

Medicare & Medicaid Data Sources Available to States

Understanding and improving care for individuals who are dually eligible for Medicare and Medicaid requires access to comprehensive and reliable data. These individuals often have complex health and social needs, and their care spans two distinct programs. To support effective care coordination, policymakers and researchers use Medicare and Medicaid data to identify beneficiaries, analyze service use, assess care quality, and inform decisions that advance health equity and program efficiency.

A variety of data sources are used to study this population. Administrative data, such as the Chronic Conditions Data Warehouse and claims data from both programs, provide insight into enrollment, utilization, and spending. Survey and assessment tools, including the Medicare Current Beneficiary Survey and Minimum Data Set, offer additional perspectives on patient-reported outcomes, functional status, and access to care. These data help inform efforts to improve care delivery models and evaluate services across different subpopulations of people who are dually eligible, including people with intellectual and developmental disabilities (I/DD).

Recent advances have enhanced the integration of Medicare and Medicaid data, and the use of person-level identifiers that enable more comprehensive analyses. Continued investment in data infrastructure and standardization is essential to support the development of coordinated care models for this high-need population.

This resource describes the various Medicare & Medicaid data sources available to states. Each source includes the following details, including the corresponding icon key:



Why Are These Data Useful?



Data Source



Data Format



Data Time Coverage and Reporting Frequency



Data Access Method



Selection Criteria



Additional Notes



Weblink

Enhanced Coordination of Benefits Agreement (COBA)



Why Are These Data Useful? COBA Data contains Medicare claims - Part A (inpatient hospital, skilled nursing facility, hospice care) and Part B (doctor visits, outpatient care, preventive services, covered medical services and supplies, limited outpatient drugs under certain conditions). This data can be used to establish a health care utilization baseline, to better understand the demographics of a state's dually eligible population and monitor trends.



Data Source. COBA standardizes the way that eligibility and Medicare claims payment information, within a claims crossover context, are exchanged. COBAs permit other insurers and benefit programs (also known as trading partners) to send eligibility information to the Centers for Medicare & Medicaid Services (CMS) and receive Medicare claims data for processing supplemental insurance benefits from CMS' national crossover contractor, the Benefits Coordination & Recovery Center (BCRC). The BCRC houses COBA trading partner's eligibility information for crossover purposes only in those instances where the information successfully matches with the in-file CMS entitlement information. COBA trading partners are apprised of situations where their eligibility information matches CMS eligibility data as well as when their submitted information does not result in a match.



Data Format. X12 (Parts A/B) / NCPDP (DME) - Electronic Data Interchange (EDI).



Data Time Coverage & Reporting Frequency.

Current, either real-time or near real-time data updates.
Daily/Weekly - Based on eligibility files from states.



Data Access Method. Secure File Transfer Protocol (SFTP) & State Connect:Direct - Secure exchange of data.



Selection Criteria. Eligibility-based file from the state to determine coverage under COBA.



Additional Notes. State's arrangement with Managed Care Organizations (MCOs) and providers affects data sources and access.



Weblink. <https://www.cms.gov/medicare/coordination-benefits-recovery/coba-trading-partners/agreement>

Historic Medicare Data and Beneficiary Summary File (MDBSF)



Why Are These Data Useful? The Historic MDBSF includes base segment data (date of birth, death, sex, race, state/county of residence, and monthly enrollment data) for Medicare Parts A & B; crosswalk files; and assessment data. It also contains duals status codes and data to indicate if a person is part of a specific sample group within the MDBSF. This data can be used to establish a baseline to better understand the demographics of a state's dually eligible population and their health care trends.



Data Source. Chronic Condition Warehouse (CCW) provides states and researchers with Medicare and Medicaid beneficiary, claims, and assessment data linked by beneficiary across the continuum of care. In the past, states and various researchers analyzing data files were required to perform extensive analysis related to beneficiary matching, deduplication, and merging of the files in preparation for their study analysis. With the CCW data, this preliminary linkage work is already accomplished and delivered as part of the data files sent to states and researchers.

The Chronic Conditions Data Warehouse (CCW) is a research database designed to make Medicare, Medicaid, Assessments, Part C, and Part D Prescription Drug Event data more readily available to support research designed to improve the quality of care and reduce costs and utilization.



