

2020 | DATA USER'S GUIDE: COST SUPPLEMENT FILE



Centers for Medicare and Medicaid Services (CMS)
Office of Enterprise Data and Analytics (OEDA)

VERSION CONTROL LOG

Date	Version	Revisions
10/20/2022	1.0	Initial version published.

TABLE OF CONTENTS

INTRODUCTION	1
Contents of the Data User's Guide – Cost Supplement File	1
What's New in 2020?	2
Questionnaire Changes	2
Data Processing	3
1. FILE STRUCTURE	4
1.1 LDS Contents.....	4
1.2 File Structure	5
2. DATA FILE DOCUMENTATION.....	11
2.1 Contents of the LDS	11
2.2 LDS Components	11
2.2.1 Codebooks	11
2.2.2 Questionnaires.....	12
2.3 Data Edits and Imputation	12
2.3.1 Data Edits	12
2.3.2 Imputation	13
3. DATA FILE NOTES.....	15
3.1 Global Information	15
3.1.1 Key Variables.....	15
3.1.2 Missing Values	15
3.1.3 Other Specify Questions	15
3.1.4 Analytic Notes for Non PM Event Segments.....	16
3.1.5 Exclusion of Home Health Utilization and Costs	16
3.2 Cost Supplement File Segment Information.....	16
3.3 Event-Level Segments	16
3.3.1 Dental Utilization Events (DUE)	17
3.3.2 Facility Events (FAE)	17
3.3.3 Hearing Utilization Events (HUE)	19
3.3.4 Inpatient Hospital Events (IPE)	19
3.3.5 Institutional Events (IUE)	19
3.3.6 Medical Provider Events (MPE).....	20
3.3.7 Outpatient Hospital Events (OPE).....	21

3.3.8 Prescribed Medicine Events (PME)	21
3.3.9 Vision Utilization Events (VUE)	22
3.4 Summary-Level Segments	23
3.4.1 Service Summary (SS).....	23
3.4.2 Person Summary (PS)	24
3.5 Weights.....	24
4. REFERENCES.....	26
5. APPENDICES.....	28
Appendix A: MCBS Common Definitions.....	28
Appendix B: MCBS Rounds by Data Year and Season	34
Appendix C: Technical Appendices	35
C.1 Using the Data	35
C.2 Matching Survey and Administrative Data	48
C.3 Imputation Information.....	56
C.4 Changes to Note in Prior Data Years	71
Appendix D: Table of Links to MCBS Documentation	73
Appendix E: 2020 MCBS Cost Supplement File Segments and Historic RIC Segments.....	74

LIST OF EXHIBITS

Exhibit 1.1.1:	Medicare Population Covered by the 2020 MCBS Data	4
Exhibit 1.1.2:	MCBS Populations in Data Products	5
Exhibit 1.2.1:	2020 Contents of Cost Supplement File Data Release	7
Exhibit 1.2.2:	2020 Cost Supplement File Segments by Data Source.....	10
Exhibit 2.3.1:	Data Review and Missing Data	13
Exhibit 2.3.2:	Summary of Event-Level Imputation for LDS Segments	14
Exhibit C.1.1:	Example 1 Recoding Variables.....	36
Exhibit C.1.2:	Example 4 Recoding Variables.....	40
Exhibit C.2.3.3:	Overview of event category matches conducted during event-level matching.....	52
Exhibit C.3.2.8:	Establishing Total Payment Amount and Allocating to Sources of Payers.....	63
Exhibit C.4.1:	Changes to Align Variables.....	72

ACRONYM LIST

ANDA	Abbreviated New Drug Application
BLA	Biologics License Application
CMS	Centers for Medicare & Medicaid Services
CPS	Charge Payment Summary Questionnaire
CSEVWGTS	Cost Supplement File Ever Enrolled weights
CSL2WGTS	Cost Supplement File Longitudinal weights (2-year)
CSL3WGTS	Cost Supplement File Longitudinal weights (3-year)
DME	Durable Medical Equipment
DRG	Diagnosis-related group
DUE	Dental Utilization Events segment
DVH	Dental, Vision, and Hearing Care Utilization Questionnaire
EOBs	Explanation of Benefit Statements
ERQ	Emergency Room Utilization Questionnaire
ESRD	End-stage renal disease
FAE	Facility Events segment
FDB	First Databank
FFS	Fee-for-Service
HCPCS	Healthcare Common Procedure Coding System
HHS	Home Health Summary Questionnaire
HHQ	Home Health Utilization Questionnaire
HICN	Health Insurance Claim Number
HMO	Health Maintenance Organization
HUE	Hearing Utilization Events segment
ID	Identification
IPE	Inpatient Hospital Events segment
IPQ	Inpatient Hospital Utilization Questionnaire
IUE	Institutional Events segment
IUQ	Institutional Utilization Questionnaire
LDS	Limited Data Set(s)
MA	Medicare Advantage
MAPD	Medicare Advantage Part D Plan
MCO	Managed Care Organization
MCBS	Medicare Current Beneficiary Survey
MP	Medical Provider events
MPQ	Medical Provider Utilization Questionnaire
MPE	Medical Provider Events segment
NCH	National Claims History repository
NDA	New Drug Application
NDC	National Drug Code
NIR	Not-In-Round
Non PM	Non Prescription Medicine
NORC	NORC at the University of Chicago
NSQ	No Statement Charge Questionnaire
OEDA	Office of Enterprise Data and Analytics
OM	Other Medical Expenses events
OMQ	Other Medical Expenses Utilization Questionnaire
OPE	Outpatient Hospital Events segment
OPQ	Outpatient Hospital Utilization Questionnaire

PDE	Part D Event
PDP	Part D Plan
PM	Prescription Medicine
PME	Prescription Medicine Events segment
PMFORM	Prescribed Medicine Form
PMQ	Prescribed Medicine Utilization Questionnaire
PPO	Preferred Provider Organization
PS	Person Summary segment
PUF	Public Use File
QMB	Qualified Medicare Beneficiary
RIC	Record Identification Code
SAS	Statistical Analysis System
SD	Separately Billing Doctor events
SL	Separately Billing Lab events
SLMB	Specified Low-Income Medicare Beneficiaries
SNF	Skilled Nursing Facility
SOP	Source of Payment
SOWMP	Survey Only With Medicare Payment
SS	Service Summary segment
STQ	Statement Cost Series Questionnaire
US	Use of Health Services Questionnaire
USCARE	Usual Source of Care segment
VA	Veterans Administration
VUE	Vision Utilization Events segment

INTRODUCTION

The Medicare Current Beneficiary Survey (MCBS)¹ Cost Supplement File provides cost and utilization data that can be linked to the MCBS Survey File to conduct analyses on health care costs and utilization for the beneficiaries in the survey. Beginning with the release of 2015 data files, Cost Supplement File users will require the Survey File for information on beneficiaries' demographic characteristics, survey-reported covariates, and health insurance information. In keeping with best practices of data management, each variable will exist on only one MCBS Limited Data Set (LDS) (i.e., the Survey File or Cost Supplement File).

The MCBS Cost Supplement File links Medicare claims to survey-reported events and aims to provide complete expenditure and source of payment data on all health care services, including those not covered by Medicare. The MCBS Cost Supplement File provides a comprehensive picture of health services received, amounts paid, and sources of payment. The file can support a broader range of research and policy analyses on the Medicare population than would be possible using either survey data or administrative claims data alone.

The MCBS Cost Supplement File provides cost and utilization data that can be linked to the MCBS Survey File to conduct analyses on health care costs and utilization for the beneficiaries in the survey.

The Cost Supplement File undergoes a careful reconciliation process to separately identify and flag health care services reported: 1) from the survey alone, 2) from the Medicare Fee-for-Service (FFS) and Part D claims data alone, and 3) from both sources. Survey-reported data include information on the cost and utilization of all types of medical services including those not covered by Medicare, as well as payments by supplementary health insurance and Medicare Part C/Medicare Advantage (MA). Medicare claims data include cost and utilization information on inpatient hospitalizations, outpatient hospital care, physician services, durable medical equipment, skilled nursing home services, hospice care, and prescription drugs.²

In general, the Cost Supplement File will be released 12 to 15 months after data collection has ended and final administrative and claims data for that calendar year become available. CMS will release the Cost Supplement File approximately three months after the Survey File, with the Survey File typically released in the summer and the Cost Supplement File in the fall.

For questions about the Cost Supplement data release, please contact MCBS@cms.hhs.gov.

CONTENTS OF THE DATA USER'S GUIDE – COST SUPPLEMENT FILE

This manual contains detailed information about the Cost Supplement File and specific background information to help data users in understanding and analyzing the data. A companion *Data User's Guide* focuses on the Survey File release and can be accessed at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks>.

Here is an overview of the contents of the *2020 Data User's Guide: Cost Supplement File*:

¹ The MCBS is authorized by section 1875 (42 USC 139511) of the Social Security Act and is conducted by NORC at the University of Chicago for the U.S. Department of Health and Human Services. The OMB Number for this survey is 0938-0568.

² Data on home health care are typically available in the Cost Supplement File LDS but are not available in 2020 due to a questionnaire redesign, which is discussed below in more detail.

- What's New in 2020 – This section describes key MCBS Questionnaire changes and other highlights for the 2020 data year.
- Section 1: File Structure – This section includes a technical description of the specifications and structure of the file and a brief description of the record types in the file.
- Section 2: Data File Documentation – This section provides information on the content of the data file including instructions on reading in the data.
- Section 3: Data File Notes – This section provides an overview of each file included in the release, a description of derived variables, and any changes from previous releases or special highlights for data users.

WHAT'S NEW IN 2020?

As in prior years, for 2020 MCBS data, CMS will release a Public Use File (PUF) and two LDS. The data within the LDS releases are organized into segments (formerly RICs). In this Guide, Appendix E provides a crosswalk from historical RIC segments to the 2020 segment names.

Questionnaire Changes

Below, data users will note highlights and updates for the 2020 data year related to the Cost Supplement File LDS. Please see the *Data User's Guide: Survey File* for additional changes that occurred in the data year and for detailed information about the 2020 MCBS Survey File LDS.

Community Questionnaire

With the emergence of the coronavirus disease 2019 (COVID-19) pandemic, data collection in Winter 2020 shifted from in-person to telephone in mid-March 2020. For the Summer 2020 round, collection of utilization and cost data in the Community Questionnaire was temporarily suspended so that additional protocol changes could be made to ease phone administration. Utilization and cost data collection was reinstated in Fall 2020 and a longer reference period was used to capture utilization data not collected during the Summer 2020 interview. To support the collection of cost information by phone in Fall 2020, a new response option was added to several screens in the Statement Cost Series (STQ), the No Statement Charge (NSQ), and the Charge Payment Summary (CPS) that allowed interviewers to skip out of the charge series in rare situations where the respondent experienced extreme frustration, fatigue, or otherwise could not provide cost information over the telephone. If selected, the questionnaire routed directly to the next section after the charge series.

In response to the COVID-19 pandemic, CMS expanded access to Medicare telehealth services so that beneficiaries can receive a wider range of services from their health care providers without having to travel to a health care facility.³ Due to the expansion of telehealth coverage, a follow-up question was added to the Medical Provider Utilization (MPQ), STQ, and NSQ sections starting in Fall 2020 to capture if a medical provider event was a telephone or video visit. This item allows data users to distinguish telemedicine visits from in-person provider visits.

Also in 2020, the Home Health Utilization (HHQ) section was redesigned and modeled after the MPQ section to support the creation of a home health event-level segment, which will first be available in the 2021 Cost Supplement File LDS release. This effort included several questionnaire changes including the following:

- To facilitate implementation of the redesigned HHQ and ease burden on analytic processing, the Home Health Summary (HHS) and HHQ sections were not fielded in the Winter or Summer 2020, thereby ensuring that all home health data collected in the 2020 data year will be from the redesigned HHQ.

³ "Medicare Telemedicine Health Care Provider Fact Sheet," Centers for Medicare & Medicaid Services, March 17, 2020, <https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet>.

- In Fall 2020, the HHS section was permanently deleted. Respondents receiving ongoing care from the same home health provider will record that information in HHQ as a new event, just as respondents do for other ongoing utilization.

Facility Instrument

Two changes were made to the Use of Health Services (US) section of the Facility Instrument in Fall 2020 in response to the COVID-19 pandemic:

- Some facilities had to temporarily or indefinitely suspend certain types of in-person health care services to prevent the spread of the virus. The US section asks if the beneficiary received health care services during the reference period for medical doctor visits, dental visits, mental health visits, physical therapist or other types of therapist visits, podiatrist visits, educational and habilitational services, and other types of health care provider visits. To measure the impact of the suspension of health care services on the beneficiary's health care utilization, a new response option was added throughout the US section for these health care services. In addition to responding "YES" or "NO", "SERVICE SUSPENDED DUE TO COVID-19" was added to these health care service questions to capture if a specific health care service type was suspended due to COVID-19.
- With the expansion of telehealth coverage due to the COVID-19 pandemic, additional question text at health care service questions and follow-up questions were added throughout the US section to measure whether any reported health care services were telehealth visits. The additional question text and new follow-up questions were added for medical doctor visits, dental visits, mental health visits, physical therapist or other types of therapist visits, podiatrist visits, educational and habilitational services, and other types of health care provider visits, in addition to already existing follow-up questions about the number of visits and duration of care.

Data Processing

Preloads

In Fall 2020 (Round 88), Winter 2021 (Round 89), and Summer 2021 (Round 90), the Community preload process was modified to prevent the questionnaire from prompting for additional information for certain charges in the CPS section. This modification was implemented to reduce potential respondent and interviewer burden associated with collecting many events and costs via phone for beneficiaries with high utilization.

Exclusion of Home Health Utilization and Costs

Because home health utilization data were not collected in Winter 2020 (Round 86) or Summer 2020 (Round 87), the 2020 Cost Supplement File does not contain home health utilization and cost data. The initial round of the redesigned HHQ section was fielded beginning in Fall 2020 (Round 88) and will be included in the 2021 Cost Supplement File.

Imputation

The imputation of missing costs and payments was improved for MA beneficiaries in the 2020 LDS. Adjustments based on MA encounter data, which were first applied for 2019, now better reflect beneficiary age and general health-related differences.

In addition, imputation for medical provider events now account for the presence of telehealth visits.

1. FILE STRUCTURE

1.1 LDS CONTENTS

The following information is represented in the MCBS Cost Supplement File:

- Survey-reported data, including information on the cost and utilization of medical services, which contains all sources of payment and out-of-pocket costs.
- Medicare FFS and prescription drug claims data including administrative and billing information on the cost and utilization of inpatient hospitalizations, outpatient hospital care, physician services, durable medical equipment, skilled nursing home services, hospice care, and prescription drugs.
- MA cost and utilization information: When a respondent reports health care events, the explanation of benefits form (EOBs) from their MA provider is used to report the payments. This is the same approach taken for services that are not covered by Medicare, such as most dental care. Actual claims-based information for MA beneficiaries, referred to as encounter data, are not currently available for these individual events.

There are two Medicare populations represented in the MCBS LDS's: the ever enrolled and the continuously enrolled population. Exhibit 1.1.1 displays the key differences between these populations.

