

Chronic Conditions Warehouse

Your source for national CMS Medicare and Medicaid research data



Chronic Conditions Warehouse Technical Guidance

Calculating Medicare Population Statistics

MAY 2022 | VERSION 1.5

This page intentionally left blank.

Revision Log

Date	Changed by	Revisions	Version
May 2022	K. Schneider A. Sisco	Adjusted references to conditions data files to reflect the addition of the 30 CCW Chronic Conditions file (MBSF_CHRONIC)	1.5
July 2021	D. Happe	Changed CMSData@gdit.com to cmsdata@gdit.com	1.4
January 2021	K. Schneider R. VanGilder	Moved to the new template and refreshed code examples to use current SAS files	1.3
January 2018	K. Schneider	MBSF A/B references updated to include C/D; updated data source information to reflect CME instead of EDB	1.2

Table of Contents

Overview	1
A. List of CCW Chronic Conditions classifications	2
B. List of the CMS OTCC.....	3
Chapter 1. CCW Medicare Data	5
A. MBSF A/B/C/D Segment.....	6
B. MBSF CCW Chronic Conditions Segment	6
C. MBSF CMS OTCC.....	7
Chapter 2. Calculating Population Statistics	8
A. Measures.....	8
1. Numerators.....	8
2. Denominators	9
B. Analysis of Rates.....	10
1. Description of Options	10
2. Calculation of Options.....	11
Chapter 3. Results	17
A. Next Steps	17
Chapter 4. Generalizing These Methods	18
A. Limitations.....	18
Chapter 5. Further Assistance with CCW Data	20
Appendix A — List of Acronyms and Abbreviations	21
Appendix B — SAS Code	22

List of Tables

Table 1. Four methods for calculating the prevalence of AMI in 2018 (5% random sample)	17
--	----

List of Code Examples

Code Example 1. Define a sample of beneficiaries with full or partial Medicare FFS coverage using variables from MBSF.....	13
Code Example 2. Define a sample of beneficiaries with Medicare FFS coverage at a point in time	15
Code Example 3. Determine total member months of Medicare FFS coverage and the corresponding condition numerator	16

Overview

The Centers for Medicare & Medicaid Services (CMS) launched the Chronic Conditions Warehouse (CCW), a research database, in response to the Medicare Modernization Act of 2003 (MMA). Section 723 of the MMA outlined a plan to improve the quality of care and reduce the cost of care for chronically ill Medicare beneficiaries. An essential component of this plan was establishing a data warehouse that contains Medicare claims data and assessments, linked by a beneficiary, across the continuum of care.

The CCW contains fee-for-service (FFS) institutional and non-institutional claims, assessment data, and enrollment/eligibility information from January 1, 1999, forward. CCW links data by a unique, unidentifiable beneficiary key, allowing researchers to analyze information across the continuum of care. CCW data 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. The CCW makes it easy to study chronic diseases by incorporating variables for common chronic conditions and other chronic or potentially disabling conditions (OTCC), identifying additional chronic health, mental health, and substance abuse conditions. These condition variables are available for two different sets of condition algorithm specifications: Chronic Conditions and OTCC. Much of the supporting documentation is available on the CCW website (e.g., chronic condition definitions, data dictionary, etc.). Reference www.ccwdata.org.

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

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, Research Data Assistance Center (ResDAC) used CMS and other national data sources to identify candidate conditions that they could define using claims-based algorithms. Next, ResDAC conducted extensive literature reviews to gather diagnosis code sets for each candidate’s condition. Finally, CMS engaged other federal agencies in a series of conversations to vet the proposed category definitions.
2. 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. The algorithm specifications are on the [Chronic Conditions](#) page of the CCW website.
3. OTCC — CMS developed additional measures for CMS OTCC specifically to enhance the Medicare-Medicaid dually enrolled population’s research. These additional conditions include variables for mental health and tobacco use conditions, developmental disorder and disability-related conditions, and other chronic physical and behavioral health conditions. Reference the OTCC algorithm specifications on the [Other Chronic Health, Mental Health, and Potentially Disabling Conditions](#) page of the CCW website.

The CCW team designed the data files to support a variety of research objectives. The predefined conditions make it easy to select a study population with a condition of interest. Researchers may refine the clinical or coverage criteria for their specific hypotheses. CCW uses Medicare claims-based utilization information to make the condition determinations (i.e., an indicator that the beneficiary received a service or treatment for the condition of interest).

We list the 30 CCW Chronic Conditions below. The algorithms for identifying these condition categories are available on the [Chronic Conditions](#) page of the CCW website.

A. 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.

Below is the list of the CMS OTCC classifications. The algorithms for identifying the condition categories are available on the [Other Chronic Health, Mental Health, and Potentially Disabling Conditions](#) page of the CCW website.

B. List of the CMS 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

* In addition to the aggregate Hepatitis condition variable, the MAX and MMLEADS V2 data files (through 2012) have fields for various hepatitis subtypes (e.g., Hepatitis A, Hepatitis B [acute], Hepatitis B [chronic]).

**In addition to the HIV/AIDS condition variable, the MAX and MMLEADS data files have fields for various HIV/AIDS screening and services, including several “monitoring” variables (largely consisting of Medicaid services).

The variables for the CMS OTCC are similar to those for the CCW conditions in that the CCW team generates them using claims-based algorithms, which indicate that treatment for a condition appears to have taken place. The algorithms examine Medicare services patterns that serve as a proxy indicating that a beneficiary is likely receiving treatment for the condition.¹ You can find more information about these algorithms, including the literature references and exact codes and claim types used to identify each condition, on the CCW’s website [Other Chronic Health, Mental Health, and Potentially Disabling Conditions](#) page.

It is important to note that the major objective of the chronic condition indicator variables is to extract the relevant clinical cohorts easily from a very large database. Therefore, definitions are somewhat broad to allow more researchers to request data extractions based on these definitions — then refine the specifications as needed to fit their own data

¹ A version of the CCW Chronic Conditions and CMS OTCC that examines both Medicare and Medicaid data for those dually enrolled in Medicare and Medicaid is available in the MMLEADS data product.

needs. Researchers communicate their specifications using the CCW Specifications Worksheet part of the CMS data request packet. For this reason, the counts published in [Medicare data tables and charts](#) on the CCW website for the various chronic conditions are likely overestimates of the useable sample for a given research project. Correspondingly, it would not be accurate to use these chronic condition counts to calculate population statistics for these conditions without first narrowing the counts further by using various criteria.

