

The unique BENE_ID field is specific to the CCW and does not apply to any other identification system or data source. Before delivering the data files to researchers, the CCW encrypts this identifier.

¹² CMS began using a new Medicare beneficiary identifier (MBI) in place of the HIC starting in 2018.

B. All Medicare HEDIS Measures RIF (File 1) Except for Plan All-Cause Readmissions

There is one record in the CCW HEDIS measures RIF (File 1 – SAS file naming convention is HEDIS_UTIL_YYYY) for each beneficiary enrolled in the contract during the year. If a beneficiary enrolls in more than one contract during the year, the beneficiary has more than one record in the file. These files are large since the number of MA beneficiaries is large.

The 2015 file (for MY 2015, submitted by MAOs in 2016) and the 2016 file use the same layout. Files change annually after that, with both measure categories and variables within the categories deleted and added. Refer to the record layout posted on the CCW [Data Dictionaries](#) website.

The CCW HEDIS RIFs include a small number of beneficiary demographic variables (e.g., sex, date of birth [DOB], state, ZIP code), contract and plan variables (e.g., contract and plan ID, plan type, organization type, and CMS region), and the number of member months the plan enrolled the beneficiary. The remainder of the variables in the files are the measure variables. The CCW team prepares various additional Medicare data files that users may request to augment the information in the HEDIS files (e.g., to obtain detailed information regarding a beneficiary or MAO characteristics). Additional information is in [Chapter 3](#).

Within a row of data, the beneficiary characteristics do not vary by measure (e.g., the beneficiary has only one sex, state, and zip code for the year). Similarly, the number of member months (member months contribution [MMC] variable called BENE_MMC) does not vary by measure.

If a beneficiary has more than one record in the HEDIS_UTIL_YYYY) RIF, then the contract, plan information, and BENE_MMC are different, and the beneficiary's geographic information may also be different (e.g., if the beneficiary moved to a different geographic location).

Within the file, the CCW team groups the variables together using the three-character alphabetical abbreviation for the measure category — and within each measure category, the denominators (or the population eligible for the measure) appear before numerators when applicable.¹³ Sometimes there is more than one numerator for a measure — reference [Table 1](#) for examples.

¹³ Some measures do not have topic-specific denominators; rather, they use an overall person-level denominator such as member months of enrollment.

Table 1. Example of SAS® naming convention for measure topics and associated metrics

Measure category	Measures	SAS variable for the denominator(s)	SAS variable for the numerator(s)
Adult BMI assessment (ABA)	ABA	ABA_DNMTR_IND	ABA_NMRTR_IND
Monitoring persistent medications (MPM)	Annual MPM — annual monitoring for members on angiotensin-converting-enzyme (ACE) inhibitors or angiotensin II receptor blockers (ARBs)	MPM_DNMTR_1_ACEI_ARB_IND	MPM_NMRTR_1_ACEI_ARB_IND
	Annual MPM — annual monitoring for members on digoxin	MPM_DNMTR_2_DIGOXIN_IND	MPM_NMRTR_2_DIGOXIN_IND
	Annual MPM — annual monitoring for members on diuretics	MPM_DNMTR_3_DRTCS_IND	MPM_NMRTR_3_DRTCS_IND

1. Types of Measures

There are different types of measures within the HEDIS file. Researchers use different calculations (formulas) depending on the type of measure with which they are working. At a high level, the types of measures include:

- **Logical or binary indicators** — have values (0,1); used to identify denominators or numerators (1 = yes, includes the beneficiary). Many of the variables in the file are “indicator” variables, including the examples in [Table 1](#). Within the RIF layout, CCW uses the SAS naming convention of *_IND for this variable type. Values of “1” or “0” indicate that the member was either included, or not included, in the numerator or denominator of the measure