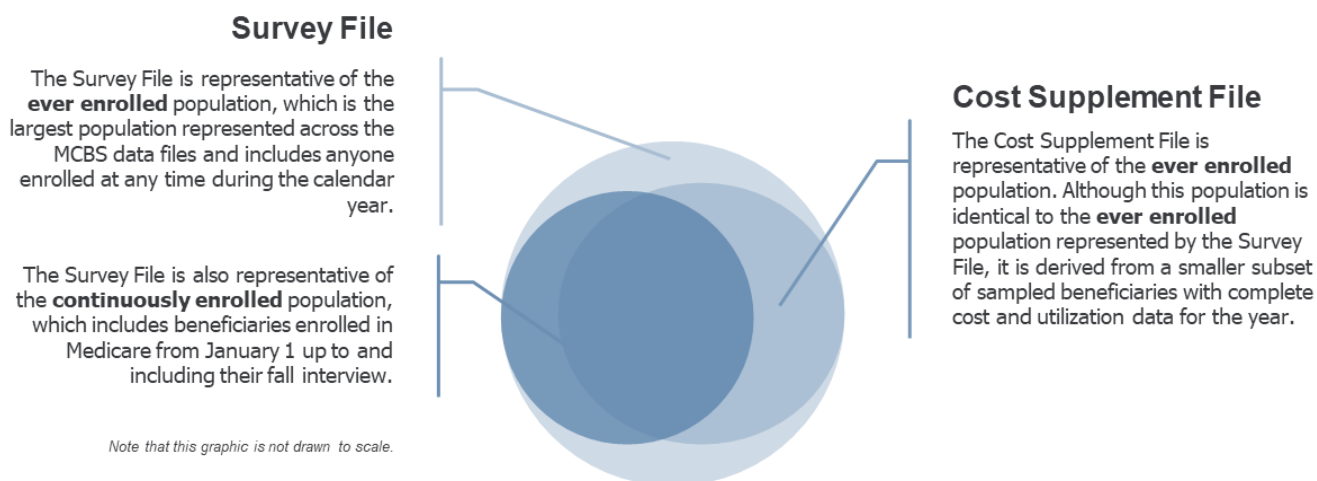
Exhibit 1.1.1: Medicare Population Covered by the 2020 MCBS Data

Ever Enrolled Population	Continuously Enrolled Population
<ul style="list-style-type: none"> ■ A random cross-section of all beneficiaries who were ever enrolled in either Part A or Part B of the Medicare program for any portion of 2020 (i.e., for at least one day at any point during 2020) 	<ul style="list-style-type: none"> ■ A subset of the ever enrolled population who were continuously enrolled from January 1, 2020 up to and including interviews conducted during Fall 2020
<ul style="list-style-type: none"> ■ Represents individuals who were enrolled in Medicare at any time during the calendar year 	<ul style="list-style-type: none"> ■ Represents only those individuals who were continuously enrolled in Medicare from January 1, 2020 up to and including their fall interview
<ul style="list-style-type: none"> ■ Includes beneficiaries who enrolled during 2020 ■ Includes beneficiaries who dis-enrolled or died prior to their fall interview 	<ul style="list-style-type: none"> ■ Excludes beneficiaries who enrolled after January 1, 2020 ■ Excludes beneficiaries who dis-enrolled or died prior to their fall interview

Exhibit 1.1.2 depicts the relationship between the beneficiaries included in the annual data releases. The ever enrolled population from the Survey File is the largest, including anyone enrolled at any time during the calendar year. The continuously enrolled population is limited to those beneficiaries who were enrolled from January 1 of the survey year through the fall interview date. The Survey File includes a weights segment that allows for subsetting the data by the ever enrolled and continuously enrolled populations. The Cost Supplement File ever enrolled population is identical to the Survey File ever enrolled population, but represents a smaller subset of beneficiaries with complete cost and utilization data for the year. The use of analytic weights ensures that each population is appropriately represented.⁴

⁴ Exhibit 1.1.2 is not drawn to scale, but provided as a visual reference for the relationship of populations between data files.

Exhibit 1.1.2: MCBS Populations in Data Products



1.2 FILE STRUCTURE

The Cost Supplement File data are provided at three different levels of summarization: Event-level, Service Summary (SS) level, and Person Summary (PS) level. The tri-level structure allows researchers to potentially avoid having to process all the detailed event records in the file when summaries may suffice. For example, an analysis of differences in total health spending per person between males and females could use the PS level and thereby avoid having to process the numerous event-level records. Similarly, an analysis of differences in use of Medicare hospital payments by race/ethnicity could use the SS records. Researchers could use event-level records for more detailed analyses, such as looking at average length of long-term facility stays or average reimbursements per prescription drug type. Users should determine whether a segment that contains already summarized costs, utilization, and payment distributions would best serve the analysis.

The Cost Supplement segments are assembled at three levels:

1. The **event-level** reports all payers, costs, and utilization at the most detailed level available. Service types at the event-level are dental (DUE), facility (FAE), hearing (HUE), medical provider (MPE), inpatient hospital (IPE), outpatient hospital (OPE), institutional (IUE), prescription medicine (PME), and vision (VUE). Note that hospice services are only included at the service and person level; there are no event-level data provided for these services. For the provided service types, the event-level data has a record for each reported medical event for each person.
2. The **Service Summary level** summarizes all payers, costs, and utilization for a person during the calendar year at the service level. There are 10 service categories: dental, facility, hearing, hospice, medical provider, inpatient hospital, institutional, outpatient hospital, prescribed medicines, and vision. The data include a record for each of the 10 service categories for each person. Within each type of service record, separate payer totals for 11 different payers are also shown.

Payer totals are summarized in two ways: one summarizing the event-level records with no adjustment, and the other is summarized adjusting for survey interview gaps that occurred for Medicare covered days within the year. For example, if a beneficiary visited the dentist for an annual checkup and again for a cavity filling, all costs and utilization for these two dental visits would be summarized in one row of data under the dental category for this beneficiary. If the dental costs were covered by different payers, the overall cost will be split by payer in the Service Summary record. The adjusted totals correct for any survey interviewing gaps during the year.

The service summaries exclude unmatched survey event records that are considered duplicative of unmatched Medicare bill record events [see Event-Level Matching and Adjusting for Missing Days and Undated Services in Appendix C: Technical Appendices]. Starting with 2019, the adjusted totals also include an upward ratio adjustment to Non PM utilization and expenditure data for beneficiaries enrolled in MA plans in the data year 2020.

3. The **Person Summary level** summarizes all payers and costs across service categories and summarizes type of service amounts. The data include only one record for each person, which shows the totals for each service and payer for that person. Again, payment amounts are shown two ways: as summarized from event records and adjusted to compensate for Medicare covered days that were not covered by interview reference periods.

As with the Service Summary level, the person summaries exclude unmatched survey event records that are considered duplicative of unmatched Medicare bill record events. In addition, starting with the 2019 data year, the adjusted totals include an upward ratio adjustment to Non PM utilization and expenditure data for beneficiaries enrolled in MA plans in the data year 2020.

All MCBS records begin with the same three variables: the survey reference year, version, and a unique number that identifies the person who was sampled (the BASEID). The BASEID provides the link to the other segments within the Cost Supplement and Survey File LDS releases. To obtain complete survey information for an individual, a researcher must link together records for that individual from the various data files using the variable BASEID. Beneficiaries may not have a record on every data file. Exhibit 1.2.1 provides an overview of the Cost Supplement File segments.

The MCBS Cost Supplement File segments contain both survey-reported data and administrative claims data on health service utilization, costs, and payers. Exhibit 1.2.2 provides a mapping of the data sources that contribute data to each event-level and summary segment, highlighting that there are differences in data source by beneficiary residence status and service type.

Exhibit 1.2.1: 2020 Contents of Cost Supplement File Data Release

Cost Supp. Segment (Abbrev)	Description	Data collection and special weights notes	Respondent Type*	Data Source**	Quex Section	Season	Panel***	Unit of Observation
Dental Utilization Events (DUE)	Contains individual dental events reported during a Community interview or created from Medicare claims data.		B	CQ, AR	DVH	All	All	One record per beneficiary per event (defined as a single visit to the dentist)
Facility Events (FAE)	Contains individual facility events reported during a Facility interview.	There is one record for each stay that occurred at least partly in the data year.	F	FI, AR	RH, US, EX	All	All	One record per beneficiary per stay in a long-term care facility
Hearing Utilization Events (HUE)	Contains individual hearing care events reported during a Community interview or created from Medicare claims data.		B	CQ, AR	DVH, US	All	All	One record per beneficiary per event (defined as a single visit to a hearing care provider)
Inpatient Hospital Events (IPE)	Contains individual inpatient hospital events reported during a Community interview or created from Medicare claims data.		B	CQ, AR	IUQ, IPQ, ERQ, OPQ	All	All	One record per beneficiary per admission
Institutional Events (IUE)	Contains individual short-term facility (usually skilled nursing facility) stays that were reported during a Community interview or created from Medicare claims data.		B	CQ, AR	IUQ, IPQ	All	All	One record per beneficiary per admission

Cost Supp. Segment (Abbrev)	Description	Data collection and special weights notes	Respondent Type*	Data Source**	Quex Section	Season	Panel***	Unit of Observation
Medical Provider Events (MPE)	Contains individual events for a variety of medical services, equipment, and supplies reported during a Community interview or created from Medicare claims data.		B	CQ, AR	ERQ, IPQ, MPQ, OMQ, OPQ	All	All	One record per beneficiary per event, defined as a separate visit, procedure, service, or a supplied item for a survey-reported event
Outpatient Hospital Events (OPE)	Contains individual outpatient hospital events reported during a Community interview or created from Medicare claims data.		B	CQ, AR	OPQ	All	All	One record per beneficiary per event (defined as a single outpatient visit)
Prescribed Medicine Events (PME)	Contains individual outpatient prescribed medicine events reported during a Community interview or created from Medicare claims data.		B	CQ, AR	PMQ, DVH, ERQ, IPQ, OPQ, MPQ	All	All	One record per beneficiary per prescribed medicine (defined as a single prescribed medicine)
Vision Utilization Events (VUE)	Contains individual vision care events reported during a Community interview or created from Medicare claims data.		B	CQ, AR	DVH, US	All	All	One record per beneficiary per event (defined as a single visit to a vision care provider)
Person Summary (PS)	Summarization of utilization and expenditures by type of service and summarization of expenditures by payer.		B	CQ, FI, AR	All utilization including HHQ, US	All	All	One record per beneficiary
Service Summary (SS)	Summarization of the nine individual event files along with hospice utilization.		B	CQ, FI, AR	All utilization including US	All	All	10 records per beneficiary

Cost Supp. Segment (Abbrev)	Description	Data collection and special weights notes	Respondent Type*	Data Source**	Quex Section	Season	Panel***	Unit of Observation
Cost Supplement Ever Enrolled Weights (CSEVWGTS)	Contains cross-sectional full-sample and replicate weights representing the 2020 ever enrolled population.		B	CQ/FI	N/A	All	All	One record per beneficiary
Cost Supplement Longitudinal Weights (CSL2WGTS CSL3WGTS)	Contains longitudinal full-sample and replicate weights for the multi-year ever enrolled population. The CSL2WGTS file includes the two-year longitudinal weights for the population ever enrolled at any time during both 2019 and 2020. The CSL3WGTS file includes the three-year longitudinal weights for the population ever enrolled at any time during 2018, 2019, and 2020.		B	CQ/FI	N/A	All	All	One record per beneficiary

* = Respondent type describes the expected setting where beneficiaries resided during the calendar year (i.e., C = respondent only completed Community interviews, F = a Facility interview was conducted, or B = respondents completed at least one Community interview and for whom at least one Facility interview was conducted). In each data year, some differences by segment will exist (i.e., data may reflect a prior or future calendar year due to the specific questionnaire and reference period used to collect the information).

** = Data source describes the source of the data on the segment. The three possible sources are the Community Questionnaire (CQ), Facility Instrument (FI), and Administrative Records (AR). Each LDS segment can have any combination of these sources.

*** = Panel describes whether the questionnaire sections that provide the data for each segment are fielded for baseline respondents, continuing respondents, or both.

Exhibit 1.2.2: 2020 Cost Supplement File Segments by Data Source

Event-Level LDS Segment	Beneficiary Residence Status		Community			Facility			Included in LDS Summary Segments (SS & PS)	
	Sources for Utilization and Cost Data		Survey Reported	Claims*		Survey Reported	Claims*			
	Beneficiary Health Insurance Status			FFS	Part D		FFS	Part D		
	Service Type									
Costs for services received outside of a Facility that are paid for by FFS or Part D are contained in the other non FAE event-level segments.	DUE	Dental utilization events		X	+			+		X
	HUE	Hearing utilization events		X	+			+		X
	IPE	Inpatient hospital events		X	X			X		X
	IUE	Institutional events		X	X			X		X
	MPE	Medical services, equipment, and supplies		X	X	X		X	X	X
	OPE	Outpatient hospital events		X	X			X		X
	PME	Prescribed medicine events		X		X			X	X
	VUE	Vision utilization events		X	+			+		X
FAE	Nursing home or other long-term care facility	• Cost of Facility Stay				X	X^		X	
		• Services provided Within the Facility				X				
		• Services received Outside of the Facility				X				
Not available at event-level	Hospice care events		X	X			X		X	

*Claims are only available for events paid for by Medicare FFS and Part D. Administrative data pertaining to events paid for by MA, Medicaid, private insurance, or out of pocket are not available to the MCBS. Historic MA Encounter Data is used during imputation for some event types; see Section 2.3.2 for more information.

+ FFS Medicare does not cover most procedures or supplies for dental, vision, and hearing segments. It only covers certain services that are received in a hospital. The segment includes claims data for the few services Medicare does cover, but the vast majority of events are Community survey-reported only.

^ Skilled Nursing Facility (SNF) and Inpatient (IP) Medicare FFS claims are used during data processing of the FAE segment to impute missing Medicare payment amounts and check the accuracy of the lengths of stays, when available. However, claims-only records are not added to FAE whereas they are added for the other event-level segments. See Appendix C for more information on claims matching and imputation processes.

2. DATA FILE DOCUMENTATION

2.1 CONTENTS OF THE LDS

The Cost Supplement File consists of a series of separate files known as segments. In addition to the segments, CMS provides technical documentation with the following resources for data users:

- Codebooks
- Questionnaires
- Data files (SAS®, CSV)
- Format control files

CMS provides technical assistance to researchers interested in using MCBS data and provides free consultation to users interested in obtaining these data products and using these data in research. Users can email MCBS@cms.hhs.gov with questions regarding obtaining and using the data.

2.2 LDS COMPONENTS

2.2.1 Codebooks

Codebooks are included with each data release and serve as the key resource for comprehensive information on all variables within a data file. The codebooks list the variables in each of the segments, the possible values, and unweighted frequencies. For variables that are associated with items in the MCBS Questionnaire, the item number and item text are provided.

The following information is provided within each Codebook:

Variable: The codebook contains the variable names associated with the final version of the data files. Certain conventions apply to the variable names. All variables that are preceded by the character "D_", such as D_SMPTYP, are derived variables. Variables preceded by the character "H_", such as H_DOB, come from CMS administrative source files.