If you want to have a control group, you can take some simple steps to identify the appropriate controls. First, researchers should consider various potential control factors, such as the presence or absence of other conditions, whether current treatment or “ever” treatment for a comorbid condition is of interest, the length of observation or surveillance periods, and certain types of Medicare coverage.

The intent of this paper is to guide researchers to:

1. appropriately identify a study group based on clinical criteria (or combination of clinical and coverage specifications), and
2. define an appropriate denominator for rate calculations, if applicable.

A list of acronyms used in this document appears in [Appendix A — List of Acronyms and Abbreviations](#).

Chapter 1. CCW Medicare Data

The CCW contains data for 100% of Medicare beneficiaries, including enrollment data and FFS claims from 1999 forward, and supports the extraction of standard samples of beneficiaries, such as:

- the CMS 5% random sample,
- an *enhanced* 5% sample (i.e., once a data user selects a beneficiary, they remain part of this sample in all subsequent years; once-in-always-in rule, from 1999 forward), or
- a user-defined cohort that the CCW team can extract from the CCW 100% Medicare files.

The CCW database contains data for all eligible beneficiaries; it does not limit to just those with a chronic condition. Managed care encounter data is available for people enrolled in Medicare Advantage (MA) plans, starting in 2015; however, the chronic condition algorithms currently do not consider encounter data. Therefore, you should view the CCW as a source of utilization and chronic condition information primarily for the FFS population. Assessment information is available for all who have received the prerequisite services and this limitation does not affect it.

The predefined chronic conditions use FFS claims-based definitions; therefore, there is no opportunity to determine whether Medicare managed care enrollees have 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 has always disseminated files that include data regarding Medicare enrollment. Historical files include the CMS Denominator File, the CCW Beneficiary Summary File (BSF), the Beneficiary Annual Summary File (BASf), and the Master Beneficiary Summary File (MBSF) that used the CMS Enrollment Database (EDB) as its source. Currently, the preferred enrollment data file is the MBSF, which uses the CMS Common Medicare Environment (CME) database as its source. The transition from the EDB to the CME database as the source for the MBSF enrollment and demographic information offers several advantages that better meet researchers' evolving data needs. The CCW team describes the rationale and impact of this data conversion in a CCW white paper, [Medicare Enrollment: Impact of Conversion from EDB to CME](#).

Like the original MBSF, the current MBSF using the CME enrollment data file contains a constellation of enrollment and other person-level variables in “segments” — that are separate components of the file researchers may purchase. You can find a description of the variables in the MBSF on the [Data Dictionaries](#) page of the CCW website.

Created annually, the MBSF contains demographic entitlement and enrollment data for beneficiaries who: 1) were part of the user-requested sample; 2) the data documented as being alive for some part of the reference year; and 3) 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 2018, the reference year.

The current MBSF — Part A/B/C/D segment, also known as the Base Beneficiary Summary file, consists of variables that identify monthly Medicare Part A, B, C, and D enrollment status and other key demographic and coverage variables. The CCW derives all fields regarding beneficiary demographics or enrollment in the MBSF from the CMS CME database (also the source for the CMS EDB). The MBSF updates each month, then at the end of the calendar year, allows a full year of additions and updates before finalizing. For example, the CCW team finalized 2018 data in December 2019 and then extracted it to populate the MBSF.

The additional segments of MBSF are: 1) CCW Chronic Conditions (file for the 27 CCW Chronic Conditions is called MBSF_CC; file for the 30 CCW Chronic Conditions is called MBSF_CHRONIC), 2) CMS OTCC (file called MBSF_OTCC), 3) Cost and Use (file called MBSF_CU), and 4) National Death Index (NDI) (file called MBSF_NDI).

The Base Beneficiary Summary file (also called the Medicare Part A, B, C, and D enrollment segment) consists of variables related to Medicare Part A, Part B, and Part D enrollment data. The CCW Chronic Condition segments and the CMS OTCC segment of the MBSF contain variables that indicate the presence of claims for common or chronic conditions using claims-based algorithms (as a proxy for evidence of treatment/identification of a condition).

A. MBSF A/B/C/D Segment

The Base MBSF is a calendar year file. It contains demographic, geographic, eligibility and Medicare Part A, B, C, and D coverage information for all alive and eligible beneficiaries for any part of the year.

This file contains a variable to ascertain whether it includes a beneficiary in the CMS 1%, 5%, or 20% sample for the year (i.e., the annual cross-sectional sample; variable called `SAMPLE_GROUP`). The file includes the `ENHANCED_FIVE_PERCENT_FLAG` variable, which indicates whether the data includes the beneficiary as a member of the *enhanced* CCW 5% sample (i.e., all-inclusive, *ever* included in the 5% sample, from 1999 forward). Researchers may use this annual person-level summary file to determine whether a beneficiary has a sufficient surveillance period (i.e., months of Medicare coverage) for inclusion in the study. Variables contained in this file include the number of months of Medicare Part A, Part B, or managed care coverage, whether the beneficiary died during the year, as well as other beneficiary demographic and geographic information. Reference the record layout on the [Data Dictionaries](#) page of the CCW website.

B. MBSF CCW Chronic Conditions Segment

The CCW Chronic Conditions segment of the MBSF contains summarized clinical information for all beneficiaries included in the requested cohort. 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 two different MBSF conditions files:

1. the 27 CCW Chronic Conditions (available 1999-2020²; delivered as the `MBSF_CC_YYYY` file)
2. the 30 CCW Chronic Conditions (2017 forward; delivered as the `MBSF_CHRONIC_YYYY` file)

These files include yearly indicator variables and the first “ever” dates for each chronic condition; the MBSF file for the 27 CCW Chronic Conditions also includes a mid-year variable for each condition. Additional details regarding these variables are in the [CCW Medicare Administrative Data User Guide](#). These variables include:

1. **Yearly indicator (or end-of-year indicator)** — indicating whether the beneficiary met each of the chronic condition definitions during the respective time period ending December 31, YYYY
2. **Mid-year indicator (available only in the `MBSF_CC_YYYY` file)** — may be useful if researchers are using a July 1 timeframe
3. **First occurrence date (or “ever” date)** — indicating the date the beneficiary first met the specifications for the condition. For the `MBSF_CC` and `MBSF_OTCC`, the earliest possible date is 01/01/1999. However, for the 30 CCW conditions, available starting with the `MBSF_CHRONIC_2017`, the earliest possible date is 01/01/2016.).

