



## **MedRIC Documentation for HRS Data Requestors**



Acumen, LLC

500 Airport Blvd., Suite 365

Burlingame, CA 94010

## Table of Contents

.....  
.....

<b>1) Project Overview .....</b>	<b>3</b>
1.1 MedRIC .....	3
1.2 MedRIC and the Health and Retirement Study.....	4
<b>2) MedRIC Summary Files .....</b>	<b>4</b>
2.1 Overview .....	4
2.2 The BISF: Specific Features .....	5
<b>3) MedRIC Claims Files .....</b>	<b>6</b>
<b>4) Understanding Claims and Summary Files .....</b>	<b>7</b>
4.1 File Information.....	8
4.2 Documentation Included on Your Media.....	10
<b>5) MedRIC Conversion File .....</b>	<b>10</b>

## List of Tables and Figures

.....  
.....

Table 1: MedRIC Surveys .....	3
Table 2.1: Beneficiary Summary File Categories.....	5
Table 2.2: Gap Type Variable Values.....	6
Table 3: File Elements .....	7
Table 4: File Type Abbreviations .....	8
Table 5: File Layout Versions.....	9
Table 6: Surrogate UPINs.....	9

# 1) Project Overview

## 1.1 MedRIC

Since 1984, the Centers for Medicare and Medicaid Services (CMS) have produced increasingly detailed confidential claims data that permit analysis of longitudinal health care utilization by the elderly at the level of specific contacts with physicians and other health providers. Current methods of survey data collection and interest in the changing demographics, health status, and economic condition of the elderly and near-elderly have resulted in an ever increasing amount of data on Medicare-eligible populations, including the disabled and certain ethnic/racial groups. These data have created great potential for understanding health care utilization among Medicare beneficiaries.

At the same time, use of these data requires substantial investment of time and resources, in addition to considerable staff knowledge of the specifics of Medicare claims processing and reimbursement regulations. Relatively few survey projects outside of CMS have linked their respondents to Medicare administrative data, and fewer even have done so on a consistent and ongoing basis. These shortcomings have limited access to only the most committed projects with expert staff and budgets large enough for data purchase and manipulation.

The Medicare Research Information Center (MedRIC) directly responds to these needs by facilitating the acquisition and linkage of CMS data to surveys and registries sponsored by federal agencies, and by promoting the use of these data for research and public policy analysis. MedRIC's intent is to establish an infrastructure that will reduce the expense and time required to obtain, create, and use research files of CMS data for individual projects. The overriding goal of MedRIC is to vastly reduce the costs and burdens of survey/registry projects on individual researchers in obtaining and using information from Medicare records. MedRIC is currently collaborating with four NIH surveys and survey users that intend to use the MedRIC files for their research:

**Table 1: MedRIC Surveys**

<b>Survey Name</b>	<b>Brief Description</b>
University of Michigan Health and Retirement Study (HRS)	A survey of more than 22,000 Americans over the age of 50 conducted every two years that paints an emerging portrait of an aging America.
Baltimore Longitudinal Study of Aging (BLSA)	Begun in 1958, a study of more than 1,400 participants ranging in age from 20 to 90 years old.
Panel Study of Income Dynamics (PSID)	Begun in 1968, a study collecting data from approximately 8,000 American families with regards to economic, health, and social behavior.
The Dynamics of Health, Aging and Body Composition (HEALTH ABC)	A study of the extent of change in body composition in 3,075 men and women aged 70-79.

## 1.2 MedRIC and the Health and Retirement Study

MedRIC's integration of Medicare and survey/registry data provides a rich set of informational resources for studying how aspects of health status influence economic circumstances and life-style of the elderly, and how these circumstances determine utilization of medical services.

The University of Michigan Health and Retirement Study (HRS), established in 1992, surveys more than 22,000 Americans over the age of fifty every two years. Supported by the National Institute on Aging and the Social Security Administration, the study paints an emerging portrait of an aging America's physical and mental health, insurance coverage, financial status, family support systems, labor market status, and retirement planning.

