HHS BARDA Solving Sepsis Framework and Algorithm Specifications

Version 3.0, February 11, 2020

|  |  |  |  |
| --- | --- | --- | --- |
| **Version No.** | **Release Date** | **Summary of Changes** | **Revised By** |
| 1.0 | July 12, 2019 | Established document | ASPR, Acumen |
| 2.0 | August 23, 2019 | Made some formatting edits | ASPR, Acumen |
| 3.0 | February 11, 2020 | Made some language revisions and formatting edits | ASPR, Acumen |

**Document Revision History**

Table of Contents

[1 Introduction 3](#_Toc32328710)

[2 Analytic Framework 3](#_Toc32328711)

[2.1 Components of The Analytic Framework 4](#_Toc32328712)

[3 Medicare Data 6](#_Toc32328713)

[3.1 Data Sources 6](#_Toc32328714)

[3.2 Medicare Claim Files 7](#_Toc32328715)

[3.3 Diagnosis, Revenue Center, and Related Condition Codes 7](#_Toc32328716)

[4 Implementation: Identifying Beneficiaries 9](#_Toc32328717)

[4.1 Analytic Stream 1 - Alive and Enrolled Medicare Beneficiaries 9](#_Toc32328718)

[4.2 Analytic Stream 2 – Inpatient Stays 10](#_Toc32328719)

[4.2.1 Collapsing Claims to Stays and Constructing Stay-Level Measures 10](#_Toc32328720)

[4.2.2 Identifying Sepsis Stays 11](#_Toc32328721)

[5 Implementation: Construct Master Inpatient Stay-Level Dataset 12](#_Toc32328722)

[6 Implementation: Develop Measures of Health History and Healthcare Service Use Prior to Inpatient Admission 12](#_Toc32328723)

[7 Implementation: Identify Healthcare Service Use Subsequent to an Acute Inpatient Discharge 12](#_Toc32328724)

[Appendix A: Acronyms and Abbreviations 14](#_Toc32328725)

HHS Solving Sepsis Framework and Analytical Specifications

# Introduction

According to the Center for Disease Control and Prevention (CDC), an estimated 1.7 million Americans develop sepsis each year, resulting in around 270,000 deaths.[[1]](#footnote-1) The World Health Organization (WHO) estimates that over 30 million people worldwide will acquire sepsis annually, potentially leading to six million deaths. While anyone can become infected with sepsis, newborns, older persons, and persons with chronic health problems are among the most vulnerable groups.[[2]](#footnote-2)

Given the severity of this life-threatening condition, the Assistant Secretary for Preparedness and Response (ASPR), the Biomedical Advanced Research and Development Authority (BARDA), and the Center for Medicare and Medicaid Services (CMS), worked together to report on sepsis among Medicare beneficiaries. The aim was to produce reports that can support efforts to reduce sepsis mortality rates and build more resilient communities.

Some of the key questions that were addressed in these analyses included identifying and quantifying sepsis prevalence, estimating the cost burden of sepsis, understanding the impact of sepsis on mortality and other health outcomes, understanding the healthcare trajectories of patients who survive sepsis as well as the factors that contribute to the risk of acquiring sepsis both in and out of a hospital. This document provides an overview of the methodology and data types used in performing these sepsis analyses.

# Analytic Framework

This section provides an overview of our general analytic framework. For our analyses of sepsis cases in the Medicare population, we rely primarily on Medicare Fee-For-Service (FFS) claims data from the Common Working File (CWF), the Medicare Enrollment Database (EDB), and the Long-Term Care Minimum Dataset (MDS).

Section 2.1 introduces the analytic framework that we use to identify and analyze cases of inpatient sepsis in the Medicare population.

## Components of The Analytic Framework

The analytic framework can be separated into two stages: i) identification of beneficiaries and inpatient stays and ii) construction of analytic datasets combining information on inpatient stays and beneficiary characteristics.