² 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.

Note that the CCW team defines these chronic condition fields by looking at a pattern of medical care utilization, as determined by Medicare FFS claims. Reference the record layout and codebooks for each file on the [Data Dictionaries](#) page of the CCW website.

C. MBSF CMS OTCC

The MBSF CMS OTCC segment of the MBSF contains summarized clinical information for all beneficiaries included in the requested cohort. This file includes two variables for each of the conditions. These variables include: 1) a yearly indicator — indicating whether the beneficiary met each of the condition definitions during the respective time period ending December 31, YYYY, and 2) a first occurrence date — indicating the date the beneficiary first met the specifications for the condition (**NOTE:** 1999 is the earliest year that will appear in this field). The CCW team defines these condition fields by looking at medical care pattern utilization, as determined by Medicare FFS claims.³ Reference the record layout on the [Data Dictionaries](#) page of the CCW website.

Refer to the CC definitions document for more details regarding reference periods and clinical specifications for individual CC definitions on the [Other Chronic Health, Mental Health, and Potentially Disabling Conditions](#) page of the CCW website.

³ A version of these condition variables that examines both Medicare and Medicaid enrollment and claims data is available in MMLEADS. This information may benefit projects involving the dually enrolled (Medicare-Medicaid) population.

Chapter 2. Calculating Population Statistics

A. Measures

The CCW data files will likely contain a slightly broader cohort of beneficiaries and claims than you need for your particular study. Researchers should consider key issues for both numerators and denominators, presented below, before identifying a final study cohort.

1. Numerators

The group of beneficiaries represented in the numerator typically consists of those with particular clinical conditions — or those who have received certain services. The condition algorithms examine service patterns in claims data, which indicate that a beneficiary is likely receiving treatment for the condition. Determining whom to include in your study group is a crucial first step in study design. The CCW allows for much flexibility in terms of refining your selection criteria easily. There are two⁴ types of condition indicator variables to consider:

- Yearly indicator** — the first option is to use the yearly indicator for the condition of interest, assuming you have an interest in one or more of the predefined CCW Chronic Conditions or the Chronic or Potentially Disabling Conditions. These indicators, which consist of a separate field for each condition (e.g., AMI, ALZH, HIP_FRACTURE), are located in the CCW Chronic Condition segments of the MBSF (MBSF_CC or MBSF_CHRONIC) and/or the CMS Other Chronic or Potentially Disabling Conditions segment (note that variable naming convention in the latter file uses “*_MEDICARE” in the file name; e.g., AUTISM_MEDICARE). The value within each field indicates whether the beneficiary received services during the time frame to indicate treatment for the condition (i.e., condition variables indicate the presence of treatment for common or chronic conditions using claims-based algorithms as a proxy for evidence of a condition). The same variable also indicates whether one can observe the beneficiary for the full surveillance period — or until the date of death (i.e., based on Medicare coverage criteria — full Part A and Part B coverage, and no HMO).

Each yearly indicator uses December 31 as the end of the reference year (e.g., the 2018 yearly indicator for an algorithm with a one-year reference period includes services between 01/01/2018–12/31/2018). The following are valid values for the yearly indicator for each of the conditions:

- 0 = Beneficiary did not meet claims criteria or have sufficient 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)

Using information from the four values (0–3) in each condition indicator field, researchers may efficiently extract their cohort of interest. Condition indicator values of 3 mean that the pattern of utilization indicates the beneficiary was likely receiving services for the condition. The beneficiary had Medicare Part A and B coverage — and no HMO

⁴ For the 27 CCW Chronic Conditions file (available in the MBSF_CC), three variables per condition per year (the midyear indicator is also available).

coverage for the entire surveillance period — or until death (i.e., anywhere from one to three years, depending on the condition of interest).

The value of 1 means that the pattern of claims indicates the beneficiary met the claims criteria for the condition; however, the beneficiary did not meet the entire eligibility criteria (not observed for the full FFS surveillance period). This limited surveillance period could be due to new accretion into the Medicare program (i.e., beneficiaries who became newly eligible), a break in Part A or B coverage, or one or more months of managed care coverage. The CCW team routinely delivers to researchers the claims and/or assessment information for values of three and one for the requested condition unless the researcher specifies otherwise.

The other two potential values in this field indicate the *absence* of sufficient claims that met the condition criteria during the reference period (condition indicator for the condition = 2 or 0). A value of 2 indicates the beneficiary had Medicare Part A and B coverage and no HMO coverage throughout the full surveillance period. A value of 0 indicates the beneficiary did not meet this coverage criteria. In both cases, there were either no claims or an insufficient number of claims (in cases where the beneficiary required more than one claim) to indicate current treatment for that particular condition.

Researchers will need to determine whether to include only beneficiaries with claims for a condition and the full surveillance period (i.e., the 3s) — or whether it is valuable to retain some or all of the beneficiaries who appear to have claims for the condition — even though coverage for the full surveillance period may be lacking (i.e., the 1s). By merging the MBSF_CC (or MBSF_CHRONIC or MBSF_OTCC) file with MBSF_ABCD file (using BENE_ID), one can efficiently determine the level of coverage during the year for each of the beneficiaries. Find more information regarding making cohort selections based on Medicare coverage criteria in this paper’s [Denominators](#) section.

- **First occurrence date** — another option for determining who has a condition is identifying whether the beneficiary *ever* had the condition. You can find these “ever” fields, a separate one for each of the CCW Chronic Conditions, in the Chronic Condition segments of the MBSF; the “ever” fields for the CMS Other Conditions are in the Other Chronic or Potentially Disabling Conditions segment of the MBSF. The presence of a date in the field (e.g., AMI_EVER) — indicates the date the beneficiary first met the *clinical FFS 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, available starting with the MBSF_CHRONIC_2017, the earliest possible date is 01/01/2016. A null value indicates this diagnosis is not in the FFS claims during the file surveillance period (1999 or 2016), or since the beneficiary became eligible for Medicare if the first coverage date occurred after that (i.e., reference the COVSTART variable in the MBSF). Note that beneficiaries who had a yearly value that indicated they met the claims criteria (i.e., either 1 or 3) will have an “ever” date for the condition.

2. Denominators

Determining who is “at-risk” for the events of interest is an important next step in the research process. Observing FFS claims that may indicate the presence (or confirm the absence) of treatment for the condition, can only occur if there is some FFS coverage exposure period. Beneficiaries have various Medicare coverage options and may not have FFS for the full surveillance period of interest. It is up to the researcher to decide how much coverage to include in the study.