MedRIC has been helping the HRS achieve one of its main goals of understanding the relationship between medical history and financial status, and how use of health care may change as people age. With MedRIC's support, the HRS endeavors to obtain information about health care costs and diagnoses from Medicare records maintained by CMS. The HRS asks all respondents who are eligible for Medicare to provide their identification numbers; over 80% of them consent do so. MedRIC summarizes respondents' Medicare information for HRS requestors, creating a series of files that improves requestors' ability to analyze Medicare data for their surveys.

## 2) MedRIC Summary Files

MedRIC has created several files that summarize a series of Medicare data across time. These files, summarized at the beneficiary level for different time periods, offer user-friendly data structures specifically designed for research purposes. Below, the Beneficiary Annual Summary File (BASF), the Beneficiary Quarterly Summary File (BQSF), and the Beneficiary Interview Gap Summary File (BISF) are described and their use explained.

### 2.1 Overview

The MedRIC summary files are derived from Medicare Parts A and B data. Medicare Part A covers hospital insurance; most participants do not pay a premium because they or a spouse already paid for it through payroll taxes while working. Part A helps cover inpatient care at hospitals, including critical access hospitals and skilled nursing facilities, but not custodial or long-term care facilities; it also helps cover hospice care and some home health care. Medicare Part B covers medical insurance for which most participants pay a monthly premium. Part B helps cover doctors' services and outpatient care, as well as certain other medical services not covered by Part A, such as some physical and occupational therapist services, and some home health care services. Part B helps pay for these covered services and supplies when they are medically necessary.

All three summary files condense Medicare Part A and B claims and enrollment information into one observation per beneficiary in a specific period of time. Individually as well as collectively, they provide comprehensive summary measures related to health and medical expenditures for each type of service and payer, along with

utilization measures associated with these services. These files do so without revealing any personally or demographically-identifiable information.

To better understand the data, the variables are grouped into five principal categories, which contain coverage information, diagnosis-related group (DRG) information, beneficiary categorical condition (CC) and categorical condition group (CCG) information by year, expenditure and utilization information, and CMS Chronic Condition Warehouse information.

**Table 2.1: Beneficiary Summary File Categories**

<b>Summary File Information</b>	<b>Brief Description</b>
Coverage information	Coverage variables for every beneficiary summarize Medicare status code and months of enrollment.
Diagnostic-related groups	Diagnostic-related groups relate types of patients to the resources they consume. Currently, there are approximately 500 DRG's.
RDDC Categorical Condition Group	All valid diagnosis codes per beneficiary are mapped to one of 184 Categorical Conditions, which are in turn mapped to 30 RDDC Categorical Condition Groups.
Expenditure and utilization	Expenditure and utilization variables describe fee-for-service reimbursements for services covered by CMS Standard Analytical File claim categories.
Chronic Condition Warehouse	CMS Chronic Condition Warehouse variables are constructed according to the algorithm provided by the Iowa Foundation for Medical Care; the algorithm consists of clinical and coverage criteria, as well as yearly and ever flag indicators

All five information types are found in the three summary files, the principal difference being the time period across which each file is summarized. The BASF summarizes beneficiary-level data for the entire calendar year into one record. The BQSF does the same, but on a quarterly level. The BISF summarizes the beneficiary-level information from one interview to the next, referred to as an interview gap. With three available summary files, requestors are able to select the file(s) that best complement and are best tailored to their research objectives.

## 2.2 The BISF: Specific Features

What distinguishes the BISF from the other two MedRIC summary files is its customization to the HRS (Health and Retirement Survey) data, as the BISF is arranged by the interview dates of the HRS. The beginning of a summary period of a beneficiary is noted by the variable *Start\_dt*; correspondingly, the conclusion of a period is given by *End\_dt*. While *Start\_dt* and *End\_dt* are formatted to include a day, month, and year, the days of each *Start\_dt* and *End\_dt* are always the first and last days of the month, respectively.

In order to provide the potential for stronger and more meaningful analysis of this information, we have added two new variables to it which differentiate between the

starting and closing dates of interviews, assigning them values based on the particular event which caused an interviewee to either first participate in the HRS or to complete the interview process. The variables *Start\_Gap\_Type* and *End\_Gap\_Type* correspond to a beneficiary's first and final dates of interview, respectively. *Start\_Gap\_Type* and *End\_Gap\_Type* may each take on only one of four potential values. These include 'N', indicating a new enrollee, 'I', signifying only an interview gap, 'C', showing that a gap has been summed up to the point of current data, and 'D', indicating death. The new addition of gap type variables was initiated for its enrichment of analysis capability beyond the base framework of the claims data structure, adding a unique nuance which could be used by researchers to potentially explore cause and effect relationships originating from the Medicare data in a more complex and salient manner.