Researchers often use these measures to calculate proportions or percentages (e.g., received a particular service among the enrollees eligible for the service, had adequate blood pressure control among enrollees diagnosed with hypertension, and had a certain level of HbA1c control among eligible diabetics). To calculate the performance rate for a study population, researchers sum (or count) the numerator divided by the denominator. For example, using variables from [Table 1](#), the formula for identifying the ABA rate is:

$$\text{ABA_NMRTR_IND} / \text{ABA_DNMTR_IND}$$

Since the numerator is binary, this calculation is analogous to taking the mean value of the numerator for a population a researcher limits by using the denominator indicator. An example of SAS code follows:

```
proc means data=HEDIS.HEDIS_UTIL_2015 N mean;
var ABA_NMRTR_IND;
where ABA_DNMTR_IND=1;
title 'Adult BMI Assessment (ABA) rate';
run;
```

- **Quantitative fields** — utilization or “count” measures; have one to three-digit numeric values to indicate the frequency of an occurrence. Measures with quantitative values include service measures, such as frequency of selected procedures (FSP) and ambulatory care (AMB). The submission shows a numerical value that indicates the

number of times the numerator of a measure includes the member. CCW uses the SAS naming convention of *_CNT for this type of utilization or “count” variables.

- The denominator for Use of Service measures is the number of months each Medicare member enrolled in the contract in the measurement year (variable called BENE_MMC). The BENE_MMC does not vary by measure and does not apply to the Effectiveness of Care or Risk Adjusted Utilization measures. The MMC only pertains to utilization measures. Each member should have a member month contribution value between 0 and 12.¹⁴ Examples of the use of service measures are in [Table 2](#).

Table 2. Example of SAS naming convention for Use of Service measure topics

Measure category	Measures	SAS variable for denominator(s)	SAS variable for numerator(s)
n/a	Beneficiary MMC to denominator	BENE_MMC	n/a
Ambulatory Care (AMB)	AMB numerator — outpatient visits	n/a	AMB_NMRTR_OP_VISIT_CNT
	AMB numerator — emergency department visits	n/a	AMB_NMRTR_ED_VISIT_CNT
Frequency of Selected Procedures (FSP)	FSP numerator — prostatectomy count	n/a	FSP_NMRTR_PROSTATCTMY_CNT
	FSP numerator — total hip replacement Count	n/a	FSP_NMRTR_TOT_HIP_RPLCMT_CNT
	FSP numerator — total knee replacement count	n/a	FSP_NMRTR_TOT_KNEE_RPLCMT_CNT
	FSP numerator — mastectomy count	n/a	FSP_NMRTR_MASTCTMY_CNT

To calculate the use of service rates for the study population, sum the numerator and divide by the sum of the denominator. CMS reports the overall total visits per 1,000 member months.¹⁵ There are two numerators for identifying the utilization of the AMB measure: AMB_NMRTR_OP_VISIT_CNT and AMB_NMRTR_ED_VISIT_CNT), as depicted in [Table 2](#). Conceptually, the formula for the overall AMB measure counts for a population is:

$$(\text{Sum of (AMB_NMRTR_OP_VISIT_CNT} \times 12)/\text{sum of BENE_MMC}) \times 1,000$$

CMS multiplies the visits (numerator) by 12 since dividing by member months adjusts the count to the number of visits *per month*. CMS expresses the measure results as “per 1,000 member years.” This means CMS multiplies the result by 1,000.

¹⁴ From 2016 HEDIS Patient-Level Data File Specifications File 1 of 2 01122016_5; downloaded from CMS website as the plan submission instructions: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData> (p.6) (Accessed 09/27/2024)

¹⁵ The CMS PUF also presents AMB rates stratified by the number of visits in the year (e.g., less than one, 1–9, 10–19, etc.).