We wish to caution researchers against using the total number of beneficiaries included in the MBSF (or even a particular subsample such as the 5% random sample) as a denominator for any condition prevalence calculation. This number is an overestimate of the number of beneficiaries with FFS and at risk for the condition at any point in time (refer to [Chapter 4: Results](#) section of this paper for an illustration). Please be deliberate in making your denominator selection to produce accurate rates.

Researchers may employ several variables in the MBSF (A/B/C/D segment) to specify which beneficiaries to include in the study. For example, you may wish to limit cohorts to certain ages, geographic locations, or a particular gender. Researchers may also want to retain beneficiaries who meet certain coverage criteria. Some of these variables include:

- **Coverage variables** — whether Medicare Part A and/or Part B during a particular month covered the beneficiary (variables BENE_MDCR_ENTLMT_BUYIN_IND_01–12)
- **Dual eligibility variables** — whether the beneficiary was eligible for both Medicare and Medicaid during the month (i.e., dually eligible; monthly variables DUAL_STUS_CD_01–12)
- **Managed care variables** — whether the beneficiary had Medicare managed care coverage (monthly variables BENE_HMO_IND_01–12)
- **Date of date** (BENE_DEATH_DT)

The CCW team creates the MBSF and all segments annually; therefore, researchers should request each calendar year of interest. Each MBSF A/B/C/D segment provides the demographic and coverage information for a given year. For researchers interested in requesting multiple years of data, they should also request multiple years of the MBSF (e.g., if you request three years of claims data, then request three years of MBSF).

B. Analysis of Rates

There are many different types of rates that you can construct using the condition indicator variables. The CCW team primarily intends this paper to discuss options for calculating population-based rates for chronic conditions.

1. Description of Options

Do not interpret the types of rates described as incidence rates. They are the prevalence of FFS claims that met the clinical criteria (e.g., identification/treatment) for a clinical condition at a point in time, using treatment/receipt of services as a proxy for having the condition of interest (e.g., claims-based point-prevalence).

These analytic options all include some examination of the extent and duration of Medicare coverage. Researchers may wish to look at enrollment in Medicare Part A and Part B since most of the condition definitions include either or both types of services (e.g., a combination of inpatient, outpatient, and/or Carrier claims). However, for a couple of conditions (i.e., AMI or hip fracture), one inpatient claim is sufficient to indicate the presence of the condition. For these conditions, the researcher should ponder whether to include subjects without Part B coverage (e.g., for studies related to follow-up care, researchers may wish to retain only subjects who also enroll in Part B or include only subjects with Part A coverage for a study related to inpatient care).

In addition, since FFS claims are central to CCW's algorithms to identify beneficiaries with conditions, the researcher may wish to allow little or no Medicare managed care coverage in the surveillance period. This will allow sufficient opportunity to experience an FFS Medicare claim, indicating the utilization of interest.

Researchers interested in calculating population-based rates (as opposed to rates of events within the study cohort — such as readmission or mortality rates) may find it desirable to use the random 5% sample (or some other randomly-selected population), as the file sizes are more manageable and require the minimum data necessary to perform the research study.

The number of beneficiaries represented in the rate calculations depends on how restrictive or lenient the researcher wishes to be in inclusion criteria. We contrast the four denominator options below. You should apply the same coverage restrictions to both numerator and denominator (**The CCW team recommends the researcher selects the population first — then, from this sample, determines who has the condition[s] of interest**):

- **Full FFS coverage** — the beneficiary has both Medicare Part A and Part B with no HMO coverage for the full surveillance period (or until the time of death if the beneficiary died in the year). This type of restriction limits analysis to those with full FFS coverage, which means that the data may exclude some beneficiaries identified as having claims for the condition of interest (i.e., because there are sufficient claims to indicate the presence of treatment for the condition) from the study. This is the most restrictive type of coverage option. This type of rate is the simplest to compute, as data values exist to provide numerators and denominators for the condition variables. However, the researcher should not presume that this cohort is either representative of the Medicare population as a whole, or the Medicare FFS population, as beneficiaries with full (all 12 months of FFS Part A and B) coverage may not be “typical” or representative of all Medicare consumers.
- **Partial FFS coverage** — the beneficiary has some Medicare Part A and/or B coverage (or much coverage, depending on how one specifies the coverage criteria) and may or may not have some managed care coverage. This option allows for a break in coverage and is less restrictive than the “full FFS coverage” option above. A common recommendation is to allow for a one-month gap in coverage per year of surveillance. In fact, by requiring 11 or 12 months of coverage (i.e., full or nearly full FFS), you will exactly match the rate calculations on the CCW website.⁵ This is an attractive option to avoid losing any/many cases with claims for the condition of interest (i.e., known cases, as indicated in claims) due to partial FFS coverage.
- **Point in time FFS coverage** — the beneficiary has FFS coverage during the month of interest — (e.g., for July, the midpoint of the year). This is an appealing option for identifying enrollment and disease burden for a typical point in time (i.e., a month).
- **Person years with FFS coverage** — this option allows each beneficiary with any FFS coverage to count toward the denominator for however many months (or proportion of a year) they have the coverage of interest. When you use a denominator such as this, all cases are in the numerator, and the denominator includes their corresponding time “at risk.” This option is attractive since beneficiaries are “under FFS observation” for various lengths of time. Technically, you do not produce rates using this type of calculation — rather, you express the extent of illness at a ratio of cases to the time at risk.

2. Calculation of Options

To capture the prevalence rates accurately using the CCW data, it requires several analytic steps. We present analytic guidance for calculating these four different types of prevalence rates. We offer the results of each type of analysis in the **Results** section of this paper so researchers may assess the empirical difference in numerators, denominators, and rates that these analytic variations produce.

We share partial SAS® code examples below to illustrate various calculation methods. We present the complete analytic code for each method in [Appendix B — SAS Code](#). Throughout, we limit our analyses to the CMS random 5% sample; our examples use 2018 data files. Since the CCW data files are very large, we recommend retaining as few rows of data and as few variables as possible to allow the programs to run efficiently. You may wish to review additional CCW Technical Guidance papers to learn more about options for specifying analytic cohorts; for example, reference the [Getting Started with CMS Medicare Administrative Research Files — A Technical Guidance Paper](#) on the CCW website.

Using the MBSF for the year of interest, retain only beneficiaries selected as part of the random 5% sample for that year (i.e., using the SAMPLE_GROUP variable — where the value = ‘01’ or ‘04’). Keep only the subset of variables of interest.