Format Name: This column identifies the format name associated with the variable in the SAS dataset.

Frequency: This column shows unweighted frequency counts of values or recodes for each variable.

Label (variable label and codes): The variable label provides an explanation of the variable, which describes it more explicitly than would be possible in only eight letters. For coded variables, all of the possible values of the variable appear in lines beneath that explanation. Associated with each possible value (in the column labeled "Frequency") is a count of the number of times that the variable had that value, and, under the column labeled "Label", a short format expanding on the coded value.

Version Number: Files may be re-released due to needed updates, which will be noted by the version number variable.

Survey Year: The Survey Year of interest is included as a variable on the file.

BASEID: The BASEID is the unique identifier assigned to each beneficiary. This identifier can be used to link data across the survey files.

Notes:

The Cost Supplement data come from summarized information from MCBS Questionnaire items as well as imputed data. On Cost Supplement File segments, data users will find imputation flag variables where applicable.

Many questionnaire items were posed to elicit simple "Yes" or "No" answers, or to limit responses to one choice from a list of categories. In these cases, the responses are "Yes" or "No" or one of the codes from the list. In other questions, the respondent was given a list of items to choose from, and all of the responses were recorded. In these cases, each of the responses is coded "Indicated"/"Selected" or "Not Indicated"/"Not Selected".

If a respondent provided an answer that was not on the list of possible choices, it was recorded verbatim. All of the verbatim responses were reviewed and categorized. New codes were added to the original list of options to accommodate verbatim responses that appeared frequently. For this reason, the list of possible values for some variables may not exactly match the questionnaire.

2.2.2 Questionnaires

Data users can view the questionnaire for each data year along with questionnaire variable names and question text on the MCBS website at: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Questionnaires>.

2.3 DATA EDITS AND IMPUTATION

2.3.1 Data Edits

A series of checks and edits are conducted in order to ensure the accuracy, completeness, and reasonableness of data within each data file. Any structural issues are addressed during either data extraction or data cleaning.

Logic checks verify that the questionnaire worked as expected, particularly with respect to questionnaire routing. Errors identified during logic checking are addressed with two categories of data edits: flagging values that were incorrectly skipped and setting incorrectly populated values to null to indicate a valid missing value.

Additional checks identify unreasonable values that are not explicitly disallowed by the questionnaire (e.g., male beneficiaries reporting female-only conditions, such as cervical cancer). After investigation, such values are then addressed with edits. The MCBS also conducts consistency checks to identify scenarios where respondents report inconsistent information (e.g., indicating that one is Medicaid eligible due to a certain condition, but not reporting having that same condition when asked about health status).

Based on a thorough data review, these types of errors are addressed with edits during data cleaning.

Certain conventions are used in coding all variables to distinguish between questions that beneficiaries would not or could not answer and questions that were not asked. These conventional codes are depicted in Exhibit 2.3.1.

Exhibit 2.3.1: Data Review and Missing Data

Value	Format	Meaning
.	INAPPLICABLE	Valid missing, inapplicable, a valid skip, missing with no expectation that a value should be present (missing is '.' in numeric variables and blank in character variables)
.R	REFUSED	Valid missing, refused survey response
.D	DON'T KNOW	Valid missing, don't know survey response
.N	INVALID SKIP	Invalid missing, not ascertained, an invalid skip, a response should be present but is not
.E	EDITING CODE	Editing code, extreme value, unreasonable or out of range survey response

2.3.2 Imputation

In order to compile the most accurate and complete LDS, there are several types of adjustments applied to the MCBS Cost Supplement data that compensate for missing information. Although a variety of methods are used in making the adjustments, adjustments of all types are governed by some basic principles. First, information reported by the survey respondent is retained, even if it is not complete, unless strong evidence suggests that it is not accurate. Second, when information is not reported during the interview, Medicare claims data and administrative data are the first choice as a source of supplementary, or in some cases surrogate, information.

Imputation occurs across three levels (see Exhibit 2.3.2). First, payment amounts for individual sources of payment (such as Medicare or private insurance) are imputed. These imputed values are reflected in the event-level data. Next, a MA Encounter Data Ratio Adjustment, which was introduced for the 2019 Cost Supplement File LDS, accounts for medical events that were not reported by survey respondents who were covered by MA. This adjustment was applied within the outpatient, inpatient, institutional, and medical procedure LDS segments. Finally, a part-year ratio adjustment is applied to account for reference period gaps within the calendar year which are usually caused by Not-In-Round (NIRs).⁵ The adjustments are applied to the event-level data and published in the service-level and person-level summaries. These processes do not impute values for Hospice events. Imputation was performed using the hot deck imputation method, and a flag was created for each imputed variable indicating whether or not the corresponding value was imputed.

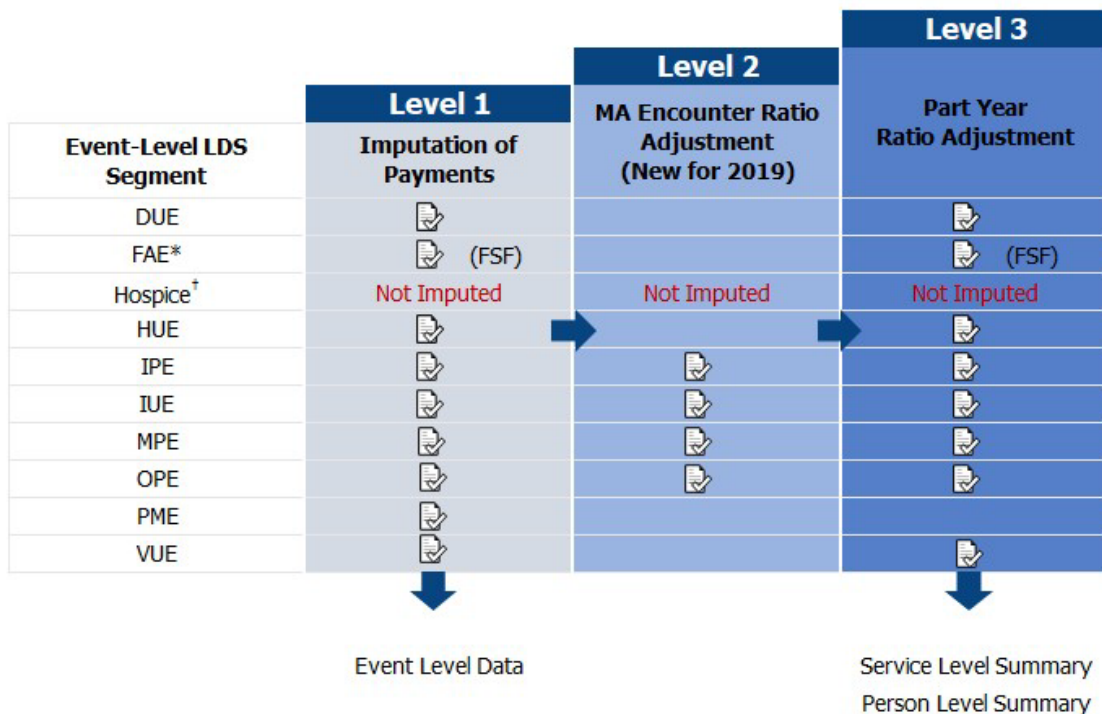
Beginning with 2019, the MA Encounter Data were utilized to improve estimation of medical events and costs for those beneficiaries enrolled in MA. The goal was to account for unreported MA medical events. To achieve that, an estimate of the ratio of utilization counts from matched event data to the utilization counts reported in the survey was calculated. Since the MA Encounter Data were not available to be matched to survey reported events for the current data year, Encounter Data from the prior three calendar years were matched to survey reported events over that period. Multipliers calculated to estimate this adjustment were applied to 2020 payment amounts for the events of MA beneficiaries during their MA enrollment periods. Beginning with the 2019 data year, the adjusted variables within the Service

***Case Example for Part-Year Ratio Adjustment:** A BASEID completed an interview on 11/15/20 and then did not complete an interview in Winter 2021. The BASEID provided cost data through 11/15/20. Payments would be adjusted in the summary files to account for costs between 11/16/20 and 12/31/20.*

⁵ Not-In-Round refers to cases in which the respondent was not available to be interviewed within the round's time frame.

Summary and Person Summary files reflect this MA adjustment combined with the part-year ratio adjustment.

Exhibit 2.3.2: Summary of Event-Level Imputation for LDS Segments



* Imputation in the Facility Stay File (FSF) occurs via a separate, simpler process.

† The MCBS adds Hospice administrative data to the files but does not conduct imputation.

Users may find additional information on how the imputation was conducted in the *2020 MCBS Methodology Report* located on CMS' MCBS website at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks>. Researchers may also find more information on the imputation process in Appendix C: Technical Appendices of this document. In particular, Appendix C.3 provides detailed information on filling in missing payments and payers for matched and non matched records as well as the types of edits made to resolve events.

3. DATA FILE NOTES

This section is a collection of information about various data fields present in the Cost Supplement File, beginning with information that is applicable globally, followed by specific information on some individual data fields, presented in the same sequence as the data fields appear in the codebook.

3.1 GLOBAL INFORMATION

3.1.1 Key Variables

There are several variables that appear on each segment in order to merge segments and/or identify each segment.

- **BASEID:** This key identifies the person interviewed. It is an 8-digit element, consisting of a unique, randomly assigned 8-digit number. The BASEID is the key to merge data segments within an LDS or across LDS releases.
- **Version Number:** Often files are re-released due to needed updates, which will be noted by the version number variable.
- **Survey Year:** The Survey Year of interest is included as a variable on each segment.

LDS segments may vary in the number of BASEIDs. This variation may occur for several reasons. First, some segments include data from Community questionnaires and others from Facility questionnaires with different numbers of beneficiaries providing responses. Second, there are also differences in the number of beneficiaries by the specific round completed. Third, the use of ever enrolled or continuously enrolled weights in constructing the segments may result in differences.

3.1.2 Missing Values

Various special values indicate the reason why some data are missing, such as .R for "refused," .D for "don't know." These conventional codes are depicted in Exhibit 2.3.1.

3.1.3 Other Specify Questions

A subset of MCBS questionnaire items include closed ended responses with "other specify" options. These options allow respondents to provide answers that are not included in the existing code frame and are useful for questions with a wide range of potential responses (e.g., types of problems experienced during attempts to obtain care). If an "other specify" option is selected, interviewers record actual responses verbatim.

In data processing, "other specify" responses that are sufficiently similar to existing code frame options are recategorized. This is accomplished by identifying keywords and misspellings corresponding to each existing response option, programmatically searching the verbatim "other specify" responses for the keywords and misspellings, and then categorizing responses into existing response option categories as appropriate. Codes are then assigned to similar responses to facilitate analysis; there are no verbatim responses provided on the released segments. Often there will be more than one answer to a single question. In these cases, responses are recoded into several variables, all of which contain categorized data. Code lists are updated, when necessary, to incorporate responses that are frequently provided in "other specify" response options.

3.1.4 Analytic Notes for Non PM Event Segments

The Inpatient Hospital Events, Institutional Events, and Outpatient Hospital Events segments include a Survey Only with Medicare Payment (SOWMP) variable. This flag variable was only set to 1 if the FFS administrative claims data were not matched to the survey-reported event data and there was a survey-reported Medicare payment. However, if there was a reported or imputed MA payment for these unmatched records with Medicare payments, the SOWMP flag was not set to 1. The conclusion is that when SOWMP is set to 1, the Medicare payment for these unmatched, survey-reported events was a traditional Medicare FFS payment, which would have already been included in the matched claims or claims-only records. Thus, these payments likely duplicate those claims, so the SOWMP flag was set to 1, and these records were excluded from the summary-level segments (PS and SS).

3.1.5 Exclusion of Home Health Utilization and Costs

Home health utilization data were not collected in Winter 2020 (Round 86) or Summer 2020 (Round 87), so the 2020 Cost Supplement File does not contain home health utilization and cost data. The initial round of the redesigned HHQ section was fielded beginning in Fall 2020 (Round 88) and will be included in the 2021 Cost Supplement File.

3.2 COST SUPPLEMENT FILE SEGMENT INFORMATION

Information regarding each segment within the Cost Supplement File release is listed below. The segments are presented in alphabetical order. The notes have been organized into three main categories of information.

1. Core Content – a description of the main subject of the data.
2. Variable Definitions – definitions of derived variables and/or variables that require additional explanation regarding their construction.
3. Special Notes – additional background information that data users may find helpful for conducting analyses.

3.3 EVENT-LEVEL SEGMENTS

The Event-level segments provide detailed information on all payers, costs, and utilization for the MCBS population.

Each Event-level segment contains the following variables.

- AMTTOT Amount paid by all payers
- AMTCARE Amount paid by Medicare FFS
- AMTCAID Amount paid by Medicaid
- AMTHMOP Amount paid by private Managed Care Organization (MCO)/HMO
- AMTMADV Amount paid by Medicare MCO/HMO
- AMTPRVE Amount paid by insurance – employer-sponsored
- AMTPRVI Amount paid by insurance – self-purchased
- AMTPRVU Amount paid by private insurance (unknown purchase)
- AMTOOP Amount paid out of pocket
- AMTDISC Discounted amount - beginning in 2013 includes sequestration reduction
- AMTOTH Amount paid by other sources, including VA

There are two additional variables that appear on the Non PM Event-level segments only.

- AMTCOV Portion of the total payment covered by Medicare
- SITCODE Whether the event occurred in a Community or Facility setting

Please note that SITCODE is not included on FAE. In addition, the only payer variables from this series available on the FAE are AMTTOT, AMTCAID, AMTPRVU, AMTTOOP, and AMTOTH. Lastly, in the FAE segment, AMTCARE is renamed to AMTUCARE to denote that AMTCARE in other segments represents a different value, FFS Medicare payments.

For both PM and Non PM events, the totals are adjusted for gap days. Gap days are periods during the calendar year in which a beneficiary was enrolled in Medicare but not covered by a survey interview. Event-level segments contain imputation flags to identify the presence of imputed amounts.

Begin date variables (D_BEGxx) and end date variables (D_ENDxx) are derived by overwriting survey-reported information with claims data when applicable. These variables are included on a majority of the event-level segments:

- DUE (D_BEGxx only)
- HUE (D_BEGxx only)
- IPE (D_BEGxx and D_ENDxx)
- IUE (D_BEGxx and D_ENDxx)
- MPE (D_BEGxx only)
- OPE (D_BEGxx only)
- VUE (D_BEGxx only)

3.3.1 Dental Utilization Events (DUE)

Core Content

This file contains individual dental events for the MCBS population. The unit of observation is a single visit to the dentist, at which time a variety of services, including cleaning, x-rays, and an exam might be rendered. Most of the information on this segment is survey-reported from the Community Questionnaire. There is no survey-reported information from the Facility Instrument on this segment.

Medicare does not cover most dental procedures or supplies, like cleanings, fillings, dentures, or other dental devices. Medicare Part A will pay for certain dental services that are delivered in the hospital. The few claims that are used on this segment are from Part A claims for qualifying procedures.