##### **IDENTIFICATION OF BENEFICIARIES AND INPATIENT STAYS**

The process for identifying inpatient sepsis stays for Medicare Fee-for-Service (FFS) beneficiaries consists of three steps. The first two steps, or analytic streams, occur independently. The third step merges the output of the two analytic streams into a single master inpatient stay-level dataset that we use for our key analyses. The fourth step creates measures of beneficiary health history and healthcare service use prior to inpatient admission. The fifth step identifies healthcare service use subsequent to an acute inpatient discharge. The implementation of each analytic stream is detailed in Section 4.

Each stream provides answers to the following questions:

1. Who is alive and enrolled in FFS Medicare in each month and could potentially have an inpatient stay?
2. Who had an inpatient stay in each month? How many of those inpatient stays included a sepsis diagnosis?
3. **Identify beneficiaries who are alive and enrolled in Fee-for-Service Medicare (Parts A and B but not C) in the month.**

In this step, we also gather information about each beneficiary’s zip code, Medicare-Medicaid dual eligibility status, address, and name information from the Medicare EDB. We supplement the EDB data with an Integrated Data Repository (IDR) dual eligibility file to specifically identify dual eligible beneficiaries in US territories.

1. **Identify inpatient stays in each month, both overall and those with a sepsis diagnosis.**

The primary source of data is CWF claims.

1. **Construct master inpatient stay-level dataset.**

In this step, we merge the output from Steps 1 and 2 and we flag inpatient stays for beneficiaries who are alive and enrolled in Medicare Parts A/B but not C.

1. **Develop measures of beneficiary health history and healthcare service use prior to inpatient admission.**

In this step, we use hierarchical condition categories (HCCs) to identify comorbidities for each beneficiary with an inpatient stay based on diagnoses reported on inpatient, outpatient, and professional claims from 12 months prior to, and inclusive of, the index admission. We used Version 22 of the CMS-HCC risk adjustment model.

We also create preceding healthcare use flags based on inpatient, outpatient, professional, home health, hospice, and SNF claims and MDS assessment data in the week prior to a beneficiary’s acute inpatient admission.

1. **Identify healthcare service use subsequent to an acute inpatient discharge.**

In this step, we use inpatient, SNF, home health, and hospice claims and MDS assessment data identify healthcare service use subsequent to an acute inpatient discharge.

# Medicare Data

This section provides an overview of the data sources and Medicare claim settings that we use in our analyses. Section 3.1 discusses the data sources at a high level, Section 3.2 provides more detailed information on Medicare claims data specifically, and Section 3.3 looks at coding systems used.

## Data Sources

We use four data sources to construct the analytic datasets. Table 1.1 provides an unofficial description of each data source and the information each source provides for our analyses.

Table 1.1 Data Sources

| **Data Source** | **Brief Data Description[[3]](#footnote-3)** | **Information Used in Analyses** |
| --- | --- | --- |
| Common Working File (CWF) Fee-For-Service (FFS) Claims | Medicare Parts A and B utilization records / claims history for individual beneficiaries | * Utilization of inpatient, outpatient, skilled nursing facility, home health, professional, and hospice services |
| Long-Term Care Minimum Data Set (MDS) Assessments | Clinical assessments of residents in nursing homes and beneficiaries using swing beds | * Admission date to nursing home * Discharge date from nursing home * Start date of swing bed use * End date of swing bed use |
| Enrollment Data Base (EDB)[[4]](#footnote-4) | Enrollment and other demographic information for all Medicare beneficiaries | * Beneficiary name * Beneficiary date of birth * Beneficiary date of death * Beneficiary mailing address * Beneficiary dual eligibility status * Beneficiary Medicare enrollment information |
| Integrated Data Repository (IDR) Dual Eligibility Data | Beneficiary dual Medicare/Medicaid eligibility status information | * Beneficiary name * Beneficiary dual eligibility status |

## Medicare Claim Files

Medicare uses two claim forms for submitting health care claims, commonly known as the institutional (CMD-1450/UB-04) and professional (CMS-1500) claim forms. Each Medicare benefit is associated with one of these claim types.

Institutional claims can be further divided based on the type of bill. Different types of facilities use different types of bills when submitting a claim. The type of bill is comprised of three digits where the first digit indicates the facility type, the second digit indicates the bill classification, and the third digit defines the claim (nonpayment, late charge, etc.). The third number is typically not considered when identifying different types of facilities.

Professional claims can be further divided based on the provider or supplier type on the claim.