⁵ Reference “Medicare Chronic Condition Statistics” [Tables B2a and B2b](#) on the CCW website.

Next, merge this restricted MBSF A/B/C/D segment with the CCW Chronic Conditions segment and/or CMS Other Chronic or Potentially Disabling Conditions segment (linkage variable = BENE_ID) for the same year and retain only the beneficiaries in the population selected from the MBSF.⁶ You are ready to examine prevalence rates using a combined file, which we create in the sample SAS code provided later in the paper (reference [Code Example 1](#); and complete code is in [Appendix B — SAS Code](#)).

- **Full coverage** — the simplest type of prevalence rate to calculate using the CCW data. For this option, you may simply use variables from the MBSF Chronic Conditions segment for the year of interest (note: although there is not a need to link with the MBSF A/B/C/D to incorporate coverage information, we recommend you limit the cohort to the random 5% sample first — as described above).

Once you have identified your condition variable of interest, you may calculate this rate using certain values within the data element.

$$\begin{aligned} \text{Denominator} &= [2\text{s}] + [3\text{s}] \\ \text{Numerator} &= [3\text{'s}] \end{aligned}$$

For example, suppose we are interested in AMI as a condition. In that case, once we have subsetting the data to obtain a strict 5% file, we simply need to look at the AMI variable (which is called AMI) — and keep the 3s and 2s (both indicate full Medicare FFS coverage — and the presence [3s] or absence [2s] of the condition in the claims). The prevalence rate = ratio of (3) / (3 + 2). Note that this calculation is identical to using the code in the example for Partial coverage (reference [Code Example 1](#), below) if you edit the code to require a full 12 months of coverage.

- **Partial coverage** — for this option, our example allows for a one-month break in Part A and/or Part B coverage, and up to one month of managed care coverage per year of surveillance (e.g., retain beneficiaries with 11 or 12 months of Part A, B, no HMO for one-year conditions; similarly, for two-year conditions you might retain beneficiaries with 22 out of 24 months of FFS coverage). Reference [Code Example 1](#), below.

Another important step for this calculation is to keep beneficiaries fully covered (Part A/B FFS) until the time of death (or covered for all but a month before the time of death). Failure to have an extra analytic step to include those who died will likely result in an undercount for both the numerator and denominator in meaningful ways (e.g., you would fail to count fatal AMIs which occurred in any month other than at the end of the year; a death could result in very few months of Medicare coverage for the beneficiary).

Using the **MBSF_ABCD** file (for the year of interest), you will need to combine information from three key variables:

1. A series of 12 variables (one for each month) called MDCR_ENTLMT_BUYIN_IND_01–12, contain the information regarding Part A and Part B coverage for each month during the year. For this example, we count only the beneficiaries with Part A and B coverage (regardless of whether they have the state buy-in). For this series of variables, we want to keep values of 3 (indicating the beneficiary has Part A and B) and C (indicating the beneficiary has state buy-in for Part A and B).
2. Next, we assume you will also want to *exclude* beneficiaries who have more than one month of managed care coverage during the timeframe. A series of 12 variables (one for each month) called HMO_IND_01–12, contain the managed care information. Any value other than 0 or 4 indicates some type of managed care coverage during the month. Hence, we exclude these beneficiaries (the 0s and 4s) from consideration in our algorithms.

⁶ This guidance applies to any of the three MBSF conditions files.

- The BENE_DEATH_DT field contains the information regarding which beneficiaries died during the year, and when the death occurred.

To construct the denominator and numerator that include beneficiaries with partial coverage, and those covered until death, we include an example of analytic code (reference [Code Example 1](#)) using SAS programming language. You may adapt this code to use whatever software you prefer.

Code Example 1. Define a sample of beneficiaries with full or partial Medicare FFS coverage using variables from MBSF

Define a sample of Medicare FFS beneficiaries with a specified duration of A and B FFS coverage, or coverage until the month of death. The first portion of the code uses the monthly coverage variables to require at least 11 months of FFS coverage; the second portion of the code identifies beneficiaries with the condition of interest (e.g., AMI). The input data source is the MBSF — only key variables from the A/B/C/D segment and the CCW Chronic Conditions segment.⁷ We read in only the data from the random 5% sample. All of our examples use data from 2018 (**NOTE:** within each MBSF file segment, the variables and variable names are the same over time).

The following SAS code assumes that you wish to create a data file to use as an example for calculating prevalence; it should have merged Medicare coverage and Condition variables for your population of interest, keeping only the necessary study variables.

We begin by developing a cohort for population statistics (called *temp1*) by using Medicare coverage specifications. Then you may use these variables to calculate rates for your population (e.g., FFS only).

```
data temp1;
merge MBSF.MBSF_ABCD_2018(keep=BENE_ID mdcr_entlmt_buyin_ind_01 -
mdcr_entlmt_buyin_ind_12 hmo_ind_01 - hmo_ind_12 bene_death_dt SAMPLE_GROUP)
MBSF.MBSF_CC_2018 (keep=BENE_ID AMI AMI_EVER);
by BENE_ID;
if SAMPLE_GROUP in ('01','04');
```

***note this input file is a merged MBSF A/B/C/D and MBSF_CC Conditions File Segment;**

*** determine # months of Part A, B, and no HMO coverage;**

```
array MemberMos_AB (12)
mdcr_entlmt_buyin_ind_01 - mdcr_entlmt_buyin_ind_12;
array MemberMos_noHMO (12) hmo_ind_01 - hmo_ind_12;
array Member_FFSMos (12) Member_FFSMos01 - Member_FFSMos12;
do i= 1 to 12;
if MemberMos_AB(i) in ('3','C') and MemberMos_noHMO(i) in
('0','4')then Member_FFSMos(i)=1;
else if MemberMos_AB(i) NOT in ('3','C') or MemberMos_noHMO(i) NOT in
('0','4')then Member_FFSMos(i)=0;
Member_Mos=sum(of Member_FFSMos:);
end;
```

⁷ Although the code examples use the MBSF_CC file, researchers may use MBSF_CHRONIC or MBSF_OTCC, depending on the condition(s) of interest.