Data Format. Fixed Length Flat Files - Structured data in predefined fields.



Data Time Coverage & Reporting Frequency.

Historic (15-month lag) and Monthly (4-month lag) - Medicare beneficiary data
Yearly/Monthly - Depending on the enrollment and MMA file criteria.



Data Access Method. External Media - Data is shared via encrypted hard drives or other secure physical media.



Selection Criteria. Enrollment and MMA (Medicare Modernization Act) file used to filter relevant records.



Additional Notes. Enrollment and MMA (Medicare Modernization Act) file used to filter relevant records.



Weblink. <https://aspe.hhs.gov/maintenance-support-chronic-conditions-warehouse-ccw-comparative-effectiveness-research-cer>

Part D Prescription Drug Event (PDE)



Why Are These Data Useful? The PDE data contains summary pharmacy claims (not individual drug claims) for Medicare beneficiaries that CMS uses to track and administer the Medicare Part D prescription drug benefit, including payments made to managed care plans. This data is very helpful to better understand the drug utilization patterns of a state's dually eligible population.



Data Source. Integrated Data Repository (IDR) is a high-volume data warehouse integrating Parts A, B, C, D, and DME claims, beneficiary and provider data sources, along with ancillary data such as contract information, risk scores, and many others. Access to this robust integrated data supports much needed analytics for states and across CMS.



Data Format. File is a subset of PDE data authorized for use - Extracted from the IDR.



Data Time Coverage & Reporting Frequency.

Historic and Monthly (1-month lag) - Recent and retrospective prescription claims data

Historic and Monthly - Regular updates based on data access agreements



Data Access Method. Connect: Direct - Automated, secure data transmission.



Selection Criteria. Enrollment/MMA file-based selection for PDE subset access.



Additional Notes. PDE data includes prescription drug claims and cost-sharing details for Medicare Part D.



Weblink. <https://www.cms.gov/about-cms/information-systems/idr>

State Medicaid Health Care Claims Data



Why Are These Data Useful? State Medicaid Health Care Claims Data includes Medicaid Encounters and Medicaid Fee For Service (FFS) Claims. These data provide a detailed view of how Medicaid beneficiaries are accessing healthcare services. This includes information on the types of services received, the frequency of those services, and the providers delivering them. By analyzing claims data, researchers and policymakers can identify patterns of healthcare utilization, revealing trends in disease prevalence, service demand, and access to care.



Data Source. State data warehouses; encounters sourced from managed care plans through contractual requirements for reporting



Data Format. X12 837 5010 for managed care encounters; CMS-1500 and 837P for fee -for -service claims.



Data Time Coverage & Reporting Frequency.
Historic, Weekly, Monthly Quarterly, Annual
Regular updates based on data access agreements with health plans and providers.



Data Access Method. It is an automated, secure data transmission.



Selection Criteria. State MMIS (member eligibility, health care claims/encounters).



Additional Notes. N/A



Weblink. N/A

Other Medicare/Medicaid Quality Data: Assessments, Measures, and Surveys



Why Are These Data Useful? Quality Data/Measures includes information that can be used to help evaluate people's health care experiences from Medicare and Medicaid programs. It can target program access and choice, service utilization and satisfaction. This also can include provider experiences. Examples of data collected by both Medicaid and Medicare include Minimum Data Set assessment data, Healthcare Effectiveness Data and Information Set (HEDIS) scores, Consumer Assessment of the Healthcare Providers and Systems (CAHPS) member surveys, and the Centers for Medicare and Medicaid (CMS) Core Measure sets. An additional potential data source for Medicare includes star ratings. Additional Medicaid data examples include home and community-based service quality assurance measures, other member surveys, such as National Core Indicators survey; and star ratings as applicable.



Data Source. Federal and state reports; a State's External Quality Review Organization (EQRO); also sourced from managed care plans through contractual requirements for reporting.



Data Format. File formats vary from one data element to another.



Data Time Coverage & Reporting Frequency.

Historic, Monthly Quarterly, Annual

Regular updates based on data access agreements with health plans and providers.



Data Access Method. It is an automated, secure data transmission.



Selection Criteria. Can determine where to start by looking at State Quality Priorities.



Additional Notes. N/A



Weblink. N/A