Table 1.2 includes all of the file settings used in our sepsis analyses, classified as institutional or professional, along with the corresponding types of bill for institutional claims. We also include an unofficial mapping to the National Claims History (NCH) as another mode of reference.

Table 1.2 Medicare Benefits Included in Analyses

| **Setting** | **Type of Claim** | **Type of Bill** | **NCH Claim Type** | **Description** |
| --- | --- | --- | --- | --- |
| Inpatient  *(IP)* | Institutional | 11x | 60, 61, 62, 63, 64 | Services provided to a patient in a hospital (e.g., surgeries) |
| Outpatient  *(OP)* | Institutional | 12x, 13x, 14x, 22x, 23x, 34x, 43x, 71x, 72x, 73x, 74x, 75x, 76x 77x, 83x, 85x | 40, 41, 42 | Non-inpatient services provided to a patient in a hospital (e.g., x-rays) |
| Skilled Nursing Facilities  *(SNF)* | Institutional | 18x, 21x | 20, 30 | Short-term nursing and rehabilitation services following inpatient stays (e.g., physical therapy) |
| Home Health *(HH)* | Institutional | 32x | 10 | Services furnished in patient’s home (e.g., assistance for mobility impaired) |
| Hospice *(HS)* | Institutional | 81x, 82x | 50 | Palliative care for people with less than six months to live (e.g., counseling, physical care) |
| Physician Services  *(PB or Carrier)* | Professional | N/A | 71, 72, 73 | Services provided by non-institutional providers (e.g., physician office visits) |

## Diagnosis, Revenue Center, and Related Condition Codes

We use International Statistical Classification of Diseases and Related Health Problems (ICD) diagnosis codes on claims to identify inpatient sepsis cases. Because our analyses span both pre- and post-2015, when coding practices changed from ICD version 9 to ICD version 10, we include both ICD-9 and ICD-10 codes in our analysis. We also use condition codes on IP claims to identify “shadow billed” no-pay claims for Medicare Advantage (MA) beneficiaries that are reported for information only. Lastly, we use revenue center codes on IP and OP claims to identify emergency room (ER) services preceding an inpatient stay. Table 1.3 provides more detail on these four code types.

Table 1.3 Diagnosis, Revenue Center, and Related Condition Codes

| **Code Type** | **Setting** | **Uses** |
| --- | --- | --- |
| ICD-9 Diagnosis Codes | IP | * Identification of sepsis inpatient stays |
| ICD-10 Diagnosis Codes | IP | * Identification of sepsis inpatient stays |
| Related Condition Codes | IP | * Identification of “shadow billed” no-pay inpatient claims for MA beneficiaries |
| Revenue Center Codes | IP, OP | * Identification of emergency room services preceding an inpatient stay |

# Implementation: Identifying Beneficiaries

This section provides specifications and steps to identifying beneficiaries in each of the following streams:

1. Alive and enrolled Medicare beneficiaries
2. Individuals with inpatient stays

## Analytic Stream 1 - Alive and Enrolled Medicare Beneficiaries

We determine who is alive and currently enrolled in Medicare FFS in each month to develop a baseline population for inclusion in our analyses. We also pull dual eligibility status and demographic information for each beneficiary at this stage.

**Steps**

1. Identify alive beneficiaries.
2. Determine enrollment status.
3. Determine Medicare-Medicaid dual eligibility status.
4. Identify other beneficiary characteristics of interest.

**Methodology**

For each month of the study period:

* Use the EDB to identify a beneficiary’s name and date of death. If it is missing, a beneficiary is considered alive. If it is not missing, a beneficiary is considered deceased.
* Use the EDB to identify beneficiary’s birth date information.
* Use the EDB to determine beneficiary enrollment status*.*
* Use the EDB and IDR dual eligibility data to identify a beneficiary’s Medicare-Medicaid dual eligibility status.
* Use the EDB to identify a beneficiary’s state of residence.
* Use the EDB to identify a beneficiary’s age, sex, and race.
* Use the EDB to identify a beneficiary’s reason for Medicare eligibility.