*** note – determine who had 11- or 12-months coverage or coverage until month before death;**

```
if (bene_death_dt=. and Member_Mos in (11,12)) or (bene_death_dt~=.
and month(bene_death_dt)<=Member_Mos+1 and Member_mos~=0) then Part1_Cov=1;
else Part1_Cov=0;
```

*** note – bring in numerator information for AMI – keep both 3s and 1s;**

```
if Part1_Cov=0 then AMI_PartRT=.;
else if Part1_Cov=1 and (ami= 0 or ami=2)then AMI_PartRT=0;
else if Part1_Cov=1 and (ami= 1 or ami=3)then AMI_PartRT=1;
label
Part1_Cov = '11 or 12 months FFS no HMO - except for those who died'
Member_Mos = 'Total Member months of A B and No HMO - per bene'
AMI_PartRT = 'Had AMI - partial coverage';
run;
```

To compute the partial coverage rate, we simply aggregate the AMI_PartRT variable created in the data step using the means procedure. The code below will produce three outputs: N will be the rate denominator, SUM will be the rate numerator, and MEAN will be the rate.

```
proc means N SUM MEAN;
var AMI_PartRT;
run;
```

With a few minor changes, you may use this analytic code as the basis for code that can specify the other types of cohorts discussed in this paper. You can easily modify this SAS code to fit your denominator specifications (e.g., two-month break in coverage; not requiring Part B coverage — only Part A coverage).

NOTE: The [condition statistics](#) on the CCW website (i.e., Medicare Table B.2) use the Partial coverage algorithm and require 11 or 12 months of coverage for inclusion in the denominator.

- **Point in time coverage** — for this type of a denominator, we need to determine which Medicare beneficiaries were alive and had Part A and Part B coverage, and no HMO coverage, during our month of interest. For our example, we use the year’s midpoint, and assess who has coverage in July of our reference year (2018). Refer to [Code Example 2](#).

If this is the only denominator you are interested in, you may simply construct your denominator using two key variables. The first is MDCR_ENTLMT_BUYIN_IND_07 (the 07 extension on this variable corresponds with the 7th month of the year, July). As in the example above, we will count only the beneficiaries with Part A and B coverage (i.e., values of ‘3’ and ‘C’ for this variable). Next, we want to exclude beneficiaries who have managed care coverage during the same month. The variable to use is HMO_IND_07 (any value other than ‘0’ or ‘4’ = managed care coverage; we also want to retain the fours since FFS claims are available).

Once you have a subset of beneficiaries who meet the denominator criteria again, you would count your cases (i.e., your numerator) as the 3s and 1s for your condition of interest. No others in your denominator have evidence (FFS claims) indicating treatment for the condition. Adding the following lines of code to the data step above will create the needed variable. Note that the code example below is only a portion of the necessary code; [Appendix B — SAS Code](#) has the complete code.

Code Example 2. Define a sample of beneficiaries with Medicare FFS coverage at a point in time

The next portion of SAS code assumes that you are working with source data that includes the MBSF A/B/C/D and CCW Chronic Conditions Segments (e.g., using the data and merge steps in [Code Example 1](#)). We have the full code for each of the examples in [Appendix B — SAS Code](#). We determine the beneficiaries enrolled in FFS for the particular month (FFS for July only), then we identify those with the condition of interest (an annual variable). We refer to the output of this data step as “temp2.”

```

if mdcr_entlmt_buyin_ind_07 in ('3','C') and hmo_ind_07 in ('0','4') then
Member_FFSMos07=1;
else if mdcr_entlmt_buyin_ind_07 NOT in ('3','C') or hmo_ind_07 NOT in
('0','4') then Member_FFSMos07=0;
if Member_FFSMos07=0 then AMI_PtTimeRT=.;
else if Member_FFSMos07=1 and (ami= 0 or ami=2) then AMI_PtTimeRT =0;
else if Member_FFSMos07=1 and (ami= 1 or ami=3) then AMI_PtTimeRT =1;

```

Again, to compute the point in time coverage rate we simply aggregate the AMI_PtTimeRT variable created in the data step using the means procedure.

```

proc means N SUM MEAN;
var AMI_PtTimeRT;
run;

```

- **Person years with coverage** — for this type of denominator, the objective is to ascertain Medicare FFS member years of coverage (i.e., cumulative member months at risk, divided by 12). For a one-year condition (AMI), we will accumulate 12 months of coverage information for each beneficiary in the 5% sample (**NOTE:** for a two-year condition you would want to accumulate 24 months of coverage, etc.). Then, to compare methods for calculating prevalence, we divide by 12 to obtain an “average” member months at risk and use this for our denominator. This is similar to the denominator calculation for a traditional “period prevalence” rate type.

You will need to create “counter” variables that accumulate the number of months each beneficiary meets your coverage criteria. Using our same definition of coverage (Part A and Part B with no HMO), we count number of months the coverage meets this definition. We illustrate this process using [Code Example 3](#) below. This example highlights only a portion of the SAS code needed to complete the analysis; refer to [Appendix B — SAS Code](#).

For the numerator, we need to determine how often events occur. Using this denominator, we count the 3s and 1s for the condition for the identified time period (**NOTE:** for this method, there is no need to link beneficiaries — but to count months at risk and events). Use caution for interpreting this type of rate. It is not technically incidence but rather indicates the number of beneficiaries at risk who had a treatment event during the time of interest.

Code Example 3. Determine total member months of Medicare FFS coverage and corresponding condition numerator

This time add the following lines of code to the end of the data step (before the Labels statement) from [Code Example 1](#). The full code example is in [Appendix B — SAS Code](#). We refer to the output of this data step as “temp3,” which we use for the proc means below.

```
Member_Years = Member_Mos/12;  
if (ami= 0 or ami=2)then AMI_Flag=0;  
else if (ami= 1 or ami=3)then AMI_Flag=1;
```

To compute the person years coverage rate, we compute the sum of the Member_Years (denominator) and AMI_Flag (numerator) variables created in the data step using the means procedure.

```
proc means SUM;  
var Member_Years AMI_Flag;  
run;
```

Chapter 3. Results

The four methods for calculating prevalence described above produce somewhat different rates. The fourth method, described in [Table 1](#), uses a type of statistic that is technically different from the others. In the fourth method, the data do not identify a particular cohort; rather, they determine the number of events for a person’s time at risk.

Table 1. Four methods for calculating the prevalence of AMI in 2018 (5% random sample)

	Denominator	Numerator	Percent
No coverage restrictions*	3,150,608	16,943	0.53%
1. Full FFS coverage	1,595,064	14,914	0.94%
2. Partial FFS coverage	1,611,261	15,024	0.93%
3. Point in time FFS coverage	1,656,967	13,959	0.84%
4. Person years with FFS coverage	1,654,735.67	16,943	1.02%

*This is not a viable method (displayed for comparative purposes only).