**Table 2.2: Gap Type Variable Values**

<b>Gap Value</b>	<b>Description</b>
N	New Enrollment
I	Interview Gap
D	Death Termination
C	Summed to point of current data

Please also note that exit interviews which occurred after death were included.

### **3) MedRIC Claims Files**

MedRIC also supplies CMS claims data to its requestors. The MedRIC claims files are constructed directly from the Medicare Standard Analytical Files (SAFs) and the Denominator File, the only difference being that certain variables are masked or removed. Using the files therefore presents the same obstacles as working directly with the claims data from CMS.

MedRIC distributes Medicare SAFs and the Denominator file, whose elements are listed in the table below (Table 3: File Elements). All files contain final action claims data, meaning that all adjustments within a calendar year have been resolved.

**Table 3: File Elements**

<b>File</b>	<b>Type of file</b>	<b>Information included</b>
Denominator file	Beneficiary enrollment information in a calendar year.	Beneficiary unique identifier, monthly entitlement indicators (A/B/Both), reasons for entitlement, state buy-in indicators, and monthly managed care indicators (yes/no).
Inpatient SAF	Final action claims data submitted by inpatient hospital providers for reimbursement of facility costs.	ICD-9 diagnosis and procedure codes, Diagnosis Related Groups (DRG's), dates of service, and reimbursement amounts.
Skilled Nursing Facility SAF	Final action claims data submitted by SNF providers.	ICD-9 diagnosis and procedure codes, dates of service, and reimbursement amounts.
Outpatient SAF	Final action claims data submitted by institutional outpatient providers.	ICD-9 diagnosis and procedure codes, CMS Common Procedure Coding System (HCPCS), dates of service, and reimbursement amounts.
Home Health Agency SAF	Final action claims data submitted by HHA providers.	Number of visits, type of visit (skilled-nursing care, home health aides, physical therapy, speech therapy, occupational therapy, and medical social services), ICD-9 diagnosis codes, dates of visits, and reimbursement amounts.
Carrier SAF (old file name: Physician/Supplier Part B)	Final action claims data submitted by non-institutional providers.	ICD-9 diagnosis and procedure codes, CMS Common Procedure Coding System (HCPCS), dates of service, and reimbursement amounts.
Hospice SAF	Final action claims data submitted by Hospice providers.	Level of hospice care received (e.g., routine home care, inpatient respite care), ICD-9 terminal diagnosis, dates of service, and reimbursement amounts.
Durable Medical Equipment SAF	Final action claims data submitted by Durable Medical Equipment suppliers.	ICD-9 diagnosis codes, services provided (CMS Common Procedure Coding System (HCPCS)), dates of service, and reimbursement amounts.
MEDPAR File	Inpatient hospital and skilled nursing facility (SNF) final action stay records.	Inpatient "stay" record summarizes all services rendered to a beneficiary from the time of admission to a facility through discharge. Each record may represent one claim or multiple claims, depending on length of beneficiary's stay and amount of inpatient services used throughout stay.

#### **4) Understanding Claims and Summary Files**

This section includes information on the files and documentation found on your media. The first section outlines useful information regarding the file naming conventions and the UPINs found in the data. The second section list the documents included on your media, such as Data Dictionaries and Record Counts.

## 4.1 File Information

### A. File Naming Conventions

Each file is identified by file type, data year, partition, and format; each file is in the form *FileType\_Year\_Partition.sas7bdat*. The following naming conventions are observed for these four categories:

- i. **FileType** indicates the type of data which is present in the data file, with abbreviations:

**Table 3: File Type Abbreviations**

Abbreviation	File Name
DM	Durable Medical Equipment
DN	Denominator
HH	Home Health Agency
HS	Hospice
IP	Inpatient
OP	Outpatient
MP	MedPAR, Calendar Year
PB	Carrier
SN	Skilled Nursing Facility
BASF	Beneficiary Annual Summary File
BQSF	Beneficiary Quarterly Summary File
BISF	Beneficiary Interview Gap Summary File

- ii. **Year** specifies the year from which the data in the file comes.
- iii. **Partition** indicates into how many partitions a file has been divided to make it more manageable. A partition value of 1 indicates a single-partitioned file. Higher numbers indicate which partition is contained in a particular file.
- iv. **Format** indicates the format of the file, which in this case is SAS, and appears as *sas7bdat*.