The CCW team presents a SAS code example to calculate the overall outpatient visit rates that are summarized and output into contract-specific rates:

```
proc sort data=_ data=HEDIS.HEDIS_UTIL_2015 out=temp;
  by HEDIS_CNTRCT_ID;
run;

proc means data=temp noprint nway;
  by HEDIS_CNTRCT_ID;
  var AMB_NMRTR_OP_VISIT_CNT BENE_MMC;
  output out=amb_ov_sum(drop=_) sum=;
run;

data amb_ov_sum;
  set amb_ov_sum;
  per_1000=(( AMB_NMRTR_OP_VISIT_CNT*12) / BENE_MMC) *1000;
run;
```

When CMS reports the FSP in the CMS PUF,¹⁶ CMS expresses the findings as the number of procedures/1,000 member months of enrollment for age-specific and sex-specific strata. This means the record includes the numerator and denominator for only one of the age/sex-specific rates. Calculate age using HEDIS_BENE_DOB and reference the last date of the measurement year (the RFRNC_YR_HEDIS) to identify the age. The exact age groups CMS reports in the PUF may vary by measure.

For utilization measures with weights (in 2015, these include inpatient hospital utilization (IHU), emergency department utilization [EDU], and hospitalization for potentially preventable complications [HPC]), the count and weight (when applicable) variables follow the denominator (or eligible population) variables. CCW uses the SAS naming convention of *_WT for this variable type — reference [Table 3](#).

¹⁶ The CMS website has the MA HEDIS Public Use Files and associated documentation available for download (within annual ZIP files). <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/MA-HEDIS-Public-Use-Files> (Accessed 09/27/2024)

Table 3. Example of SAS naming convention for File 1 utilization measure with weights

Measure category	Measures	SAS variable for denominator	SAS variable for utilization metrics	SAS variable for predicted probability of discharge (PPD) weights	SAS variables for predicted unconditional count of discharge (PUCD) weights
Inpatient Hospital Utilization (IHU)	IHU	IHU_DNMTR_IND	n/a	n/a	n/a
	IHU: Observed IP Discharges, Surgery	n/a	IHU_IP_SRGRY_DSCHRG_CNT	IHU_PPD_SRGRY_CMRRBDTY_WT IHU_PPD_SRGRY_AGE_SEX_WT IHU_PPD_SURG_BASE_RISK_WT	IHU_PUCD_SRGRY_CMRRBDTY_WT IHU_PUCD_SRGRY_AGE_SEX_WT IHU_PUCD_SRGRY_BASE_RISK_WT
	IHU: Observed IP Discharges, Medical	n/a	IHU_IP_MDCL_DSCHRG_CNT	IHU_PPD_MDCL_CMRRBDTY_WT IHU_PPD_MDCL_AGE_SEX_WT IHU_PPD_MDCL_BASE_RISK_WT	IHU_PUCD_MDCL_CMRRBDTY_WT IHU_PUCD_MDCL_AGE_SEX_WT IHU_PUCD_MDCL_BASE_RISK_WT
	IHU: Observed IP Discharges, Total	n/a	IHU_IP_TOT_DSCHRG_CNT	IHU_PPD_TOT_CMRRBDTY_WT IHU_PPD_TOT_AGE_SEX_WT IHU_PPD_TOT_BASE_RISK_WT	IHU_PUCD_TOT_CMRRBDTY_WT IHU_PUCD_TOT_AGE_SEX_WT IHU_PUCD_TOT_BASE_RISK_WT

Since these metrics apply to contracts, the measure specifications call for summarizing the calculations for the applicable CMS contract ID. To calculate the observed hospitalization rate for the study population, begin by limiting the analysis to the eligible population (where IHU_DNMTR_IND=1). Then calculate the rates separately for surgical, medical, and total discharges. Within these three IHU strata (depicted in [Table 3](#)), the PUF reports the rates¹⁷ using six age-specific sub-strata. Conceptually the formula for observed acute inpatient surgical discharge rates is:

$$\text{Sum of IHU_IP_SRGRY_DSCHRG_CNT when IHU_DNMTR_IND=1}$$

Weights apply to some of the measures, like utilization variables. They are risk adjustment weights. Depending on the analyses investigators are conducting, CMS recommends using these risk adjustment weights so that the data accurately portrays the contract-level utilization.