The first row in [Table 1](#) (representing no coverage restrictions and shaded to contrast this from the more deliberate and accurate rate computations) has a larger denominator than the other viable methods — and a correspondingly low rate. The denominator is the sample size for the entire random 5% sample for 2018. The numerator includes all the 3s and 1s for the condition. ***We do not recommend ever describing prevalence of a condition using this technique,*** because it includes beneficiaries with no FFS coverage (i.e., those who are not “at risk” for having an FFS claim indicating the presence of the condition).

Determining which rate is the correct rate, depends on the research question and the purpose of the analysis. For example, you may wish to include only those with full FFS coverage in your study if it is essential to obtain a thorough description of all services used by beneficiaries with a certain condition. This maximizes the FFS surveillance opportunity; however, you lose some of the known cases in the cohort (i.e., it does not include the 1s).

Allowing a partial break in coverage enables you to keep more known cases. If we understand that full FFS coverage may not necessarily be “typical” of all Medicare beneficiaries, then this type of rate may be more generalizable. The CCW team suggests completely ascertaining patterns of care, costs, and outcomes for this cohort, as the surveillance period is still extensive.

Point in time coverage is a method often appealing to those trying to extrapolate rates to the entire population — FFS beneficiaries for a typical month (and costs for these beneficiaries for a typical month, etc.). However, this type of information includes fewer cases in the numerator since it considers only those enrolled in July.

Including all known events is part of the appeal of the person-time method for calculating these rates — and CCW counts every month of FFS “at risk.” This method is useful for constructing prevalence-like ratios; however, it is not particularly helpful if the researcher wishes to examine claims or patterns of care for people in the denominator. We count beneficiaries in the denominator if there is even a single month of coverage, yet there is not a strictly defined cohort.

A. Next Steps

Perhaps the purpose of your study goes well beyond simply determining the prevalence of the condition of interest. Once you have defined your denominator, you may proceed to the rest of your analysis (e.g., assessing utilization, health outcomes, etc.) — including only this subset of beneficiaries (i.e., your cohort of interest).

Chapter 4. Generalizing These Methods

Some conditions in the CCW require a two-, three-, or five-year surveillance period (e.g., diabetes requires two years, Alzheimer’s [when in the MBSF_CC file] requires three years, and sickle cell disease requires five years). Researchers can generalize the methods we described to these conditions. For prevalence calculation method #1 (Full FFS coverage), you would simply need to look at the yearly indicator variable for the condition of interest. By definition, the 3s and 2s indicate that the beneficiary had full coverage for the surveillance period (not just the calendar year represented by the MBSF). For the other methods, you would need to obtain multiple years of the MBSF A/B/C/D segment to accumulate months of coverage for the entire duration of the surveillance period. Researchers can specify which years of the MBSF they desire during the data request process.

Our examples used the yearly condition indicator variable for all numerator calculations. You may choose to use the mid-year variable⁸ (e.g., AMI_MID) or the first occurrence date (e.g., AMI_EVER) for numerator determination. Some conditions included may not require active treatment, yet it is helpful to know the disease history (e.g., breast cancer, stroke) to obtain a “clean” control group. CCW designs the “ever” variables (i.e., first occurrence date) for this purpose.

Additional potential use of the “ever” variables is to look at the population subset with a recurrence of the disease. For example, perhaps you are interested in studying stroke or breast cancer. Beneficiaries identified as having the condition using the current year variable (called STROKE_TIA or CANCER_BREAST in the MBSF_CC file) may consist of subgroups who have never had the condition before — and those who have a prior history of the condition (i.e., perhaps the current year indicates a recurrence of the condition rather than an incident case). Researchers can identify these subgroups by determining whether the ever date was before the year of the data file (e.g., if the date value in the STROKE_TIA_EVER variable is before 2018, then the stroke indicated in the 2018 data file is a recurrent stroke). The earliest possible ever date for anyone in the CCW database is January 1, 1999,⁹ although some beneficiaries become eligible for Medicare before then. 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 A/B/C/D segment). *The CCW team cautions investigators that the absence of an ever date that precedes the current year does not mean the provider treated the person for the disease; it simply means that we have no FFS Medicare claims history (from 1999 forward) indicating the presence of claims for the condition.*

For conditions that only require an inpatient diagnosis code for us to classify them according to the condition definition, you may wish to construct more lenient coverage criteria than we discuss here. For example, for hip fracture (variable called HIP_FRACTURE), the presence of one inpatient diagnosis code is sufficient for meeting the claims-based definition. You should determine whether you wish to look at cohorts with Part A coverage and/or those with Part B coverage. By keeping the coverage criteria as lenient as possible, we retain more of the known cases. Investigators can adapt the SAS code displayed above to accommodate these scenarios.

A. Limitations

The condition definitions in CCW are claims-based — determined by documentation of diagnosis or procedure codes for the condition of interest in the Medicare FFS claims data. As a result of using these administrative data to infer the presence of conditions, population prevalence rates derived from this data source may differ somewhat from

⁸ The mid-year condition variables are available only in the MBSF_CC file.

⁹ For the MBSF_CHRONIC file, the earliest “ever” date possible is 01/01/2016.

prevalence rates constructed from other data sources — particularly those derived from survey data or other types of clinical data.

The claims used to make the condition determinations are for Medicare FFS only. As a result, there is missing data in the CCW due to managed care coverage. Since claims for most services provided to Medicare beneficiaries in managed care do not reach the FFS claims data files, data users should view the CCW Medicare claims as providing utilization information primarily for the FFS population. The managed care population may differ in important ways from the FFS population (e.g., they could potentially be younger and healthier). Researchers should be cautious when making population-level generalizations using CCW data.

Researchers have a variety of hypotheses and objectives. The intent of this paper is not to be prescriptive, but rather descriptive of useful tools for refining cohorts and calculating rates. The objective is to make it easy for you to accomplish your study objectives and to ensure you have a thorough understanding of the data available from the CCW. Using the methods described in this paper, we can better understand the magnitude of chronic conditions and their effect on the population through appropriate and accurate data analysis techniques.