### B. File Layout Versions

MedRIC ships Standard Analytical Files (SAFs) in two different layout versions known as Versions H and I. Most files are shipped in the Version I layout, but several file types from older years are shipped in the Version H layout. Below is listed the years in which each layout is used, by file element. This table is useful when referencing the Data Dictionary for a specific file and year.

**Table 4: File Layout Versions**

<b>File element:</b>	<b>Version H files:</b>	<b>Version I files:</b>
Skilled Nursing Facility	1998, 1999	1991-1997, 2000-current
Inpatient	1998, 1999	1991-1997, 2000-current
Home Health	1998, 1999	1991-1997, 2000- current
Outpatient	1997, 1998	1991- current
Hospice	None	1991- current
Carrier	None	1991- current
Durable Medical Equipment	None	1991- current

**C. Surrogate Unique Physician Identification Numbers**

Unique Physician Identification Numbers (UPINs) are 6-place alpha numeric identifiers assigned to all physicians. MedRIC does not encrypt Surrogate UPINs, since they are temporarily assigned and therefore do not identify an individual provider. The table below maps Surrogate UPINs to provider type.

**Table 5: Surrogate UPINs**

<b>Surrogate UPIN</b>	<b>Provider Type</b>
INT000	for each intern
RES000	for each resident
PHS000	for Public Health Service physicians, includes Indian Health Services
VAD000	for Department of Veterans Affairs physicians
RET000	for retired physicians
SLF000	for providers to report that the patient is self-referred
OTH000	for all other unspecified entities not included above
AA0000	Anesthesia Assistant
CNA000	Certified Registered Nurse Anesthetist
CNM000	Certified Nurse Midwife
CNS000	Clinical Nurse Specialist
CP0000	Clinical Psychologist
CSW000	Clinical Social Worker
FOR000	Foreign Doctor (for all non-United States physicians)
MD0000	Medical Doctor (includes DO, CH, DDM, DDS, DPM, OD)
NP0000	Nurse Practitioner
OT0000	Occupational Therapist
PA0000	Physician Assistant
PT0000	Physical Therapist

## 4.2 Documentation Included on Your Media

### A. Data Dictionaries

The data dictionaries list and define all variables and include code values and meanings, variable attributes, variable masking and special notes. The data dictionaries are provided in both .xls and .html formats. Record counts and numbers of unique beneficiaries are also included on this document. The shipment should be verified against the record counts as soon as they are received.

### B. Segmented Claims

MedRIC claims-level data are *flat files*, meaning the number of variables per record is fixed. That is, for a particular year and file type (such as Inpatient or Carrier), each record contains a common set and number of variables, but some variables may be blank.

In some cases, a claim is segmented into two or more records. This occurs when there are more values in a particular variable group than available spaces. Sometimes a claim will be segmented while showing empty variables in the first record; this occurs because claim-segment files are built to accommodate different versions of CMS data with inconsistent numbers of variables stored into a single record.

### C. Decryption Document

The data files are compressed and encrypted using a 256-bit encryption algorithm. This document describes the software used to encrypt the data files and instructions on how to access them.

### D. Cover Letter

The cover letter provides contact information for any questions or concerns regarding accessing your data files.

## 5) MedRIC Conversion File

Some HRS data users may have studies that require a larger beneficiary population where as some may require a more confident beneficiary linkage. Additional variables have been added to the conversion file displaying whether there was a match between the Medicare enrollment information and HRS finder file. The match variables include the beneficiary's first name, last name, year of birth, month of birth, day of birth, zip code and gender and indicate whether the information matched (1), did not match (0), or was missing (M). These variables allow the data users to apply their own precise specifications to build their beneficiary cohorts.