C. HEDIS Readmissions RIF (File 2)

The CCW HEDIS readmissions file RIF (File 2 — SAS file naming convention is HEDIS_INP_YYYY) includes a small number of demographic (e.g., sex, age at discharge, state, and zip code) and contract/plan variables. The admission and discharge date for the stay (ADMSN_DT and DSCHRG_DT), and the readmission indicator (READMSN_IND), provide key information regarding the row of data. The remainder of the variables are for adjusting or weighting the discharges and readmissions.

Files for 2015–2017 (corresponding with MAO data submission years 2016–2018) have the same layout. In the 2018 file, CMS required six new variables (for socioeconomic status [SES] stratification variables), and in 2020 one field was retired and seven new fields added. Reference [Table 4](#).

Table 4. Count of variables in readmissions RIF over time

File 2 (data year)	2015–2017	2018	2019*	2020+
Number of variables	22	28	n/a	34

* No file available; MAOs did not submit data due to COVID PHE.

Each record (RECORD_ID) is for an inpatient stay for a beneficiary enrolled in an MA plan during the year. The record includes the associated admission and discharge dates, as well as a key variable, the readmission indicator (READMSN_IND). The readmission indicator has a value of 1 if this admission (IHS) has a readmission (numerator event), and the value 0 if this admission (IHS) does not have a readmission. There are also various weight variables [Table 5](#).

¹⁷ The CMS MA HEDIS Public Use Files and associated documentation available for download (within annual ZIP files). <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDENrolData/MA-HEDIS-Public-Use-Files> (e.g., 2016 PUF “HEDIS2016Doc”, p. 35) (Accessed 09/27/2024)

Table 5. Example of SAS naming convention for readmissions file weighting variables

Type of weighting variable	SAS variable for weights
Age/sex weight	AGE_SEX_WT
Base risk weight*	BASE_RISK_WT
Surgery weight	SRGRY_WT
Discharge weight	DSCHRG_WT
Comorbidity weight	CMRBDTY_WT

* Retired after 2018.

Starting with the 2018 file, there are six variables for the SES stratification, and they are:

- SES stratification — non-Part D low-income subsidy (LIS)/non-dual Medicare and Medicaid eligibility (DE), No disability
- SES stratification — LIS or DE
- SES stratification — disabled
- SES stratification — LIS or DE and disabled
- SES stratification — other eligible population
- SES stratification — unknown eligible population

Starting with the 2020 file, there are seven variables for discharge or transfer to skilled nursing care; one is an indicator variable, and the others are weights.

Chapter 3: Linkage with Other CCW Medicare Data

Researchers may use the HEDIS files alone or in combination with other CCW data products, such as the Medicare enrollment data, the plan characteristics file, or MA encounter data.

A. The Master Beneficiary Summary File (MBSF)

The Medicare enrollment file produced by the CCW is the Master Beneficiary Summary File (MBSF). CCW creates it using the CMS Common Medicare Environment (CME) database as its source. The MBSF includes many enrollment and other person-level variables contained in file “segments.” These segments are separate components of the file researchers may elect to request. On the CCW website, the [CCW Medicare Administrative Data User Guide](#) describes the segments available to researchers. The CCW website [Data Dictionaries](#) tab has a description of the variables contained in the MBSF.

The CCW team creates the MBSF for each calendar year. The MBSF contains demographic, entitlement, benefit, and plan enrollment data for beneficiaries who: 1) in CMS documents a being alive for some part of the reference year, and 2) enrolled in the Medicare program during the file’s reference year. Reference year refers specifically to the calendar year accounted for in the MBSF. For example, the 2018 MBSF covers the year 2018, which is the reference year. When joining information from MBSF to a HEDIS population, link using the BENE_ID. Presumably, all people in the MAO-submitted HEDIS files are in the MBSF.