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

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

Appendix A — List of Acronyms and Abbreviations

Acronym	Definition
AMI	Acute Myocardial Infarction
BASF	Beneficiary Annual Summary File
BSF	Beneficiary Summary File
BETOS	Berenson-Eggers Type of Service
CC	Chronic Condition
CCW	Chronic Conditions Warehouse
CMS	Centers for Medicare & Medicaid Services
COPD	Chronic Obstructive Pulmonary Disease
FFS	Fee-for-service
HIV/AIDS	Human Immunodeficiency Virus/Acquired Immunodeficiency Disorder
HMO	Health Maintenance Organization
MAX	Medicaid Analytic eXtract
MBSF	Master Beneficiary Summary File
MMA	Medicare Modernization Act of 2003
MMLEADS	Medicare-Medicaid Linked Enrollee Analytic Data Source
PTSD	Post-Traumatic Stress Disorder
ResDAC	Research Data Assistance Center

Appendix B — SAS Code

In this section, we present the full SAS code we used to generate the results; whereas, in the main body of the document, we highlighted only the key components of the code for each of the code examples.

1. Full FFS coverage in a 5% sample.

```
data temp;
merge MBSF.MBSF_ABCD_2018(keep=BENE_ID SAMPLE_GROUP)
      MBSF.MBSF_CC_2018 (keep=BENE_ID AMI AMI_EVER);
  by BENE_ID;
  if SAMPLE_GROUP in ('01','04');
run;

proc freq;
tables AMI;
run;
```

2. Partial FFS coverage in a 5% sample.

This is the complete SAS code used in Code Example 1.

```
data temp1;
merge MBSF.MBSF_ABCD_2018(keep=BENE_ID mdcr_entlmt_buyin_ind_01 -
mdcr_entlmt_buyin_ind_12 hmo_ind_01 - hmo_ind_12 bene_death_dt SAMPLE_GROUP)
      MBSF.MBSF_CC_2018 (keep=BENE_ID AMI AMI_EVER);
  by BENE_ID;
  if SAMPLE_GROUP in ('01','04');
```

***note this input file is a merged MBSF A/B/C/D and MBSF CC Conditions file segment;**

*** determine # months of Part A, B, and no HMO coverage;**

```
array MemberMos_AB (12)
mdcr_entlmt_buyin_ind_01 - mdcr_entlmt_buyin_ind_12;
array MemberMos_noHMO (12) hmo_ind_01 - hmo_ind_12;
array Member_FFSMos (12) Member_FFSMos01 - Member_FFSMos12;
do i= 1 to 12;
if MemberMos_AB(i) in ('3','C') and MemberMos_noHMO(i) in
('0','4')then Member_FFSMos(i)=1;
else if MemberMos_AB(i) NOT in ('3','C') or MemberMos_noHMO(i) NOT in
('0','4')then Member_FFSMos(i)=0;
Member_Mos=sum(of Member_FFSMos:);
end;
```

*** note – determine who had 11- or 12-months coverage or coverage until month before death;**

```

if (bene_death_dt=. and Member_Mos in (11,12)) or (bene_death_dt~=.
and month(bene_death_dt)<=Member_Mos+1 and Member_mos~=0) then Partl_Cov=1;
else Partl_Cov=0;
* note – bring in numerator information for AMI – keep both 3s and 1s;
if Partl_Cov=0 then AMI_PartRT=.;
else if Partl_Cov=1 and (ami= 0 or ami=2)then AMI_PartRT=0;
else if Partl_Cov=1 and (ami= 1 or ami=3)then AMI_PartRT=1;
label
Partl_Cov = '11 or 12 months FFS no HMO – except for those who died'
Member_Mos = 'Total Member months of A B and No HMO – per bene'
AMI_PartRT = 'Had AMI – partial coverage';
run;

proc means N SUM MEAN;
var AMI_PartRT;
run;

```

3. Point in time FFS coverage in a 5% sample.

This is the complete code used in Code Example 2.

```

data temp2;
merge MBSF.MBSF_ABCD_2018(keep=BENE_ID mdc_r_entlmt_buyin_ind_01 -
mdc_r_entlmt_buyin_ind_12 hmo_ind_01 - hmo_ind_12 bene_death_dt
SAMPLE_GROUP)
MBSF.MBSF_CC_2018 (keep=BENE_ID AMI AMI_EVER);
by BENE_ID;
if SAMPLE_GROUP in ('01','04');

if mdc_r_entlmt_buyin_ind_07 in ('3','C') and hmo_ind_07 in ('0','4')then
Member_FFSMos07=1;
else if mdc_r_entlmt_buyin_ind_07 NOT in ('3','C') or hmo_ind_07 NOT in
('0','4')then Member_FFSMos07=0;
if Member_FFSMos07=0 then AMI_PtTimeRT=.;
else if Member_FFSMos07=1 and (ami= 0 or ami=2)then AMI_PtTimeRT =0;
else if Member_FFSMos07=1 and (ami= 1 or ami=3)then AMI_PtTimeRT =1;

proc means N SUM MEAN;
var AMI_PtTimeRT;
run;

```

4. Person years with FFS coverage in a 5% sample.

This is the complete code used in Code Example 3.

```

data temp3;
merge MBSF.MBSF_ABCD_2018(keep=BENE_ID mdc_r_entlmt_buyin_ind_01 -
mdc_r_entlmt_buyin_ind_12 hmo_ind_01 - hmo_ind_12 bene_death_dt SAMPLE_GROUP)
      MBSF.MBSF_CC_2018 (keep=BENE_ID AMI AMI_EVER);
  by BENE_ID;
  if SAMPLE_GROUP in ('01','04');

*note this input file is a merged MBSF A/B/C/D and Conditions File segment;
* determine # months of Part A, B, and no HMO coverage;
  array MemberMos_AB (12)
mdc_r_entlmt_buyin_ind_01 - mdc_r_entlmt_buyin_ind_12;
  array MemberMos_noHMO (12) hmo_ind_01 - hmo_ind_12;
  array Member_FFSMos (12) Member_FFSMos01 - Member_FFSMos12;

do i= 1 to 12;
if MemberMos_AB(i) in ('3','C') and MemberMos_noHMO(i) in
('0','4')then Member_FFSMos(i)=1;
else if MemberMos_AB(i) NOT in ('3','C') or MemberMos_noHMO(i) NOT in
('0','4')then Member_FFSMos(i)=0;
Member_Mos=sum(of Member_FFSMos:);
end;

Member_Years = Member_Mos/12;
if (ami= 0 or ami=2)then AMI_Flag=0;
else if (ami= 1 or ami=3)then AMI_Flag=1;

proc means SUM;
var Member_Years AMI_Flag;
run;

```