Table 2.1 Alive and Enrolled Beneficiaries Key Variables

| **Variable Name** | **Description** | **Data Source** |
| --- | --- | --- |
| BENE\_1ST\_NAME | First Name of Beneficiary | EDB |
| BENE\_LAST\_NAME | Last Name of Beneficiary | EDB |
| BENE\_MLG\_CNTCT\_ADR | Beneficiary Mailing Contact Address | EDB |
| DUAL\_MDCR | Dual Eligibility Status | EDB |
| ENR\_A, ENR\_B, ENR\_C, ENR\_D | Beneficiary Medicare Enrollment Plans | EDB |
| BENE\_MDCD\_ELGBLTY\_BGN\_DT | Start Date of Dual Eligibility Period | IDR |
| BENE\_MDCD\_ELGBLTY\_END\_DT | End Date of Dual Eligibility Period | IDR |
| BENE\_DEATH\_DT | Death Date | MBD |
| BENE\_BIRTH\_DT | Birth Date | MBD |

**Outcome**

After the conclusion of these steps, we have a dataset at the beneficiary-month level with the following information: beneficiary name, date of birth, date of death, enrollment plan (FFS or MA), dual eligibility status, and U.S. Territory/state of residence information.

## Analytic Stream 2 – Inpatient Stays

We identify inpatient stays in each month, both overall and those with a sepsis diagnosis. We also classify inpatient stays by provider type.

### Collapsing Claims to Stays and Constructing Stay-Level Measures

The inpatient file setting in CWF contains claims for provider types including acute inpatient facilities, long-term care hospitals, inpatient rehabilitation facilities, and inpatient psychiatric facilities. While the vast majority of acute inpatient stays are billed using a single claim, some other inpatient provider types, such as inpatient rehabilitation facilities, submit claims on a monthly consolidated billing schedule. This means that a stay spanning multiple months would be represented by multiple claims. In cases when multiple claims are submitted, we group those that share the same beneficiary identifier, provider number (PROVIDER), and admission date (ADMSN\_DT). Medicare payment data is summed across all of the claims in the stay by aggregating the claim “PMT\_AMT” variable, while diagnosis information is retained only from the final claim in the stay.

We also construct a length of stay measure. If the discharge date is available, then the length of stay is set equal to the difference between the discharge date and admission date or ‘1’, whichever is greater. If the discharge date is not available, then the length of stay is set equal to the difference between the final claim’s through date and admission date or ‘1’, whichever is greater.

We identify ER services in the IP claims data using ER revenue center codes. To do this, we check for the presence of a revenue condition code starting with “045” and ending with “0”, ”1”, “2”, “6”, or “9”. If the ER code is not found in the IP data, then we repeat this step using outpatient claims data to check for ER services that are within three days prior to the IP stay admission, including the day of admission.

To identify “shadow billed” no-pay claims for Medicare Advantage (MA) beneficiaries, we look for the presence of related condition code “04” on the IP claims.[[5]](#footnote-5)

The final stay-level data is restricted to paid stays only where total payment (the sum of PMT\_AMT, PRPAYMT, DED\_AMT, COIN\_AMT, BLDDEDAM) is greater than zero.

*Key Variables***:** The table below lists the key variables we use in the inpatient stay creation algorithm.

Table 2.2 Inpatient Stay Creation Algorithm Key Variables

| **Variable Name** | **Description** | **Data Source** |
| --- | --- | --- |
| PROVIDER | Provider Number | CWF |
| ADMSN\_DT | Claim Admission Date | CWF |
| FROM\_DT | Claim From Date | CWF |
| THRU\_DT | Claim Thru Date | CWF |
| DSCHRGDT | Claim Discharge Date | CWF |
| PMT\_AMT | Claim Payment Amount | CWF |
| DRG\_CD | Claim Diagnosis-Related Group | CWF |
| RLT\_COND | Claim Related Condition Code Array | CWF |
| RVCNTR | Claim Revenue Center Code Array | CWF |
| DGNSCD | Claim Diagnosis Code Array | CWF |
| POAINDCD | Claim “Present on Admission” Diagnosis Flag Array | CWF |

### Identifying Sepsis Stays

After collapsing claims to stays and constructing relevant stay-level metrics, we identify sepsis stays in the inpatient data using ICD-9 and ICD-10 codes listed in Table 2.3.

