# mbsf_ccw (Processed CCW medicare dataset documentation)

## 1. Data description overview

This dataset contains beneficiary-level chronic condition indicators derived from the CMS Chronic Conditions Warehouse (CCW) Master Beneficiary Summary File – Chronic Conditions segment (mbsf_cc) and (mbsf_chronic). The data includes all eligible beneficiaries in a given year, not only those with chronic conditions The dataset is intended for research use and provides standardized indicators for selected chronic conditions defined by CCW algorithms.

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 chronic conditions files:

1. the 27 CCW Chronic Conditions (available in NSAPH from 2000-2016 as the mbsf_cc_yyyy file)
2. the 30 CCW Chronic Conditions (available in NSAPH from 2017-2022 as the mbsf_chronic_yyyy file)

**Comparison of 27 CCW and 30 CCW chronic condition algorithms**

The move from the 27-condition to the 30-condition algorithms includes methodological updates such as:

- adding or retiring diagnosis (DX) codes;
- expanding the number of diagnosis code fields inspected on claims;
- changing which claim file types contribute evidence; and
- modifying the lookback period length.

For condition-specific definitions and change details, see the CCW page: Chronic Condition Categories.

This README provides documentation and processing details that serve as references for the processed dataset generated from the raw files received from ResDAC.

For dataset location details on ReD computing environement, see our internal data catalog: CCW datasets.

### 2. File structure and relationships

**Summary unit:** Beneficiary–year (one row per beneficiary per calendar year).

**Core tables:** `mbsf_cc` (27 CCW conditions) and `mbsf_chronic` (30 CCW conditions).

**Join keys:** The CCW-encrypted beneficiary identifier (`bene_id`) and `year`. These keys are used to link with other CCW-derived files within the ReD computing environment, including the Master Beneficiary Summary File (`mbsf_mortality`), MEDPAR, MDS, Part D, outpatient, and carrier files.

**Cohort creation:** Essential information for defining most study denominators is found in the Base A/B/C/D segment of the MBSF (referred to as `mbsf_mortality` in NSAPH). This segment primarily contains beneficiary demographic characteristics and Medicare coverage information. Variables from the MBSF

A/B/C/D segment are commonly used to define study cohorts (e.g., by age, geography, or sex) and to apply coverage eligibility criteria, including:

- **Coverage variables:** Monthly Medicare entitlement and buy-in indicators for Part A and/or Part B (`buyin_indicators`).
- **Dual eligibility:** Monthly indicators of Medicare–Medicaid dual status (`dual_indicators`).
- **Managed care / Medicare Advantage / HMO:** Monthly indicators of Medicare managed care enrollment (`hmo_indicators`).

A key consideration when merging with the MBSF is that the NSAPH data science team has preprocessed and cleaned the data : datasets are restricted to beneficiaries aged 65+ years.The data can be further merged with medpar_denom to identify FFS beneficiaries, and the buyin_indicators can then be used to assess monthly Part A and/or Part B coverage in more detail.

## 3. Data dictionary

For more information on dataset schemas and variables, see the catalog pages for mbsf_cc and mbsf_chronic.

Each defined chronic condition is represented by:

**Yearly indicator**: denotes whether the beneficiary met the claims-based algorithm within the reference year. Coding uses four values reflecting both claims evidence and fee-for-service (FFS) coverage sufficiency over the surveillance window (12/24/36 months, or until death, depending on the condition):

- **0:** Did not meet claims criteria and did not have sufficient FFS coverage.
- **1:** Met claims criteria but did not have sufficient FFS coverage (e.g., partial Part A/B months, HMO months, or new eligibility).
- **2:** Did not meet claims criteria but had sufficient FFS coverage.
- **3:** Met claims criteria and had sufficient FFS coverage (all required Part A and Part B months and no HMO, or until death).

Each yearly indicator uses December 31 as the end of the reference year; for a one-year reference period, services span 01/01–12/31 of that year.

**First occurrence (ever)**: date the beneficiary first met the clinical FFS claims criteria (coverage criteria not applied). Earliest possible dates: 01/01/1999 for mbsf_cc and 01/01/2016 for mbsf_chronic. A null indicates no qualifying FFS claims in the surveillance period, or since the beneficiary's Medicare eligibility start.

Both `mbsf_cc` and `mbsf_chronic` include these fields as appropriate for their condition sets.