The MBSF includes monthly Medicare Part C contract and plan IDs (variables called PTC_CNTRCT_ID_01–12 and PTC_PBP_ID_01–12) for the beneficiary, when applicable. The HEDIS file includes the CMS contract number for the plan that submitted the HEDIS data; analysts can find it as HEDIS_CNTRCT_ID. The HEDIS file also includes the CMS plan ID (HEDIS_PLAN_ID).

B. Plan Characteristics File

CCW produces an annual Medicare plan characteristics data file comprised of a six-file set per year with detailed information on MA plan characteristics. Researchers can join these plan characteristics to the MBSF to better understand the benefits available to enrollees. The files describe the plan benefit packages within each contract; therefore, researchers must use both the contract and plan identifiers.

Although nearly all MA plans include both Part C and Part D (the prescription drug benefit — or MA-PD plans), a small number of plans may be for Part C only. CCW includes an indicator in the plan characteristics base file that allows users to subset by Medicare program type (use the SAS variable called PTCD_INDICATOR; Part D plans are those where PTCD_INDICATOR = 2 [Part D only plan] or 3 [both Part C and D plan]).

The six plan characteristics files are:

1. **Plan benefit base file or “base” plan file** — contains key information about the managed care or drug benefit offered by the plan sponsor. Many of the variables in this file apply only to the Part D benefit and are 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 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** — useful to analysts interested in examining changes over time to the plans available to enrollees. Researchers can identify new plans, terminated in the prior year, or renewed or consolidated with other plans. The file includes information for all plans in the plan characteristics files for the current year or the preceding 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.

Additional details regarding this file are available in the [CCW Medicare Administrative Data User Guide](#) on the CCW website. The [Data Dictionaries](#) tab on the CCW website contains a description of the variables in the plan characteristics file.

C. Medicare Encounter RIFs

MA (Part C) encounter data RIFs are available to researchers starting with 2015. MAOs that provide services to beneficiaries under the Medicare Part C benefit, submit data to CMS, which 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.

Presumably, all hospital discharges in the MAO-submitted HEDIS plan all-cause readmission file (File 2) are in the MA inpatient (IP) encounter data. The people in both files have the same BENE_ID.

Chapter 4: Format, Content, and Encryption of CCW Output Files

The CCW provides the HEDIS files to academic researchers and certain government agencies to conduct approved research studies under a CMS Data Use Agreement (DUA). The CCW HEDIS data contain identifiable information and are subject to the Privacy Act and other federal government rules and regulations (reference the ResDAC website for information on requesting data <http://www.resdac.org/>). Anyone seeking approval for access to the CCW HEDIS data must ensure that the data can reliably support their research proposal.

Researchers have two options for accessing the data files — they may access them directly from the CCW Virtual Research Data Center (VRDC), or have their data shipped. For CCW VRDC users, the CCW team begins processing the data request upon receipt of the approved DUA and payment.

As stated previously, the file layout for each year may vary (although the layout for 2015 is the same as 2016; the rest of the HEDIS_UTIL_YYYY (File 1) layouts are different each year). Each year the layout specifically represents the HEDIS PLD data plans submitted to CMS in the calendar year following the MY, and the CCW files do not retain variables from a previous year that have been retired.

All RIF record layouts, codebooks, and this user guide, are available on the ccwdata.org website.

This section describes the content and format of the CCW HEDIS output package (the CCW data physically delivered to researchers). The CCW team delivers files to the researcher in the following format. [Table 6](#) lists descriptions of each of these contents. In this example, the CCW team assumes the researcher receives the MBSF file in addition to the HEDIS files.