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

N/A

3.3.2 Facility Events (FAE)

Core Content

This segment includes individual facility events for the MCBS population recorded during a Facility interview. There is one record for each stay that occurred at least partly in the data year (i.e., the stay begins, ends, or continues through the calendar year). The unit of measurement of facility services is a "stay" in a nursing

home or other long-term care facility. Stays are measured in terms of days of residence in that facility. If a person is still in the facility at the end of 2020, the stay is not complete, but all data through the end of 2020 are included.

Variable Definitions

AMTUCARE: The amount paid by Medicare to the facility that is not included in any of the other Event records. For instance, most doctor visits that occurred while the person was in the facility and were paid for by Medicare FFS are in the MPE segment; however, if the facility reported an amount received by Medicare that exceeded the total Medicare amounts on the Event segments, then the Medicare amount reported by the facility that is in excess of the other events' Medicare amounts is reported here.

AMTTOT: The sum of the five facility payer types (AMTUCARE, AMTCAID, AMTPRVU, AMTOOP, AMTOTH). Note that according to the above explanation of AMTUCARE, this amount is not duplicated in the other Event segments.

AMTOTH: Given the definition of TOTCARE, AMTOTH is the total amount paid for the person while in the facility by other payers (i.e., payments not attributed to Medicare, Medicaid, or other private payers). In the case of missing information where imputed amounts are needed, AMTOTH could contain either the total payment amount or the additional amount by which payment amounts were increased.

REFBEGYY, REFBEGMM, REFBEGDD: The earliest date in the calendar year when the beneficiary was in the facility.

REFENDYY, REFENDMM, REFENDDD: The last date in the calendar year when the beneficiary was in the facility.

TOTCARE: The total amount paid by Medicare FFS while the person was in the facility, which includes all Medicare amounts from other Event segments that occurred during the person's facility stay. Additionally, it includes any amount reported by the facility that is in excess of the other events' Medicare amounts (AMTUCARE).

TOTALL: The sum of TOTCARE, AMTCAID, AMTPRVU, AMTOOP, AMTOTH.

Special Notes

In 2020, a new item series related to telehealth services was added to the Use of Health Services (US) section of the Facility Instrument as part of the beneficiary-level COVID-19 Facility Supplement (see the *What's New in 2020* section). This resulted in four new variables on the FAE segment indicating if the beneficiary had any telehealth visits during a given stay with a medical doctor (MDTEL), dentist (DENTTEL), mental health professional (MENTTEL), or other provider (OTHRTEL).

Stays are defined as any period of time when the beneficiary lived in a facility for one or more days and had complete Facility interview data. New stays are generated for a beneficiary whenever they move to a new facility and complete an interview with the new facility. If the beneficiary left the facility for a period greater than 30 days and returned to the facility, a separate stay record was created.

3.3.3 Hearing Utilization Events (HUE)

Core Content

This segment contains individual hearing events for the MCBS population. The unit of observation is a single visit to a hearing care provider, such as an ear and nose throat doctor or audiologist. A variety of services may be rendered during a hearing event, including a hearing exam, a hearing aid fitting, repair, or purchase, or hearing rehabilitative services. Note that any hearing aid purchases are classified as OM events rather than HU events, similar to durable medical equipment purchases. Only the HU visits themselves are assigned an HU event type. The majority of the information on this segment is survey-reported from the Community Questionnaire. There is no survey-reported information from the Facility Instrument on this segment.

Medicare does not cover most hearing procedures or supplies, like hearing exams or hearing aids. Medicare Part A will pay for certain hearing services that are delivered in the hospital. The few claims that are used on this segment are from Part A claims for qualifying procedures.

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

N/A

3.3.4 Inpatient Hospital Events (IPE)

Core Content

This segment contains individual inpatient hospital events for the MCBS population that are reported during a Community interview or created from Medicare claims data. The unit of observation of inpatient hospital services is a single admission. If the beneficiary was still hospitalized at the end of the year, the inpatient event record is not complete, but all data through the end of 2020 are present.

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

Note that only the principal diagnosis code, PRINDIAG, is included on the IPE and IUE segments, and the principal procedure code, PRCDRCD1, is only included on the IPE segment. Please consult the CMS claims included in the Survey File LDS for any additional diagnosis or procedure codes associated with the event records. Please note that research claims only include FFS events.

3.3.5 Institutional Events (IUE)

Core Content

This segment contains individual short-term facility (usually SNF) stays for the MCBS population that are reported during a Community interview or created from Medicare claims data. The unit of observation is a single admission. If the beneficiary was still in the institution at the end of the year, the institutional event is not complete, but all data for 2020 are present.

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

Note that the IPE and IUE segments report only the principal diagnosis code, PRINDIAG. Please consult the CMS claims included in the Survey File LDS for any additional diagnosis or procedure codes associated with the event records. Please note that research claims only include FFS events.

3.3.6 Medical Provider Events (MPE)

Core Content

This segment contains individual events for a variety of medical services, equipment, and supplies reported during a Community interview or created from Medicare claims data and Medicare Part D claims data. The unit of observation is a separate visit, procedure, service, or a supplied item for a survey-reported event. For Medicare claim-only events, it may represent 1) single or multiple visits; 2) single or multiple procedures; 3) single or multiple services; or 4) single or multiple supplies, depending on the number of items pulled together on the bill.

MPE is a combination of medical provider events collected in the Community Questionnaire: medical provider (MP), separately billing doctor (SD), separately billing lab (SL), and other medical expenses (OM).

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

The EVNTTYPE variable distinguishes between the different event types. The classifications of EVNTTYPEs are determined by how the respondent reported the event during the survey. For example, a respondent may report an MP event type and total costs associated with it. This may match a Medicare claim with a line-item cost for the physician visit and a line-item cost for a lab service. In this case, there would not be a SL event.

When an event matches a Medicare claim or Medicare Part D claim, efforts are made to preserve some of the cost classifications that the claims line items explain through the Healthcare Common Procedure Coding System (HCPCS) code. These groupings are found in several variables:

- PAMTMED (physician costs)
- PAMTSURG (surgical costs)
- PAMTLABX (laboratory and x-ray costs)
- PAMTOM (other medical costs such as DME)
- PAMTPM (prescribed medicine costs)

The costs above reflect total reimbursements and sum to AMTTOT. These variables will only have data for matched survey events and claim-only events.

A new variable related to telephone or video visits, also called telehealth visits, was added to the questionnaire in Fall 2020 (Round 88). TELEHLTH was added to the MPQ, NSQ, and STQ sections and asks the beneficiary whether any of their doctor visits reported in those sections were telehealth visits.

3.3.7 Outpatient Hospital Events (OPE)

Core Content

This segment contains individual outpatient hospital events for the MCBS population reported during a Community interview or created from Medicare claims data. The unit of observation is a separate visit to any part of the outpatient department for a survey-reported event. For Medicare claim-only events, it may represent 1) a single visit; 2) multiple procedures or services within one visit; or 3) multiple visits billed together.

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

The primary diagnosis code, PRINDIAG, is present on the file. For any other additional diagnosis codes associated with an event, please consult the CMS claims included in the Survey File LDS.

As discussed above under MPE, a new variable related to telehealth visits was added in Fall 2020 to the MPQ, NSQ, and STQ sections and asks whether any of the beneficiary's medical provider visits were telephone or video visits. A variable with this information (TELEHLTH) is also released on OPE for outpatient hospital events that were originally reported as medical provider events but reassigned based on claims data.

3.3.8 Prescribed Medicine Events (PME)

Core Content

This segment contains individual outpatient prescribed medicine events for the MCBS population reported during a Community interview or created from Medicare Part D claims data. The unit of observation is a single purchase/fill of a single drug in a single container. The segment also contains the names of the prescribed medicines, the form, and unit of strength.

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

Some of the variables in this record are only applicable in certain situations during the interview. The following variable is only applicable when the form of the medication is a pill or a patch:

- TABNUM Number of tablets/patches in the container

Note: STRNNUM1 and STRNUNI1 were not released on the 2020 Cost Supplement File.

The following questions are asked of the respondent when the medication's dosage form is not a pill, a patch, or a suppository:

- AMTUNIT Amount unit
- AMTNUM Amount number
- SUPPNUM Inapplicable unless the dosage form is a suppository

Often, drug characteristics are imputed to assist in assigning pricing data. The imputed dosage form was only imputed when there was no match between what was reported and the possible dosage forms found in First Databank (FDB) or if the form was missing. The value of PMFORM (Prescribed Medicine Form, which is the type of medicine prescribed, such as pills, liquid, injection, etc.) was also changed when the imputed dosage form was present. The imputed strength and the amount number were imputed using various criteria and contributed to determining a unit price only. The presence of imputed amounts is identified via flags on this segment.

The following variables are unadjusted totals for the beneficiary. These totals only partially account for any gap days (days not covered by interview). While survey data was not available, Part D administrative claim amounts were included, thus any interview gap period would be partially covered by these Prescription Drug Event (PDE) administrative claims.

■ AMTTOT	Amount paid by all payers
■ AMTCARE	Amount paid by Medicare FFS/Part D
■ AMTCAID	Amount paid by Medicaid
■ AMTHMOP	Amount paid by private MCO/HMO
■ AMTMADV	Amount paid by Medicare MCO/HMO
■ AMTPRVE	Amount paid by insurance - employer sponsored
■ AMTPRVI	Amount paid by insurance - self purchased
■ AMTPRVU	Amount paid by private insurance (Unknown Purchase)
■ AMTOOP	Amount paid out of pocket
■ AMTDISC	Discounted amount
■ AMTOTH	Amount paid by other sources, including VA

Part B drug information: A small number of Part B drugs are collected as survey-reported data in the PME. However, the data added from claims are only from Part D. There are no survey-reported drugs administered by a physician matched from the Part B administrative claims data.

In order to determine whether a drug is brand name or generic, data users can compare the FDB FDB_BN field with the FDB_GNN field. If these fields differ, then it is potentially a brand name drug (or at least has a trademarked name).

Data users can also use the PDE National Drug Code (NDC) and use an external drug information database (like FDB) to determine brand vs. generic status of the drug.

If data users do not have a drug database, they could use the Food and Drug Administration's NDC SPL Data Elements File. CMS uses this for the Manufacturer drug discount program to determine what products are not eligible for the Manufacturer discount. Based on the Marketing category, drugs can be classified as follows:

Brand: NDA (New Drug Application), NDA authorized generic, BLA (Biologics License Application)

Generic: ANDA (Abbreviated New Drug Application)

3.3.9 Vision Utilization Events (VUE)

Core Content

This segment contains individual vision events for the MCBS population. The unit of observation is a single visit to a vision care provider, such as an optometrist or an optician. A variety of services may be rendered during a vision event, including a vision exam, a contact lens fitting or purchase, an eye glass frame fitting or purchase, and different kinds of surgeries (e.g., cataract, corneal, etc.). Note that any vision purchases (e.g., contacts, eyeglasses, etc.) are classified as OM events rather than VU events, similar to durable medical equipment

purchases. Only the VU visits themselves are assigned a VU event type. The majority of the information on this segment is survey-reported from the Community Questionnaire. There is no survey-reported information from the Facility Instrument on this segment.

Medicare does not cover most vision procedures or supplies, like eye exams or contact lenses. Medicare Part A will pay for certain vision services that are delivered in the hospital. The few claims that are used on this segment are from Part A claims for qualifying procedures.

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

N/A

3.4 SUMMARY-LEVEL SEGMENTS

The summary-level segments summarize all payers, costs, and utilization for a beneficiary at the service-level or person-level, respectively.

3.4.1 Service Summary (SS)

Core Content

This segment provides a summary of the nine individual event files along with hospice utilization, yielding a total of 10 summary records per person.

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

For every person, there are 10 records: one record for each of the nine event types plus one additional record which is not present at the event-level for hospice services. The records are identifiable by the EVNTTYPE variable:

- DU - Dental
- FA - Facility
- HP - Hospice
- HU - Hearing
- IP - Inpatient hospital
- IU - Institutional
- MP - Medical provider
- OP - Outpatient hospital
- PM - Prescribed medicine
- VU - Vision

When linking event-level Non PM data to service-level data, any survey-reported event that specified traditional Medicare as a payer and was not matched to a FFS Medicare claim was excluded from the Type of Service summary. The analysis showed that either 1) the survey event's monies are bundled with a Medicare

claim that already matched another survey event or 2) the respondent was incorrect in reporting Medicare as a payer.

3.4.2 Person Summary (PS)

Core Content

This segment provides a summarization of utilization and expenditures by type of service and a summarization of expenditures by payer, yielding one record per person.

Variable Definitions

Please see the Codebook for information regarding variables in this segment.

Special Notes

N/A

3.5 WEIGHTS

Core Content

The Cost Supplement weights segments include full-sample and replicate weights representing the 2020 ever enrolled population (CSEVWGTS). There are also segments containing two-year longitudinal weights (CSL2WGTS) and three-year longitudinal weights (CSL3WGTS).

Ever Enrolled Cross-Sectional Weights (CSEVWGTS)

Given that the Cost Supplement population represents an ever enrolled population enrolled in Medicare on at least one day at any time in 2020, the Cost Supplement LDS contains cross-sectional weights for the ever enrolled population only (CSEVWGTS). The population represented by the sum of CSEVRWGT on the CSEVWGTS segment in the Cost Supplement is identical to the population represented by the sum of the ever enrolled Survey File weight, but it is populated for a smaller subset of beneficiaries from the 2017, 2018, 2019, and 2020 Panels who provided utilization and cost data for 2020. The CSEVWGTS segment includes full-sample and replicate weights and may be used to conduct joint analyses of MCBS 2020 Survey File data and MCBS 2020 Cost Supplement data.

Two-Year Longitudinal Weights (CSL2WGTS)

The two-year longitudinal weights are populated for members of the 2017, 2018, and 2019 Panels who were ever enrolled in Medicare at any time during both 2019 and 2020 and provided utilization and cost data for both years. Please note that data users cannot use the Survey File longitudinal weights with Cost Supplement data.

Three-Year Longitudinal Weights (CSL3WGTS)

The three-year longitudinal weights are populated for members of the 2017 and 2018 Panels who were ever enrolled in Medicare at any time during 2018, 2019, and 2020 and provided utilization and cost data for all three years. Please note that data users cannot use the Survey File longitudinal weights with Cost Supplement data.

Variable Definitions

CSEVWGTS contains BASEID, which can be used to merge the weights to other data files, in addition to variables for variance stratum and unit defined for variance estimation using Taylor series linearization (SUDSTRAT, SUDUNIT), and the final full-sample weights and replicate weights.