## 4. Dataset motivation

### a. Motivation

- This dataset was created to transform the CCW Chronic Conditions segment into a standardized,analysis-ready format by harmonizing variable structure across years and converting the data to an efficient columnar (Parquet) format. The processing enables consistent use of CCW chronic condition indicators across studies and supports reproducible, scalable analysis of Medicare data.

**b. Funding**

- Supported by NSAPH-affiliated research projects and collaborating grants; acknowledge specific award numbers in study outputs where required.

**c. Contributors**

- Generated and processed by the NSAPH (National Study on Air Pollution and Health) data science team and collaborators.

## 5. Dataset composition

**a. Completeness**

- The MBSF Chronic Conditions segments include a record for all eligible beneficiaries in a given year, not just those with chronic conditions. Although CMS began disseminating Medicare Advantage encounter data in 2015, the CCW algorithms currently do not use encounter data. Therefore, the CCW is primarily a source of utilization and chronic condition information for the Medicare FFS (Part A and Part B) population.

**b. Limitations and considerations**

**Claims-Based and FFS-Only:** CCW condition indicators are derived from administrative Medicare fee-for-service (FFS) claims rather than clinical records, so prevalence estimates may differ from those based on surveys or other clinical data sources. The chronic condition algorithms do not incorporate Medicare Advantage or managed care encounter data; therefore, CCW data primarily reflect utilization and conditions among the FFS population. Because beneficiaries enrolled in managed care may differ systematically from those in FFS (e.g., younger or healthier), population-level generalizations should be made with caution.

**Enrollment and coverage considerations:** Identification of chronic conditions depends on the extent and timing of Medicare Part A and Part B enrollment, as most CCW condition definitions require qualifying inpatient, outpatient, or carrier claims. Partial-year enrollment or gaps in coverage may reduce the likelihood that conditions are observed. For some conditions, such as acute myocardial infarction or hip fracture, a single inpatient claim may be sufficient; researchers should consider whether inclusion of beneficiaries without Part B coverage is appropriate based on study objectives.

**Historical cepth and "Ever" dates:** The 30 CCW Chronic Conditions file relies exclusively on ICD-10-CM codes, limiting historical claims to 2016 onward and making January 1, 2016 the earliest possible "ever" date. In contrast, the 27-condition file includes history back to 1999, which can result in differences in "ever" dates across files. Consequently, "ever" dates in the 30 CCW file may not represent the true first occurrence of a condition if it occurred prior to 2016; these dates are most reliable for beneficiaries newly enrolled in Medicare in 2016 or later, as indicated by the `COVSTART` variable.

**c. External sources and citations**

The raw files are developed and maintained by the Centers for Medicare & Medicaid Services (CMS).

- Centers for Medicare & Medicaid Services (CMS). Chronic Conditions Data Warehouse (CCW) Chronic Condition Segments: Master Beneficiary Summary File (MBSF) Add-On, 1999–2021. Baltimore, MD: CMS. Available at: https://www2.ccwdata.org/web/guest/condition-categories

- Research Data Assistance Center (ResDAC). CMS Data Request and Access Procedures. Minneapolis, MN: University of Minnesota. Available at: https://www.resdac.org

**d. Data de-sensitized**

- No data is provided with this entry. Processed analytic files remain within secure, DUA-governed environments and exclude direct identifiers. When publishing tabulated results, users must apply CMS small-cell suppression and complementary suppression rules (e.g., suppress counts < 11) consistent with the approved DUA and institutional policies.

## 6. Processing and Access

**a. Preprocessing and cleaning**

All CCW Chronic Conditions segment files for each year were ingested, converted to Parquet format, and standardized with proper data types and structure. All beneficiaries were included without exclusions and the data were checked for duplicates. Two outputs were produced for each year: mbsf_cc_YYYY (27-CCW condition) and mbsf_chronic_YYYY (30-CCW condition).

**b. Is the source data included in this dataset or otherwise preserved and accessible?**

No. The original raw source files are not included in this dataset.

Access to the underlying CMS Chronic Conditions Warehouse (CCW) data is restricted and governed by a CMS Data Use Agreement (DUA).

**Access for NSAPH collaborators**

Students, postdoctoral fellows, and researchers conducting research in collaboration with principal investigators at NSAPH or affiliated institutions with an approved DUA may request access to the data if they meet the following requirements:

- Submit a project initiation form;
- Be listed on an approved Institutional Review Board (IRB) protocol;
- Complete all required data security and privacy training.

**Independent access via ResDAC**

Researchers not affiliated with NSAPH may request access to the source data directly through the Research Data Assistance Center (ResDAC):

https://www.resdac.org