Table 6. Format and naming convention for the CCW files

File	File description
readme_first_req000_2016.txt	This is a text file that describes the files contained in the output package. Filename example: readme_first_req232_2016.txt
res000011111req000232_2016_HEDISUTIL.exe	This is the executable that researchers must run to decrypt and uncompress the HEDIS all-measures (File 1) data file. In this example, 000011111 is the DUA number, 232 is the request number, and 2016 is the measurement year. This executable includes v8 SAS read-in programs, the .psv data file, and .fts file containing the layout and record counts.
res000011111req000232_2016_HEDISINP.exe	This is the executable that researchers must run to decrypt and uncompress the HEDIS inpatient/all-cause readmission (File 2) data file. In this example, 000011111 is the DUA number, 232 is the request number, and 2016 is the measurement year. This executable includes v8 SAS read-in programs, the .psv data file, and the .fts file containing the layout and record counts.
res000011111req000232_2016_MBSFABCD.exe	This is the executable that researchers must run to decrypt and uncompress the Part ABCD segment of the MBSF. In this example, 000011111 is the DUA number, 232 is the request number, and 2016 is the year of the data. This executable includes v6 and v8 SAS read-in programs, the .dat data file, and .fts file containing the layout and record counts.




HEDIS files*

File	File description
hedis_util_res000011112_req000161_2016.psv hedis_util_res000011112_req000161_2016.fts hedis_util_read_v8.sas	This set of files includes file1 — the HEDIS PLD All measures (except readmission) pipe-separated data file (.psv) file, the .fts (layout and record counts) file, and the version 8 SAS read-in programs.
hedis_inp_res000011112_req000161_2016.psv hedis_inp_res000011112_req000161_2016.fts hedis_inp_read_v8.sas	This set of files includes file 2 — the HEDIS PLD inpatient/plan all cause readmission pipe-separated data file (.psv) file, the .fts (layout and record counts) file, and version 8 SAS read-in programs.

* Beginning with data for MY2015 (submitted by MAOs in 2016).

In addition to the specific data files the researcher requested, CCW includes a variety of resource files in the deliverable package. [Table 7](#) shows these files.

Table 7. CCW resources accompanying data files

File	Description
 Code reference sets.xls	Code lists for ICD-9/ICD-10 diagnosis and procedure codes, HCPCS codes, Revenue Center, and other codes contained in the extracted files.
 Decryption instructions.pdf	This document contains instructions for decrypting/uncompressing the data files.
 Tips on getting started with data.pdf	This document contains tips for using the CCW data.

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

After the CCW team ships the data to the researcher, they send the password to decrypt the archive to the researcher via email. Each researcher request has a unique encryption. The CCW team never packages the password and the data media together. To decrypt the data files, the researcher accesses the email containing the decryption password. The data package contains detailed instructions for using this password.

Chapter 5. 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 directs the researcher to the appropriate person. If researchers 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 (excluding most federal holidays). Contact the CCW Help Desk at ccwhelp@ccwdata.org or 1-866-766-1915.

Appendix A — List of Acronyms

Acronym	Definition
ABA	Adult BMI assessment
AMB	Ambulatory care
CCW	Chronic Conditions Warehouse
CMS	Centers for Medicare & Medicaid Services
DE	Dual eligibility (Medicare and Medicaid)
DUA	Data Use Agreement
EDU	Emergency department utilization
EOC	Effectiveness of care
FFS	Fee-for-service
FSP	Frequency of selected procedures
HEDIS	Healthcare Effectiveness Data and Information Set
HIC	Medicare health insurance claim number
HPC	Hospitalization for potentially preventable complications
IHU	Inpatient hospital utilization
IP	Inpatient
LIS	Low-income subsidy
MA	Medicare Advantage
MAO	Medicare Advantage Organizations
MBI	Medicare Beneficiary Identifier
MBSF	Medicare Master Beneficiary Summary File
MDS	Minimum Data Set
MMC	Member months contribution
MPM	Monitoring persistent medications
MY	Measurement year (also known as the year of the data file)
NCQA	National Committee for Quality Assurance
OASIS	Outcome and Assessment Information Set
PCR	Plan all-cause readmissions
PHE	Public health emergency (COVID-19)
PLD	Patient-level detail
PPD	Predicted probability of discharge
PUCD	Predicted unconditional count of discharge
PUF	Public Use File
RIF	Research Identifiable File
SES	Socioeconomic status
SNP	Special-needs plan
VRDC	Virtual Research Data Center