Special Notes

CSEVWGTS includes records for beneficiaries who were sampled in the 2017, 2018, 2019, and 2020 Panels. The 2020 cross-sectional weights are populated for all records, including members of all four panels. The 2017, 2018, and 2019 Panels are referred to as "Continuing Panels" and provided survey-reported cost and utilization for 2020 through participation in the MCBS during Fall 2019 through Winter 2021 rounds. Members of the 2020 Panel who were first selected for participation in the MCBS in 2020 are referred to as Incoming Panel respondents. They were first interviewed in Fall 2020 and did not provide cost and utilization data for the period of time between enrollment and completion of the Fall 2020 interview; cost and utilization data for the period between the Fall 2020 interview and the end of 2020 were collected in Winter 2021. A combination of the survey-collected data for the end of the year and Medicare claims data was used to impute beneficiary-level data for the entire period of enrollment in 2020. The resulting final cross-sectional weights (CSEVRWGT/CSEVR001-CSEVR100), which include both the Continuing and Incoming Panel respondents, represent the population of beneficiaries that were ever enrolled in Medicare for at least one day at any time during 2020.

Users do not need to apply formats to the datasets as the weights are real numbers and do not need to be grouped or labeled.

The Survey File longitudinal weights cannot be used for analysis of Cost Supplement data because the beneficiaries eligible for inclusion in the file are different. The Cost Supplement File is representative of the ever enrolled population of beneficiaries (i.e., ever enrolled in Medicare for at least one day at any time during the data year), but contains a smaller subset of beneficiaries with complete cost and utilization data for the year.

A detailed discussion of the weights construction process is provided in the *2020 MCBS Methodology Report* available on the CMS MCBS website at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks>. Additionally, examples of using the MCBS weights are provided below in *Appendix C: C.1 Using the Data*.

4. REFERENCES

Centers for Disease Control and Prevention. "Telehealth Interventions to Improve Chronic Disease." Last modified May 11, 2020. <https://www.cdc.gov/dhdsp/pubs/telehealth.htm>.

Centers for Disease Control and Prevention. "How to Protect Yourself & Others." Last modified February 25, 2022. <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html>.

Centers for Medicare & Medicaid Services. "Medicare Telemedicine Health Care Provider Fact Sheet." March 17, 2020. <https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet>.

Eppig, F. J., and Brad Edwards. "Computer matching of Medicare current beneficiary survey data with Medicare claims." In *Library of Congress Cataloging-in-Publication Data*, 191. 1996.

HealthIT.gov. "What is telehealth? How is telehealth different from telemedicine?" Last reviewed October 17, 2019. <https://www.healthit.gov/faq/what-telehealth-how-telehealth-different-telemedicine>.

APPENDICES

5. APPENDICES

APPENDIX A: MCBS COMMON DEFINITIONS

Baseline interview: The initial questionnaire administered in the fall round of the year the beneficiary is selected into the sample (interview #1).

Beneficiary: Beneficiary refers to a person receiving Medicare services who may or may not be participating in the MCBS.⁶ Beneficiary may also refer to an individual selected from the MCBS sample about whom the MCBS collects information.

Claim-only event: A claim-only event is a medical service or event known only through the presence of a Medicare FFS claim from administrative data. This means that the event represented in the data could not be reconciled with a corresponding survey-reported event.

Community component: Survey administered for beneficiaries living in the community (i.e., not in a long-term care facility such as a nursing home) during the reference period covered by the MCBS interview. An interview may be conducted with the beneficiary or a proxy.

Company clinic: A doctor's office or clinic, which is operated principally for the employees (and sometimes their dependents) of a particular company or business.

Continuing interview: The questionnaire administered as beneficiaries progress through the study (interviews #2-11).

Continuously enrolled (aka always enrolled): A Medicare beneficiary who was enrolled in Medicare from the first day of the calendar year until the fall interview and did not die prior to the fall round. This population excludes beneficiaries who dis-enrolled or died prior to their fall interview, residents of foreign countries, and residents of U.S. possessions and territories.

Core sections: These sections of the MCBS Questionnaire are of critical purpose and policy relevancy to the MCBS. They may be fielded every round or on a seasonal basis.

Coronavirus (COVID-19 or SARS-CoV-2): An illness caused by a new coronavirus that can spread person to person. Symptoms range from mild (or no symptoms) to severe illness.⁷ The virus has been named "severe acute respiratory syndrome coronavirus 2" (SARS-CoV-2) and the disease it causes has been named "coronavirus disease 2019" ("COVID-19").

COVID-19 Fall 2020 Community Supplement: A nationally representative, cross-sectional telephone survey of Medicare beneficiaries living in the community on topics pertaining to the COVID-19 pandemic that was administered from October 2020 through November 2020.

COVID-19 Fall 2020 Facility Supplement: A nationally representative, cross-sectional telephone survey of Medicare beneficiaries living in a facility on topics pertaining to the COVID-19 pandemic that was administered from October 2020 through December 2020.

⁶ <https://www.cms.gov/Medicare/Medicare-General-Information/MedicareGenInfo/index.html>

⁷ "How to Protect Yourself & Others," Centers for Disease Control and Prevention, Last modified February 25, 2022, <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html>.

COVID-19 Summer 2020 Community Supplement: A nationally representative, cross-sectional telephone survey of Medicare beneficiaries living in the community on topics pertaining to the COVID-19 pandemic that was administered from June through July 2020.

COVID-19 Winter 2021 Community Supplement: A nationally representative, cross-sectional telephone survey of Medicare beneficiaries living in the community on topics pertaining to the COVID-19 pandemic that was administered from March to April 2021.

Crossover: A respondent who enters a long-term care facility setting (e.g., nursing homes) or who alternates between a community and a facility setting.

Current-year enrollee: Beneficiaries who were eligible and enrolled in Medicare (Parts A or B) anytime from January 1 to December 31 of the year the sample was selected.

Doctor: This includes both medical doctors (M.D.) and doctors of osteopathy (D.O.). It does not include chiropractors, nurses, technicians, optometrists, podiatrists, physician's assistants, physical therapists, psychologists, mental health counselors, or social workers. Generic specialties shown in parentheses following one of the specialties were coded as the specialty. For example, if the respondent mentioned a "heart" doctor, cardiology was coded. Generic answers not listed were not converted to specialties.

Doctor's office or group practice: This refers to an office maintained by a doctor or a group of doctors practicing together; generally, the patient makes an appointment to see a particular physician.

Ever enrolled: A Medicare beneficiary who was enrolled at any time during the calendar year including people who dis-enrolled or died prior to their fall interview. Excluded from this population are residents of foreign countries and of U.S. possessions and territories.

Exit interview: Conducted in the winter round, this final interview completes the respondent's participation in the MCBS (interview #11) and captures any unreported utilization and cost information from the prior year.

Facility component: Survey administered for beneficiaries living in facilities, such as long-term care nursing homes or other institutions, during the reference period covered by the MCBS interview. Interviewers conduct the Facility component with staff members located at the facility (i.e., Facility respondents); beneficiaries are not interviewed if they reside at a facility.

Fee-for-Service (FFS) payment: FFS is a method of paying for medical services in which each service delivered by a provider bears a charge. This charge is paid by the patient receiving the service or by an insurer on behalf of the patient.

Field interviewer: The principal contact for collecting and securing respondent data.

Field manager: A supervisor who motivates and manages a group of field interviewers to meet the goals of high-quality data collection on time and within budget limits.

Free-standing surgical center: A facility performing minor surgical procedures on an outpatient basis, and not physically connected to a hospital. Note that a unit performing outpatient procedures connected with a hospital (either physically or by name) is referred to as a hospital outpatient department/clinic.

Gap days: Gap days are periods during the calendar year in which a sample person was enrolled in Medicare but was not covered by a survey interview.

Home: This includes situations where the doctor comes to the beneficiary, rather than the beneficiary going to the doctor. Here, "home" refers to anywhere the beneficiary was usually staying at the time of the medical provider's visit. It may be the beneficiary's home, the home of a friend, a hotel room, etc.

Hospital emergency room: This means the emergency room of a hospital. "Urgent care" centers are not included. (NOTE: All hospital emergency room visits were included, even if the beneficiary went there for a "non-emergency" condition such as a cold, flu, or intestinal disorder.) A physician, nurse, paramedic, physician extender, or other medical provider may administer the health care.

Hospital outpatient department: A unit of a hospital, or a facility connected with a hospital, providing health and medical services, health education, health maintenance, preventive services, diagnosis, treatment, surgery, and rehabilitation to individuals who receive services from the hospital but do not require hospitalization or institutionalization. Outpatient clinics can include obesity clinics; eye, ear, nose and throat clinics; alcohol and drug abuse clinics; physical therapy clinics; kidney dialysis clinics; and radiation therapy clinics. The outpatient department may or may not be physically attached to a hospital, but it must be associated with a hospital.

Incoming Panel Sample (formerly known as Supplemental Panel): A statistically sampled group of beneficiaries that enter the MCBS in the fall of a data collection year. One panel is retired at the conclusion of each winter round, and a new panel is selected to replace it each fall round. Panels are identified by the data collection year (e.g., 2015 Panel) in which they were selected.

Long-term care facility: A facility that provides rehabilitative, restorative, and/or ongoing skilled nursing care to patients or residents in need of assistance with activities of daily living.

Medical clinic: This refers to any group of doctors or other health professionals who have organized their practice in a clinic setting and work cooperatively; generally, patients either come in without an appointment or make an appointment and see whatever health professional is available.

Medicare: Medicare is the federal health insurance program for people who are 65 and over, certain younger people with disabilities, and people with End-Stage Renal Disease (permanent kidney failure requiring dialysis or a transplant, sometimes called ESRD). The different parts of Medicare help cover specific services:

- Hospital Insurance (Part A): covers inpatient hospital stays, care in a skilled nursing facility, hospice care, and some home health care.
- Medical Insurance (Part B): covers certain doctors' services, outpatient care, medical supplies, and preventive services.
- Medicare Advantage (Part C): an alternative to coverage under traditional Medicare (Parts A and B), a health plan option similar to a Health Maintenance Organization (HMO) or Preferred Provider Organization (PPO) administered by private companies.
- Prescription Drug Coverage (Part D): additional, optional coverage for prescription drugs administered by private companies.

For more information, please visit the Medicare.gov website at <https://www.medicare.gov/sign-up-change-plans/decide-how-to-get-medicare/whats-medicare/what-is-medicare.html>.

Medicare Advantage (MA): Medicare Advantage Plans, sometimes called "Part C" or "MA Plans," are offered by private companies approved by Medicare. An MA provides, or arranges for the provision of, a comprehensive package of health care services to enrolled persons for a fixed capitation payment. The term "Medicare Advantage" includes all types of MAs that contract with Medicare, encompassing risk MAs, cost MAs, and health care prepayment plans (HCPs).

Medicare beneficiary: See Beneficiary.

Medicare Managed Care Organization (MCO)/Health Maintenance Organization (HMO): This is an organization that provides a full range of health care coverage in exchange for a fixed fee/co-pay. Some managed care plans require that plan members receive all medical services from one central building or location. Formerly referenced only as HMOs, these organizations are now referred to with terms such as Medicare MCOs/HMOs/MA/Part C.

Neighborhood/family health center: A non-hospital facility which provides diagnostic and treatment services, frequently maintained by government agencies or private organizations.

Other clinic: A non-hospital facility clinic that is not already listed in the other clinic categories. Some examples include a "free" clinic, a family planning clinic, or military base clinic.

Panel: See Incoming Panel sample.

Personal health care expenditures: Personal health care expenditures consist of health care goods and services purchased directly by individuals. They exclude public program administration costs, the net cost of private health insurance, research by nonprofit groups and government entities, and the value of new construction put in place for hospitals and nursing homes.

Prescription drugs: The basic unit measuring use of prescription drugs is a single purchase of a single drug in a single container. Prescription drug data are included for beneficiaries living in the community and in a facility; Prescription drugs administered during an inpatient hospital stay are not included.

Proxy: Beneficiaries who were too ill, or who could not complete the Community interview for other reasons, were asked to designate a proxy, someone very knowledgeable about the beneficiary's health and living habits. In most cases, the proxy was a close relative such as the spouse or a son or daughter. In a few cases, the proxy was a non-relative like a close friend or caregiver. In addition, a proxy was utilized if a beneficiary had been reported as deceased during the current round's reference period or if a beneficiary who was residing in the community in the previous round had since entered into a long-term care facility. Proxy interviews are only used for the Community interview, as the Facility interview is conducted with a staff member located at the facility (see definition of "Facility component").

Race/ethnicity: Hispanic origin and race are two separate and distinct categories. Persons of Hispanic origin may be of any race or combination of races. Hispanic origin includes persons of Mexican, Puerto Rican, Cuban, Central and South American, or Spanish origin. For the MCBS, responses to beneficiary race and ethnicity questions are reported by the respondent. More than one race may be reported. For conciseness, the text, tables, and figures in this document use shorter versions of the terms for race and Hispanic or Latino origin specified in the Office of Management and Budget 1997 Standards for Data on Race and Ethnicity. Beneficiaries reported as White and not of Hispanic origin were coded as White non-Hispanic; beneficiaries reported as Black/African-American and not of Hispanic origin were coded as Black non-Hispanic; beneficiaries reported as Hispanic, Latino/Latina, or of Spanish origin, regardless of their race, were coded as Hispanic. The "Other" race category includes other single races not of Hispanic origin (including American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander), Two or More Races, or Unknown Races.

Reference Period: The timeframe to which a questionnaire item refers.

Residence status: Medicare beneficiaries who only completed Community interviews during the calendar year are categorized as living only in the community. Medicare beneficiaries for whom only Facility interviews were completed during the calendar year are categorized as living only in facilities. Beneficiaries who

completed at least one Community interview and for whom at least one Facility interview was conducted during the year are classified as living in both community and facility

Respondent: Respondent refers to a person who answers questions for the MCBS; this person can be the beneficiary, a proxy, or a staff member located at a facility where the beneficiary resides (i.e., the Facility respondent).

Round: The MCBS data collection period. There are three distinct rounds each year; winter (January through April); summer (May through August); and fall (September through December).

Rural health clinic: A clinic that provides outpatient services, routine diagnostic services for individuals residing in an area that is not urbanized and is designated as a health staff shortage area or an area with a shortage of personal health services. The clinic can also provide outpatient services that include physician services, services and supplies provided under the direction and guidance of a physician by nurse practitioner, physician assistants, and treatment of emergency cases. These services are usually provided at no charge except for the amount of any deductible or coinsurance amount.

Sample person: An individual beneficiary selected from MCBS' Incoming Panel sample to participate in the MCBS survey.

Skilled nursing facility (SNF): A facility (which meets specific regulatory certification requirements) which primarily provides inpatient skilled nursing care and related services to patients who require medical, nursing, or rehabilitative services but does not provide the level of care or treatment available in a hospital. (Source: <https://www.cms.gov/apps/glossary/default.asp?Letter=S&Language=English>)

Survey-reported event: A survey-reported event is a medical service or event reported by a respondent during an interview. The event may have been matched to a Medicare FFS claim from administrative data, or it may be a survey-only event, in which case it was not matched to a Medicare claim and is only known through the survey.

Telehealth: The use of electronic information and telecommunications technologies to support long-distance clinical health care, patient and professional health-related education, public health and health administration. Technologies include videoconferencing, the internet, store-and-forward imaging, streaming media, and terrestrial and wireless communications. Telehealth is different from telemedicine because it refers to a broader scope of remote healthcare services than telemedicine. While telemedicine refers specifically to remote clinical services, telehealth can refer to remote non-clinical services, such as provider training, administrative meetings, and continuing medical education, in addition to clinical services.⁸

Telemedicine: The use of remote clinical services, such as videoconferencing for consultations with health professionals.⁹

Topical sections: Sections of the MCBS Questionnaire that collect information on special interest topics. They may be fielded every round or on a seasonal basis. Specific topics may include housing characteristics, drug coverage, and knowledge about Medicare.