Table 2.3 ICD9 and ICD10 Codes Used to Identify Inpatient Sepsis

|  |  |  |
| --- | --- | --- |
| **Code Set** | **Code Set Description** | **Diagnosis Code List** |
| ICD-9 | Explicit sepsis using ICD-9 codes | 038, 995.91, 995.92, 785.52 |
| ICD-10 Crosswalk | Explicit sepsis using ICD-10 codes obtained using the CMS General Equivalence Mappings (GEMs) for the ICD-9 codes above | A409, A412, A4101 A4102, A411, A403, A414, A4150, A413, A4151, A4152, A4153, A4159, A4189, A419, A419, R6520, R6521 |
| ICD-10 SEP-1 metric | Sepsis defined using ICD-10 codes for the SEP-1 metric | A021, A227, A267, A327, A400, A401, A403, A408, A409, A4101, A4102, A411, A412, A413, A414, A4150, A4151, A4152, A4153, A4159, A4181, A4189, A419, A427, A5486, B377, R6520, R6521 |

We categorize sepsis stays by whether at least one sepsis diagnosis on the claim was flagged as “Present on Admission” as described in Table. 2.4.

Table 2.4 Sepsis “Present on Admission” Assignment Rules

|  |  |
| --- | --- |
| **POA Category** | **Code Set Description** |
| Sepsis Present on Admission | At least one sepsis DGN on the last claim in the stay has a POA indicator of “Yes” (POAINDCD = Y), indicating present on admission. |
| Sepsis Not Present on Admission | All sepsis DGNs on the last claim in the stay have POA indicator of “No” (POAINDCD = N), indicating not present on admission. |
| Sepsis Unknown Presence on Admission | No sepsis DGN on the last claim in the stay has POA indicator of “Yes” (POAINDCD = Y) and not all sepsis DGNs have POA indicator of “No” (POAINDCD = N). This means at least one sepsis DGN had unknown status on admission (POAINDCD has a value other than Y or N). |

We also group IP sepsis stays by severity tiers based on the most severe codes present on the claim as described in Table 2.5.

Table 2.5 Sepsis “Present on Admission” Assignment Rules

|  |  |
| --- | --- |
| **Sepsis Severity Category** | **Code Set Description** |
| Septic Shock | Specified by ICD- code 785.2 or ICD-10 code R6521 |
| Severe Sepsis without Shock | Specified by ICD-9 code 995.92 or ICD-10 code R6520, and no septic shock codes |
| Non-Severe Sepsis, Organism-Specific | Specified by organism-specific sepsis codes and no severe sepsis or septic shock codes |
| Non-Severe Sepsis, Unspecified | Specified by ICD-9 codes 0389 or 995.91 or ICD-10 code A419, and no severe, shock, or organism-specific sepsis codes |

# Implementation: Construct Master Inpatient Stay-Level Dataset

We combine data from Analytic Streams 1 and 2 to construct a master inpatient stay-level dataset that also includes beneficiary enrollment, demographic, and mortality information.

# Implementation: Develop Measures of Health History and Healthcare Service Use Prior to Inpatient Admission

We use hierarchical condition categories (HCC) to estimate the influence of comorbidities on the probability of sepsis hospitalization and the outcomes of hospitalization. CMS-HCC risk adjustment models are used to calculate risk scores that predict individual beneficiaries’ healthcare expenditures relative to the average beneficiary. CMS adjusts Part C payments and bids using risk scores to account for differences in health status (historic diagnoses data) and demographic characteristics (such as gender and age) of enrollees across health plans. The risk adjusted payment is largely based on assignment of diagnoses to disease groups (also known as Condition Categories (CCs)). The Condition Categories are placed into hierarchies (HCCs), reflecting severity and cost dominance. Beneficiaries get credit for the condition with the highest severity or that subsumes the costs of other conditions in the same group. Hierarchies allow for payment based on the most serious conditions when less serious conditions also exist in the respective groups. The associated comorbidities list for each inpatient stay was constructed using diagnoses reported on inpatient, outpatient, and professional claims from 12 months prior to, and inclusive of, the index admission. We use Version 22 of the CMS-HCC risk adjustment model and the diagnosis codes used to define each comorbidity are publically available on CMS’ website [here](https://www.cms.gov/medicare/health-plans/medicareadvtgspecratestats/risk-adjustors.html).