⁸ "What is telehealth? How is telehealth different from telemedicine?" HealthIT.gov, last reviewed October 17, 2019, <https://www.healthit.gov/faq/what-telehealth-how-telehealth-different-telemedicine>.

⁹ "Telehealth Interventions to Improve Chronic Disease," Centers for Disease Control and Prevention, last modified May 11, 2020, <https://www.cdc.gov/dhds/pubs/telehealth.htm>.

Walk-in urgent center: A facility not affiliated with a nearby hospital, offering services for acute conditions (e.g., flu, virus, sprain). Typically, people are seen without appointments (i.e., walk-ins).

APPENDIX B: MCBS ROUNDS BY DATA YEAR AND SEASON

	Winter	Summer	Fall
1991	n/a	n/a	1
1992	2	3	4
1993	5	6	7
1994	8	9	10
1995	11	12	13
1996	14	15	16
1997	17	18	19
1998	20	21	22
1999	23	24	25
2000	26	27	28
2001	29	30	31
2002	32	33	34
2003	35	36	37
2004	38	39	40
2005	41	42	43
2006	44	45	46
2007	47	48	49
2008	50	51	52
2009	53	54	55
2010	56	57	58
2011	59	60	61
2012	62	63	64
2013	65	66	67
2014	68	69	70
2015	71/72	71/72	73
2016	74	75	76
2017	77	78	79
2018	80	81	82
2019	83	84	85
2020	86	87	88
2021	89	90	91
2022	92	93	94

APPENDIX C: TECHNICAL APPENDICES

C.1 Using the Data¹⁰

Example 1: Ever Enrolled Population

Below is the SAS code needed to create an analytic file that includes the Survey File and Cost Supplement variables. This example uses the Cost Supplement ever enrolled general purpose weights required to analyze out-of-pocket spending for Medicare beneficiaries by age. Because the analysis involves annual out-of-pocket costs, the adjusted total out-of-pocket spending variable (PAMTOOP) in the Person Summary (PS) segment is used. PAMTOOP summarizes out-of-pocket spending for 10 types of events (dental services, hearing services, hospice, inpatient, institutional, medical provider, outpatient, prescribed medication, vision services, and facility) for each beneficiary.

To produce estimates that are representative of the ever enrolled population, data users need to merge the Person Summary segment from the Cost Supplement File with the Demographics (DEMO) segment from the Survey File and limit the observations to respondents listed in the weights segment (CSEVWGTS). Inclusion of "(in=a)" and "if a then output" in the SAS code statement ensures that all observations in the weights segment are preserved, regardless of whether or not the other segments include the entire sample, and that the resulting dataset is restricted to the observations included in the weights file.

```
data merged_costfile;
    merge      costYY.CSEVWGTS (in = a)
              surveyYY.DEMO (keep = BASEID D_STRAT)
              costYY.PS (keep = BASEID PAMTOOP);
    by BASEID;
    if a then output;
run;
```

This merged file now includes the relevant Survey File and Cost Supplement variables that will allow for an examination of out-of-pocket spending for Medicare beneficiaries by age.

The code below transforms the age variable (D_STRAT) from a seven-category variable into a four-category variable. Exhibit C.1.1 presents information regarding recoding this age variable.

```
data mcbs_analyticfile;
    set merged_costfile;

    /*AGE*/
    if D_STRAT in (1,2) then AGECAT = 0;
    else if D_STRAT in (3,4) then AGECAT = 1;
    else if D_STRAT in (5,6) then AGECAT = 2;
    else if D_STRAT = 7 then AGECAT = 3;

run;
```

¹⁰ The "YY" in "costYY" and "surveyYY" refers to the data year of the Cost Supplement File and Survey File, respectively. Longitudinal code is represented with the convention of Y1, Y2, etc.

Exhibit C.1.1: Example 1 Recoding Variables

Measure	Original Variable	Recoded Variable
Age	D_STRAT 1: 0-44 2: 45-64 3: 65-69 4: 70-74 5: 75-79 6: 80-84 7: 85 +	AGECAT 0: 0-64 1: 65-74 2: 75-84 3: 85+

The sample code below demonstrates how to use the constructed analytic dataset to estimate out-of-pocket spending for all Medicare beneficiaries stratified by age. Although the MCBS includes variables to obtain weighted estimates and estimated standard errors using Taylor-series linearization approach, the balanced repeated replication (BRR or Fay's method) method provides more analytic flexibility when analyzing subgroups. Therefore, the examples presented in this section use BRR (Fay's method) variance estimation method.

* Estimate Total Out-of-Pocket Spending for Medicare Beneficiaries (using balanced repeated replication (Fay's method));

```
proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
    var PAMTOOP;
    weight CSEVRWGT;
    repweights CSEVR001-CSEVR100;
    domain AGECAT;
run;
```

Example 2: Event-Level Data

Additionally, a researcher might be interested in using event-level data, for example to estimate spending on cardiovascular drugs costs for Medicare beneficiaries in the given year. In this case, data users need separate records for each prescription drug event a beneficiary had in a given year, which allows for identifying beneficiaries who were prescribed cardiovascular drugs and total spending during the year. This example uses an Event Summary (ES) segment (i.e., the Prescribed Medicine Event segment (PME)), which allows for analysis of each prescription drug event for a given beneficiary.

To identify Medicare beneficiaries who were prescribed cardiovascular drugs during the year, data users need to limit the Prescribed Medicine Events (PME) file to prescription drug events related to drugs with a cardiovascular therapeutic classification (i.e., for which variable THERCC=41). Next, data users need to calculate total spending on cardiovascular drugs during the year for each Medicare beneficiary.

```
proc sql;
create table drug_spending_bybene AS
    select unique BASEID, sum(AMTTOT) as drug_spending
    from costYY.PME
    where thercc = "41"
    group by BASEID;
quit;
```

To produce estimates that are representative of the ever enrolled population, data users need to merge the weights segment (CSEVWGTS) and limit the dataset to the population included in the `drug_spending_bybene` file. Inclusion of "(in=a)" and "if a then output" in the SAS code statement ensures that observations that are in both the weights segment and the analytic file are preserved.

```
data mcbs_analyticfile;
    merge          costYY.CSEVWGTS
                drug_spending_bybene (in = a);
    by BASEID;
    if a then output;
run;
```

The sample code below demonstrates how to use the constructed analytic dataset to estimate average spending on cardiovascular drugs during the year for Medicare beneficiaries who had at least one prescription for a cardiovascular drug during the year. Balanced repeated replication (BRR or Fay's method) method variance estimation method is used in this example.

* Estimate Total Spending on Cardiovascular drugs per user for Medicare Beneficiaries (using balanced repeated replication (Fay's method));

```
proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
    var DRUG_SPENDING;
    weight CSEVRWGT;
    repweights CSEVR001-CSEVR100;
run;
```

Example 3: Service Summary Data

Finally, a data user might want to learn about what services beneficiaries use in a given year. This example shows the extent to which the use of outpatient hospital services varies by age. The analysis utilizes the Service Summary (SS) segment, which allows researchers to identify the total number of outpatient hospital events each beneficiary had in a given year. Data users need to subset the SS segment by event type. The sample code below demonstrates this by creating an analytic file that only includes service summary records of outpatient events.

To produce estimates that are representative of the ever enrolled population, data users need to merge the `op_summary` analytic file with the Demographics (DEMO) segment, and limit the observations to respondents listed in the weights segment (CSEVWGTS). Inclusion of "(in=a)" and "if a then output" in the SAS code statement ensures that all observations in the weights segment are preserved, regardless of whether or not the other segments include the entire sample, and that your dataset is restricted to the observations included in the weights file.

```
data op_summary;
    set costYY.SS (keep = BASEID AEVENTS EVNTTYPE);
    if EVNTTYPE = "OP" then output;
run;

data merged_costfile;
    merge          costYY.CSEVWGTS (in = a)
                surveyYY.DEMO (keep = BASEID D_STRAT)
                op_summary;
    by BASEID;
```

```

        if a then output;
run;

```

The merged file now includes the relevant Survey File and Cost Supplement variables that will allow for an examination of the number of outpatient episodes for Medicare beneficiaries by age.

The sample code below transforms the age variable (D_STRAT) into a four-category variable. Exhibit C.1.1 presents information regarding recoding this age variable.

```

data mcbs_analyticfile;
    set merged_costfile;

    /*AGE*/
    if D_STRAT in (1,2) then AGECAT = 0;
    else if D_STRAT in (3,4) then AGECAT = 1;
    else if D_STRAT in (5,6) then AGECAT = 2;
    else if D_STRAT = 7 then AGECAT = 3;

run;

```

The sample code below demonstrates how to use the constructed analytic dataset to estimate the average number of outpatient episodes for Medicare beneficiaries by age. Balanced repeated replication (BRR or Fay's method) method variance estimation method is used in this example.

* Estimate Number of Outpatient Episodes for Medicare Beneficiaries by Age (using balanced repeated replication (Fay's method));

```

proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
    var AEVENTS;
    weight CSEVRWGT;
    repweights CSEVR001-CSEVR100;
    domain AGECAT;

run;

```

Example 4: Combining Survey File and Cost Supplement File data

The next set of examples uses a hypothetical analysis plan to examine out-of-pocket costs for Medicare beneficiaries with diabetes whose usual source of care is a hospital or emergency room. This analysis requires merging the Survey File to the Cost Supplement File. As in Example 1 above, the analysis involves annual out-of-pocket costs. Thus, the adjusted total out-of-pocket spending variable (PAMTOOP) in the Person Summary (PS) segment is used, which summarizes spending for 10 types of events (dental services, hearing services, hospice, inpatient, institutional, medical provider, outpatient, prescribed medication, and vision services and facility) for each beneficiary, including by source of payment.

To generate estimates using the data from one of the Topical Questionnaire sections (including Usual Source of Care [USCARE]) on their own or merged with another Survey File segment that does not contain special non-response adjustment weights, the researcher must always use the special non-response adjustment general and replicate weights included in the Topical segment INSTEAD of using the general and replicate weights that appear in the separate weight segments (CENWGTS, EVRWGTS, CSEVWGTS). In this case, since we are merging the Person Summary segment from the Cost Supplement File, we should use the USCARE Cost Supplement ever enrolled weights (USCEWT).

```

data merged_surveycostfile;
  merge
    surveyYY.DEMO (keep = BASEID INT_TYPE D_STRAT)
    surveyYY.CHRNCOND (keep = BASEID D_OCDTYP)
    surveyYY.USCARE (in = a keep = BASEID PLACEPAR PLACEKND USCEWT USCE1-
    USCE100)
    costYY.PS (keep = BASEID PAMTOOP);
  by BASEID;
  if a then output;
run;

```

This merged file now includes the relevant Survey File and Cost Supplement variables that will allow for an examination of out-of-pocket costs for Medicare beneficiaries living in the community with diabetes.

There are numerous ways to convert these raw variables into analytic variables. The approach below creates variables for age (AGECAT), residence status (COMMONLY), the presence of Type 1 or Type 2 diabetes (DIABETES), and an indicator for the beneficiary's usual source of care (US_SOC). Exhibit C.1.2 presents information regarding the recoding and creation of these analytic variables.

```

data mcbs_analyticfile;
  set merged_surveycostfile;
  keep baseid commonly diabetes us_soc uscewt usce: pamtoop;
  /* AGE */
  If D_STRAT in (1,2) then AGECAT = 0;
    else if D_STRAT in (3,4) then AGECAT = 1;
    else if D_STRAT in (5,6) then AGECAT = 2;
    else if D_STRAT = 7 then AGECAT = 3;
  /* RESIDENCE STATUS */
  if INT_TYPE='C' then commonly=1;
    else commonly=0;
  /* DIABETES */
  if D_OCDTYP in (1,2) then diabetes=1;
  /* indicator variable for Type 1 or Type 2 diabetes */
  else diabetes=0;
  /* USUAL SOURCE OF CARE */
  US_SOC = 999; /* MISSING */
  if PLACEPAR = 2 then US_SOC = 0; /* NONE */
    else if PLACEPAR = 1 then do;
  if PLACEKND = 1 then US_SOC = 1; /* DOCTOR'S OFFICE */
    else if PLACEKND = 2 then US_SOC = 2; /* MEDICAL CLINIC */
    else if PLACEKND IN (11,12) then US_SOC = 3; /* HOSPITAL/OPD/ER */
    else if PLACEKND IN (3,4,6,7,8,9,10,13,14,91) then US_SOC = 4;
  /* OTHER */
  end;
run;

```

Exhibit C.1.2: Example 4 Recoding Variables

Measure	Original Variable	Recoded Variable
Age	D_STRAT 1: 0-44 2: 45-64 3: 65-69 4: 70-74 5: 75-79 6: 80-84 7: 85 +	AGECAT 0: 0-64 1: 65-74 2: 75-84 3: 85+
Residence Status	INT_TYPE B: Both C: Community F: Facility	COMMONLY 1: Community 0: Not community
Diabetes	D_OCDTYP 1: Type 1 2: Type 2 3: Pre-diabetes or borderline 4: Gestational (pregnancy-related) 91: Other type of diabetes	DIABETES 1: Type 1 or Type 2 0: Not Type 1 or Type 2
Usual Source of Care	PLACEKND 1: Doctor's office or group practice 2: Medical clinic 3: Managed care plan center/HMO 4: Neighborhood or family health center 6: Rural Health Clinic 7: Company clinic 8: Other clinic 9: Walk-in urgent care center 10: At home 11: Hospital emergency room 12: Hospital outpatient department 13: Veterans' Administration facility 14: Mental health center 91: Other, specify <i>Note: Applies only if PLACEPAR = 1</i>	US_SOC 1: Doctor's office 2: Medical Clinic 3: Hospital/OPD/ER 4: Other

Example 5: Variance Estimation

The sample code below provides examples of how to produce correct estimates for continuous and categorical variables and for the entire sample as well as for a subgroup. The examples use the file created in Example 4 above (i.e., mcbs_analyticfile) to estimate out-of-pocket spending for all Medicare beneficiaries (Example 5.1), out-of-pocket spending for Medicare beneficiaries living in the community whose usual source of care is a hospital or emergency room (Example 5.2), usual source of care for all Medicare beneficiaries (Example 5.3), and usual source of care for Medicare beneficiaries living in the community with diabetes (Example 5.4).

Although the MCBS includes variables to obtain weighted estimates and estimated standard errors using Taylor-series linearization approach, the balanced repeated replication (BRR or Fay's method) method provides more analytic flexibility when analyzing subgroups.¹¹ Therefore, the examples presented in this section pertaining to subgroup analysis (Examples 5.2 and 5.4) use BRR (Fay's method) variance estimation method.

The examples below provide sample code for SAS, Stata®, and R® to generate estimates, both for the total population and for subgroups.

Example 5.1. Total Out-of-Pocket Spending for Medicare Beneficiaries

SAS

```
data mcbs_analyticfile;
  merge      costYY.CSEVWGTS (in = a)
            costYY.PS (keep = BASEID PAMTOOP);
  by BASEID;
  if a then output;
run;
```

* Estimate Total Out-of-Pocket Spending for Medicare Beneficiaries (using balanced repeated replication (Fay's method));

```
proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
  var PAMTOOP;
  weight CSEVRWGT;
  repweights CSEVR001-CSEVR100;
run;
```

* Estimate Total Out-of-Pocket Spending for Medicare Beneficiaries (using Taylor-series Linearization approach);

```
proc surveymeans data= mcbs_analyticfile varmethod = TAYLOR;
  var PAMTOOP;
  weight CSEVRWGT;
  cluster SUDUNIT;
  strata SUDSTRAT;
run;
```

¹¹ For example, performing analysis on small subgroups could lead to instances where one or more strata contain a single observation. In such situations, calculating intra-strata variances using Taylor-series Linearization approach is not possible. Therefore, standard errors cannot be computed.

Stata¹²

```

* declare survey dataset (using balanced repeated replication (fay's method));
svyset [pweight= CSEVRWGT], brrweight(CSEVR001-CSEVR100) fay(.3) vce(brr)

* estimate total out-of-pocket spending for Medicare beneficiaries;
svy brr, fay(.3): mean PAMTOOP

* declare survey dataset (using taylor-series linearization approach);
svyset [pweight= CSEVRWGT], strata(SUDSTRAT) psu(SUDUNIT)

* estimate total out-of-pocket spending for Medicare beneficiaries;
svy: mean PAMTOOP

```

R¹³

```

# specify survey design object (using balanced repeated replication (fay's method))
mcbs <- svrepdesign(
  weights = ~CSEVRWGT,
  repweights = "CSEVR[001-100]+",
  type = "Fay",
  rho = 0.3,
  data = mcbs_analyticfile,
  combined.weights = TRUE
)

# specify survey design object (using taylor-series linearization approach)
mcbs <- svydesign(
  weights = ~CSEVRWGT,
  id = ~BASEID,
  strata = ~SUDSTRAT,
  nest= TRUE,
  data = mcbs_analyticfile
)

# estimate total out-of-pocket spending for Medicare beneficiaries;
svymean(~PAMTOOP, design=mcbs)

```

Example 5.2. Total Out-of-Pocket Spending for Medicare Beneficiaries Living in the Community Whose Usual Source of Care is Hospital or Emergency Room

SAS

```

data merged_surveycostfile;
  merge
    surveyYY.DEMO (keep = BASEID INT_TYPE)
    surveyYY.USCARE (in = a keep = BASEID PLACEPAR PLACEKND USCEWT USCE1-
    USCE100)

```

¹² Note that the mcbs_analyticfile for Stata and R code must include the appropriate weights and the stratum variables.

¹³ The survey package in R is required to run this R code.

```

        costYY.PS (keep = BASEID PAMTOOP);
    by BASEID;
    if a then output;
run;

data mcbs_analyticfile;
    set merged_surveycostfile;
    keep baseid commonly us_soc uscewt usce: pamtoop;
    /* RESIDENCE STATUS */
    if INT_TYPE='C' then commonly=1;
    else commonly=0;
    /* USUAL SOURCE OF CARE */
    US_SOC = 999; /* MISSING */
    if PLACEPAR = 2 then US_SOC = 0; /* NONE */
    else if PLACEPAR = 1 then do;
    if PLACEKND = 1 then US_SOC = 1; /* DOCTOR'S OFFICE */
    else if PLACEKND = 2 then US_SOC = 2; /* MEDICAL CLINIC */
    else if PLACEKND IN (11,12) then US_SOC = 3; /* HOSPITAL/OPD/ER */
    else if PLACEKND IN (3,4,6,7,8,9,10,13,14,91) then US_SOC = 4;
    /* OTHER */
    end;
run;

```

* Total Out-of-Pocket Spending for Medicare Beneficiaries Living in the Community Whose Usual Source of Care is Hospital or Emergency Room (using balanced repeated replication (Fay's method));

```

proc surveymeans data= mcbs_analyticfile varmethod = brr (fay=.30);
    var PAMTOOP;
    weight USCEWT;
    repweights USCE1-USCE100;
    domain COMMONLY * US_SOC;
run;

```

Stata

```

* declare survey dataset (using balanced repeated replication (fay's method))
svyset [pweight= USCEWT], brrweight(USCE1-USCE100) fay(.3) vce(brr)

```

```

* total out-of-pocket spending for Medicare beneficiaries living in the community whose usual source of
care is hospital or emergency room
svy brr, fay(.3) subpop(if commonly==1 & US_SOC==3) : mean PAMTOOP

```

R

```

# specify survey design object (using balanced repeated replication (fay's method))
mcbs_ussoc <- svrepdesign(
    weights = ~USCEWT,
    repweights = "USCE[1-100]+",
    type = "Fay",
    rho = 0.3,
    data = mcbs_analyticfile,

```

```

        combined.weights = TRUE
    )

```

#subset survey design object to Medicare beneficiaries living in the community whose usual source of care is hospital or emergency room

```
mcbs_ussoc <- subset(mcbs, commonly==1 & US_SOC==3)
```

total out-of-pocket spending for Medicare beneficiaries living in the community whose usual source of care is hospital or emergency room

```
svymean(~PAMTOOP, design=mcbs_ussoc)
```

Example 5.3. Number of Medicare Beneficiaries by Usual Source of Care

SAS

```

data merged_surveycostfile;
    merge
        surveyYY.USCARE (keep = BASEID PLACEPAR PLACEKND USCEWT USCE1-USCE100)
        surveyYY.EVRWGTS (in = a keep = BASEID SUDSTRAT SUDUNIT)
        costYY.PS (keep = BASEID PAMTOOP);
    by BASEID;
    if a then output;
run;

```

```

data mcbs_analyticfile;
    set merged_surveycostfile;
    keep baseid us_soc pamtoop uscewt usce: sudstrat sudunit;
    /* USUAL SOURCE OF CARE */
    US_SOC = 999; /* MISSING */
    if PLACEPAR = 2 then US_SOC = 0; /* NONE */
    else if PLACEPAR = 1 then do;
    if PLACEKND = 1 then US_SOC = 1; /* DOCTOR'S OFFICE */
    else if PLACEKND = 2 then US_SOC = 2; /* MEDICAL CLINIC */
    else if PLACEKND IN (11,12) then US_SOC = 3; /* HOSPITAL/OPD/ER */
    else if PLACEKND IN (3,4,6,7,8,9,10,13,14,91) then US_SOC = 4;
    /* OTHER */
    end;
run;

```

* Number of Medicare Beneficiaries by Usual Source of Care (using balanced repeated replication (Fay's method));

```

proc surveyfreq data= mcbs_analyticfile varmethod = brr (fay=.30);
    table US_SOC;
    weight USCEWT;
    repweights USCE1-USCE100;
run;

```

* Number of Medicare Beneficiaries by Usual Source of Care (using Taylor-series Linearization approach);

```
proc surveyfreq data= mcbs_analyticfile varmethod = TAYLOR;
```

```

table US_SOC;
weight USCEWT;
cluster SUDUNIT;
strata SUDSTRAT;

run;

```

Stata

```

* declare survey dataset (using balanced repeated replication (fay's method))
svyset [pweight= USCEWT], brrweight(USCE1-USCE100) fay(.3) vce(brr)

* number of Medicare beneficiaries by usual source of care
svy brr, fay(.3) : tabulate US_SOC, se

* declare survey dataset (using taylor-series linearization approach)
svyset [pweight= USCEWT], strata(SUDSTRAT) psu(SUDUNIT)

* number of Medicare beneficiaries by usual source of care
svy : tabulate US_SOC, se

```

R

```

# specify survey design object
mcbs <- svrepdesign(
  weights = ~USCEWT,
  repweights = "USCE[1-100]+",
  type = "Fay",
  rho = 0.3,
  data = mcbs_analyticfile,
  combined.weights = TRUE
)

# specify survey design object (using taylor-series linearization approach)
mcbs <- svydesign(
  weights = ~USCEWT,
  id = ~SUDUNIT,
  strata = ~SUDSTRAT,
  nest = TRUE,
  data = mcbs_analyticfile
)

# number of Medicare beneficiaries by usual source of care
svytable(~US_SOC, design=mcbs)

```

Example 5.4. Number of Medicare Beneficiaries Living in the Community with Diabetes by Usual Source of Care

SAS

```

data merged_surveycostfile;
  merge
    surveyYY.DEMO (keep = BASEID INT_TYPE)
    surveyYY.CHRNCOND (keep = BASEID D_OCDTYP)

```

```

        surveyYY.USCARE (in = a keep = BASEID PLACEPAR PLACEKND USCEWT USCE1-USCE100)
        costYY.PS (keep = BASEID PAMTOOP);
    by BASEID;
    if a then output;
run;

data mcbs_analyticfile;
    set merged_surveycostfile;
    keep baseid commonly diabetes us_soc uscewt usce: pamtoop;
    /* RESIDENCE STATUS */
    if INT_TYPE='C' then commonly=1;
    else commonly=0;
    /* DIABETES */
    if D_OCDTYP in (1,2) then diabetes=1;
    /* indicator variable for Type 1 or Type 2 diabetes */
    else diabetes=0;
    /* USUAL SOURCE OF CARE */
    US_SOC = 999; /* MISSING */
    if PLACEPAR = 2 then US_SOC = 0; /* NONE */
    else if PLACEPAR = 1 then do;
    if PLACEKND = 1 then US_SOC = 1; /* DOCTOR'S OFFICE */
    else if PLACEKND = 2 then US_SOC = 2; /* MEDICAL CLINIC */
    else if PLACEKND IN (11,12) then US_SOC = 3; /* HOSPITAL/OPD/ER */
    else if PLACEKND IN (3,4,6,7,8,9,10,13,14,91) then US_SOC = 4;
    /* OTHER */
    end;
run;

* Number of Medicare Beneficiaries Living in the Community with Diabetes by Usual Source of Care
(using balanced repeated replication (Fay's method));

proc surveyfreq data= mcbs_analyticfile varmethod = brr (fay=.30);
    table DIABETES * COMMONLY * US_SOC/ row;
    weight USCEWT;
    repweights USCE1-USCE100;
run;

```

Stata

```

* declare survey dataset (using balanced repeated replication (fay's method))
svyset _n [pweight= USCEWT], brrweight(USCE1-USCE100) fay(.3) vce(brr)

* number of Medicare beneficiaries living in the community with diabetes by usual source of care
svy brr, fay(.3) subpop(if diabetes==1 & commonly==1) : tab US_SOC

```

R

```

# specify survey design object (using balanced repeated replication (fay's method))
mcbs <- svrepdesign(
    weights = ~USCEWT,
    repweights = "USCE[1-100]+",

```

```

        type = "Fay",
        rho = 0.3,
        data = mcbs_analyticfile,
        combined.weights = TRUE
    )

#subset survey design object to Medicare beneficiaries living in the community with diabetes by usual
source of care
mcbs_subgrp <- subset(mcbs, diabetes==1 & commonly==1)

# number of Medicare beneficiaries living in the community with diabetes by usual source of care
svytable(~US_SOC, design=mcbs_subgrp)

```

Example 6: Longitudinal Analysis

The sample code below demonstrates the steps involved in constructing an analytic dataset and performing longitudinal analysis. The example estimates percent change in out-of-pocket costs between CY1 and CY2 for Medicare beneficiaries enrolled in the Medicare program during both years. Although the MCBS includes variables to obtain weighted estimates and estimated standard errors using Taylor-series linearization approach, the balanced repeated replication (Fay's method) method provides more analytic flexibility when performing analysis using longitudinal data.¹⁴

CMS generally recommends the BRR method of variance estimation to MCBS users because it requires neither the specification of strata and cluster definitions nor the specification of domain or subgroup definitions in subpopulation analyses, which are required for Taylor-series estimation and are common inadvertent omissions. However, the Taylor series method of variance estimation is also appropriate for experienced users who prefer this method or in instances where the BRR method is not possible in the available software. For these reasons, the MCBS data files include the variables SUDSTRAT and SUDUNIT, which are needed for Taylor-series estimation. The SAS functions %surveyglm and %surveygenmod appropriately allow for strata and cluster definitions. When using these functions (and in any other instances where Taylor series estimation is used), specify SUDSTRAT as the strata definitions and SUDUNIT as the cluster definitions.

The examples presented in this section involve multiple years of MCBS data and use replicate weights – a form of the BRR technique.

```

/* Create Analytic Dataset for Longitudinal Analysis */
/* Merge Y1 Person Summary (PS) file with Y2 longitudinal weights (CSL2WGTS) file */
data mcbs_analytic_file;
    merge costY2.CSL2WGTS (in = a)
          costY1.PS (keep = BASEID PAMTOOP);
    by BASEID;
    rename PAMTOOP = PAMTOOPY1;
    if a;
run;

/* Merge Y2 Person Summary file with analytic file created in example [include here] */
data mcbs_analytic_file;
    merge mcbs_analytic_file (in = a)
          costY2.PS (keep = BASEID PAMTOOP rename=(PAMTOOP=PAMTOOPY2));

```

¹⁴ Given the rotating panel design of the MCBS, performing longitudinal analysis using Taylor-Series Linearization method of variance estimation will require additional adjustments to account for non-independence of beneficiaries across years in a multi-year dataset.

```

        by BASEID;

        PAMTOOP_PDIF = PAMTOOPY2 - PAMTOOPY1;

        if a;
run;

```

SAS

* Estimate percent change in out-of-pocket costs between CY1 and CY2 for beneficiaries enrolled in Medicare during both years (using balanced repeated replication (Fay's method));

```

proc surveymeans data=mcbs_analytic_file varmethod = brr (fay = .30);
    var PAMTOOP_PDIF;
    weight CSL2YWGT;
    repweights CSL2Y001-CSL2Y100;
run;

```

Stata

* Declare survey dataset (using balanced repeated replication (Fay's method))

```
svyset _n [pweight= CSL2YWGT], brrweight(CSL2Y001-CSL2Y100) fay(.3) vce(brr)
```

* Estimate percent change in out-of-pocket costs between CY1 and CY2 for beneficiaries enrolled in Medicare during both years

```
svy brr, fay(.3) : mean PAMTOOP_PDIF
```

R

Specify survey design object (using balanced repeated replication (Fay's method))

```

mcbs <- svrepdesign(
  weights = ~ CSL2YWGT,
  repweights = "CSL2Y[001-100] +",
  type = "Fay",
  rho = 0.3,
  data = mcbs_analytic_file,
  combined.weights = TRUE
)

```

```
svymean(~PAMTOOP_PDIF, design=mcbs)
```

C.2 Matching Survey and Administrative Data

The Cost Supplement File brings together survey information, which can only be obtained directly from a beneficiary or proxy respondent with reliable information on services used, and Medicare payments made from administrative bill files. Survey-reported cost data include information on the use and costs of health care services, including the amount paid by private health insurance and other payers/payer types (if applicable). The survey also collects information on health services not covered by Medicare, most notably, long-term facility care and dental, vision, and hearing services.