To calculate HCCs, we must be able to observe all healthcare service use for the beneficiary in the year leading up to and including inpatient admission. Thus, beneficiaries are required to have continuous enrollment in Medicare Parts A and B but not C for the 12 months prior to their inpatient admission and the duration of their hospital stay.

We also create preceding healthcare use flags (i.e., preceding nursing facility use, preceding home health care claim) by assessing whether beneficiaries have any inpatient, outpatient, professional, home health, hospice, or SNF claims in the week prior to their acute inpatient admission. We also determine if they have any Long-Term Care Minimum Data Set (MDS) Assessments in this time period indicating that they are receiving unskilled care in a nursing facility in the week prior to their acute inpatient admission. All service types (except for professional claims, so that we could observe professional services associated with a global surgery period, and MDS assessments, which do not include payment data) are restricted to those with positive payment amounts.

# Implementation: Identify Healthcare Service Use Subsequent to an Acute Inpatient Discharge

For two of our analyses, the cost estimation and post-sepsis trajectory analyses, we identify healthcare service use subsequent to an acute inpatient discharge. For the cost analysis, we first require that a beneficiary be eligible to receive subsequent services by restricting to beneficiaries who did not die within a week of their inpatient discharge. For these beneficiaries, we then identify any subsequent SNF stays that have a start date within 90 days of inpatient discharge. We use the following payment variables to calculate total subsequent SNF payment: PMT\_AMT, PRPAYMT, DED\_AMT, COIN\_AMT, BLDDEDAM. We also calculate total SNF beneficiary payments using the following variables: DED\_AMT, COIN\_AMT, BLDDEDAM. Note that we do not impose a clean period restriction that would require beneficiaries to have had no prior IP sepsis stay.

For the trajectories analysis, we restrict IP stays at the beginning of a trajectory to Acute IP stays with start dates between 2012 and 2017 and those for which the beneficiary has no IP stays in the prior year. Subsequent IP, SNF, hospice and home health stays as well as MDS assessments are tracked over the 6-month period after the end of the initial Acute IP stay. All stays are restricted to those with positive payment amounts. The post-sepsis trajectories are restricted to those in which the beneficiary is enrolled in Medicare Parts A and B but not Part C starting from the year prior to the start of their trajectory and continuing up until the beneficiary's death or until end of their 6-month trajectory period.

# Appendix A: Acronyms and Abbreviations

This section lists the acronyms and abbreviations referenced throughout this document.

CMS Centers for Medicare & Medicaid Services

CWF Common Working File

DME Durable Medical Equipment

EDB Enrollment Database

EDS Encounter Data System

ESRD End-Stage Renal Disease

FFS Fee-For-Service

HH Home Health

HS Hospice

IDR Integrated Data Repository

IP Inpatient

MA Medicare Advantage

MDS Long-Term Care Minimum Dataset

MMA Major Metropolitan Area

NPPES National Plan and Provider Enumeration System

OP Outpatient

PB Physician Services

1. Data & Reports.” Centers for Disease Control and Prevention. Centers for Disease Control and Prevention, August 23, 2016. https://www.cdc.gov/sepsis/datareports/index.html. [↑](#footnote-ref-1)
2. “Sepsis.” World Health Organization. World Health Organization. Accessed February 11, 2020. https://www.who.int/news-room/fact-sheets/detail/sepsis. [↑](#footnote-ref-2)
3. Not an official description of the database. [↑](#footnote-ref-3)
4. We consolidate data from the EDB, Common Medicare Environment (CME), and Medicare Beneficiary Database (MBD) to obtain beneficiary information. For simplicity, we collectively refer to data from these sources as EDB information. [↑](#footnote-ref-4)
5. Providers are required to include condition code “04” on claims for MA beneficiaries who are receiving inpatient services. These claims are not paid and instead serve as information purposes only to report the diagnosis codes and services rendered to MA beneficiaries. [↑](#footnote-ref-5)