Medicare bill data include use and cost information on prescription drugs, inpatient hospitalizations, outpatient hospital care, physician services, home health services, durable medical equipment, skilled nursing home services, and hospice services. This combination Cost Supplement File can support a much broader range of research and policy analyses on the Medicare population than would be possible using either survey collected cost data or administrative bill data alone.

Use and costs of Medicare-covered services are reported on both the MCBS survey and in the CMS administrative bill files. This overlap in reporting from the two sources is used to verify the accuracy of survey reports of health service use. Survey reports are matched with administrative bill data to adjust for survey under-reporting using more complete administrative bill data, and to fill in and correct survey reported payment amounts with more accurate information from bills submitted to and paid by Medicare. Note that this can only be done for FFS (or Original Medicare) beneficiaries accessing services covered by Medicare such as inpatient hospital services, outpatient hospital services, physician services, acute skilled nursing facility services, durable medical equipment, and other covered services as well as Medicare Part D beneficiaries accessing prescription drug services. For health services not covered by Medicare FFS or Part D such as long-term facility care, there is no independent source to which survey reports could be matched.

Under-reporting of medical services is an enduring problem in personal interview surveys. While respondents can usually recall significant events like hospitalizations for several months, they often fail to recall more routine care like physician visits after a few weeks. In general, as the time interval between the interview date and the medical event increases, the probability decreases that the event will be recalled and reported accurately in the interview. The MCBS interviews sampled beneficiaries up to three times per year over a four year period, and the average interview recall period is about four months. (More frequent interviews would reduce the recall problem, but it would greatly increase both survey administration costs and the reporting burden on beneficiaries). Given normal rates of memory decay and the frequency with which aged and disabled persons use medical care, it was reasonable to assume that matching survey events to administrative bills would be helpful in identifying medical events that the respondent could not recall during the interviews.

C.2.1 File Building

In order to get a complete and accurate file of services used and payments made from the event-level segments, all 40,427 matched service records should be added to all unmatched 91,713 Medicare claim-only records. In addition, unmatched survey reports, excluding the 7,942 records with a Medicare payment amount and no MA payment amount, should be added to the matched and Medicare claim-only records. The records which are to be excluded are identified with the SOWMP flag on the IUE, OPE, MPE, and IPE segments. This file will be the most complete and accurate file possible, and this combination minimizes the risk of double counting unmatched records.

C.2.2 Filling the Gaps

The 2020 Cost Supplement File is designed to provide person-level data for estimating total use of, and total payments for, all health care services, covered and non-covered, received by Medicare beneficiaries during calendar year 2020.

This section describes the adjustments that are made to the MCBS data to attain the goal of estimating total use of, and total payments for all health care services (covered and non-covered), received by Medicare beneficiaries during calendar year 2020. The adjustments made are as follows:

- **EVENT-LEVEL MATCHING:** These operations identified services paid for by Medicare that were not reported on the survey and corrected Medicare payment data reported inaccurately on the survey. A discussion of match results and instructions for building a complete file and avoiding duplication is provided in Technical Appendices C.2.3.3-C.2.3.4.
- **MISSING PAYMENTS AND PAYERS:** These adjustments compensate for missing payment data and for when the respondent did not know how much an event cost and/or how the event was paid for (by whom, and how much by each payer).
- **PRESCRIPTION DRUGS:** Describes the particular problems encountered in creating the prescription drug event file. Also includes adjustments to compensate for missing payment data and for when the

respondent did not know how much a prescription cost or how the prescription was paid for (by whom, and how much by each payer).

- **ADJUSTMENTS FOR MISSING DAYS AND UNDATED USE:** These adjustments compensate for data that are missing because some periods of time were not covered by interviews and because some types of health services use (particularly prescription drugs) are undated.
- **ADJUSTMENTS FOR UNREPORTED HEALTH CLAIMS OF MA BENEFICIARIES:** These adjustments compensate for known under-reporting of health events for beneficiaries with MA coverage. For beneficiaries with FFS coverage, similar under-reported events were obtained from the administrative claims data records that were not matched to survey reported events.

Adjustments made to records in the Cost Supplement LDS are constrained in two ways. First, because CMS administrative data are used to fill in much of the missing information, all adjustments to MCBS utilization, cost, and source of payment data are consistent with CMS administrative data. For example, if CMS records indicate that the beneficiary is dually entitled to both Medicare and Medicaid, then Medicaid must be considered a possible source of payment when source of payment is missing, even if the respondent did not volunteer that information. Second, adjusted data must be consistent with other information for the same person. For example, the source of payment for individual events must be consistent with the beneficiary's health insurance information.

C.2.3 Event-Level Matching

There are two primary objectives in matching survey reports to Medicare administrative bill records: to correct for under-reporting of events on the survey, and to correct errors in payment information collected in the survey.

The first step in matching survey-reported medical events to Medicare bill records is gathering all events for a person together. Because the MCBS sample is drawn from CMS' Medicare Administrative enrollment data, matching the Medicare paid claims and bills with the correct beneficiary is a reasonably straightforward process. The beneficiary's Medicare number (i.e., health insurance claim number (HICN) or Medicare Beneficiary Identifier (MBI)) is part of the information collected from the enrollment data when the sample is drawn. The beneficiary's HICN/MBI is verified in the first MCBS interview. Prior to the match, Medicare paid claims are retrieved from the Medicare National Claims History (NCH) repository, by HICN/MBI. The search file includes all cross-reference numbers and different beneficiary identification codes associated with each beneficiary, ensuring that all bill records are recovered.

Linking and reconciling the retrieved Medicare claims with individual events reported in the survey is a much more complicated process than matching Medicare paid bills with the correct beneficiary. There is no data element, or combination of elements, that provide a consistent basis for matching survey data to Medicare claims across all types of services. There are significant differences in the ways in which medical goods and services are characterized in the survey and in the Medicare claims records.

Neither the MCBS nor CMS claims records capture a consistent set of data elements for all service types. For example, the MCBS does not capture total reimbursement for inpatient hospital services because the respondent is not likely to know that information; it is not typically included on the notice of utilization, and thus, this information cannot be used in matching. In other categories, especially Part B services, the total charge of the service is known because it appears on the explanation of benefits, and it is a key match field. Similarly, CMS claims data do not always have the same data elements for different claim types. The carrier control number for each claim is included in CMS' claims history files and the MCBS attempts to collect the carrier control number from the beneficiary's explanation of benefits in the interviews. As a result, this item is extremely useful in matching survey reported utilization to Part B claims. On the other hand, the intermediary control number (Intermediaries process claims for Part A of Medicare) is not available in CMS' files, so even

though it is collected in the survey, this data element is not helpful in matching the survey data to Part A bill records. Additional details regarding matching are provided in Technical Appendix C.2.3.4.

C.2.3.1 Survey-reported utilization

In the utilization sections of the MCBS Community Questionnaire, respondents are asked about all medical events, including visits to practitioners of all types, prescriptions, and any medical equipment or supplies used. (Please find copies of the survey instruments and exact wording of the questions at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Questionnaires>).

Types of utilization collected in the MCBS:

- DU: Dentist visits, including cleaning, x-rays and repair, purchase or repair of dentures, and orthodontic procedures
- VU: Vision care including vision exams, contact lens fittings or purchases, eye glass frame fittings or purchases, and different kinds of surgeries (e.g., cataract, corneal, etc.).
- HU: Hearing care including hearing exams, hearing aid fittings, repairs, or purchases, or hearing rehabilitative services
- ER: Hospital emergency room visits
- IP: Inpatient hospital stays
- IU: Other short-term institutional stays, such as skilled nursing home stays or rehabilitation hospital stays
- MP: Doctor visits, including visits with medical doctors (MD); practitioners, such as chiropractors, podiatrists, audiologists, and optometrists; mental health professionals such as psychiatrists, psychologists, and clinical social workers; therapists, such as physical therapists, speech therapists, occupational therapists, and intravenous and respiratory therapists; other medical practitioners, such as nurses and paramedics; and other places offering medical care, such as clinics, neighborhood health centers, infirmaries, and urgent care centers.
- OP: Outpatient visits, including visits to the outpatient department or outpatient clinic of a hospital.
- OM: Other medical expenses, including purchase or rental of a variety of items: eyeglasses or contact lenses and hearing aids; orthopedic items such as canes, walkers, wheelchairs, and corrective shoes; diabetic supplies; oxygen supplies and equipment; kidney dialysis equipment; hospital beds, commodes, and disposable supplies such as disposable diapers and bandages.
- PM: All prescription medications except those provided by the doctor or practitioner as samples and those provided in an inpatient setting.

In addition to these categories, the Community Questionnaire is also designed to collect some types of utilization that the respondent may unintentionally omit. This utilization is captured when the beneficiary's Medicare and private health insurance statements are reviewed and is classified as SD and SL. The SD and SL categories typically include such things as anesthesiology administered while the beneficiary was an inpatient, lab tests not done at the doctor's office, and the radiologist's interpretation of an x-ray.

The MCBS Facility Instrument captures similar information about people living in long-term care facilities, with the exception of prescribed medicines, which is not collected in the Facility Instrument. CMS administrative data for inpatient hospital stays and short-term skilled nursing home stays are integrated with survey data to ensure date alignment. These administrative data are also used to properly account for the Medicare payments related to the short-term skilled nursing home stays that occur within the encompassing stay at the long-term care facilities.

C.2.3.2 CMS-reported utilization

Medicare claims are organized by type of provider. The categories of Medicare claims records are as follows:

- Inpatient hospital, psychiatric hospital, Tuberculosis hospital, Christian Science facility, and skilled nursing facility bills: Although these records all share the same format, they contain codes that allow them to be separated into these subcategories. For purposes of the match, bills from skilled nursing facilities were separated from the other types of bills, but no further subdivisions were made.
- Hospice bills
- Outpatient hospital bills
- Part B physician/supplier claims for physician services, diagnostic laboratory and radiology, durable medical equipment, and some prescription medicines.
- Part D bills

C.2.3.3 Match categories

In matching the survey-reported utilization to the Medicare claims data, MCBS staff frequently must match a Medicare claim category to multiple MCBS categories, and vice versa. Although there are some clear relationships between the categories of utilization collected in the survey and CMS claims categories, not all categories match neatly.

Event-level matching is actually a series of matches between different categories of Medicare claims and MCBS service types. In conducting these matches MCBS staff employ different match algorithms, depending on the data elements available for the particular categories being matched. Matches are arranged in sequence, so that the most similar survey-reported and Medicare claims categories are compared first. Exhibit C.2.3.3 presents an overview of the categorical matches.

Exhibit C.2.3.3: Overview of event category matches conducted during event-level matching

Matches between similar service types	Matches between less similar service types
IP to Inpatient hospital bills	ER to Inpatient hospital bills
MP, OM, SD, SL to Part B physician/supplier	OP to Inpatient hospital bills
OP to Outpatient hospital bills	IU to Inpatient hospital bills
IU to SNF bills	IP to SNF bills
DU to Part B physician/supplier claims	IP to Outpatient hospital bills
ER to Outpatient hospital bills	OP to Part B physician/supplier claims
	MP, OM, SD, SL to Outpatient hospital bills

Each match algorithm employs a hierarchy of match criteria which are progressively less restrictive. For example, reported doctor visits are initially compared to claims records by physician's name, date of service, and total charge. If there is not an exact match, the algorithm checks for a match on physician's name and date of service, or total charge and date of service. If there is still no match, the program looks for an exact match on physician's name and total charge, with the date of service match relaxed to dates within one week of each other. The match algorithms not only link survey-reported utilization and Medicare claims records, but also code the records to indicate the strength of the link.

MCBS staff designed the match algorithms to allow survey-reported utilization to be linked to multiple Medicare claims, and vice versa, for two reasons. First, multiple links are often valid. For example, a survey-reported

doctor visit may be linked to both a Medicare claim for the physician's services and a Medicare claim for laboratory services connected with the visit. Second, a stronger match may occur later in the series of matches. A survey-reported doctor visit may have a weak link to a Medicare Part B physician/supplier claim and a strong link to a Medicare outpatient claim. MCBS' staff uses the link-strength indicator to resolve situations where the multiple matches are logically inconsistent.

Hospice bills are excluded from the match because there is no clean "hospice" category in the survey data. Facility utilization is matched in only a summary fashion to improve the accuracy of Medicare payment data.

Three outcomes are possible from the attempted match of survey data to Medicare claims data for events where claims are available (i.e., paid for by Medicare FFS or Part D): the information from the two sources agrees (a match); or, information reported in the survey is not present in the Medicare claims data; or, information is present in the Medicare claims data which was not reported in the survey.

- 2020 Cost Supplement File "events"

- ▶ The matching programs produce a set of records which reflect the best combination of survey and Medicare claims categories, and present records from both sources (matched and unmatched) in a uniform format. Since the categories of utilization in the Medicare claims do not match the survey categories, utilization groups in the 2020 Cost Supplement File are a combination of the two sources.
- ▶ The most disaggregate level of utilization records in the 2020 Cost Supplement File is the "event" level record. Event records combine survey-reported information and Medicare claims data in the nine categories presented: IPE, OPE, IUE, DUE, HUE, MPE, PME, VUE, and FAE. Event records contain a variable to indicate the source of information—Medicare claims data, survey data, or both. An additional variable provides the link from the event record to the bill data, if both sources provided the information.

- Emergency room

- ▶ The emergency room (ER) survey category is split between IPE and OPE. Under the prospective payment system, emergency room services that result directly in a hospital admission are included in the diagnosis-related group (DRG) payment for the inpatient stay, and thus are not associated with any separate charges or claims (see DRG variable on IPE segment). Emergency room visits that are not immediately followed by an inpatient admission are classified as outpatient services. For this reason, survey-reported emergency room (ER) utilization is matched to outpatient, then inpatient bill records, and is reflected in the 2020 Cost Supplement File as either OPE or IPE records. Several other survey categories (MP, SD, SL, and OM) have been combined to make up the single MPE category. Hospice services do not exist as a separate category of utilization in the survey data, so this category derives from the Medicare claims data.

- Post-match edits

- ▶ For most types of services, the MCBS collects a date of service to assist in matching survey-reported data to claims records. Respondents may not always recall exact dates, so dates are collected in three independent parts (i.e., month, day, and year).
- ▶ Since the year portion of a survey date may be missing or incorrect, records for services in 2019 and 2021 were not eliminated from the survey file until the match was concluded. Similarly, respondents may "telescope" events, believing them to have taken place recently when in reality they occurred a year or more in the past. As matching Medicare claims might help to identify and eliminate these responses, the Medicare records were also not edited on date until after the match; for the match records included services rendered in 2019 and 2020, as well as 2021. After matching, the event file was edited to exclude all services that were rendered outside of calendar year 2020.

- ▶ If the survey-reported data matched Medicare claims data, the dates of service on the Medicare record were carried into the event record. Dates of service were used as a match criterion in most of the matches, so in many cases, the dates of service in the event record did not change from those reported.

C.2.3.4 Summary of Match Results

A total of 125,075 Medicare bill events for beneficiaries living in the community were matched against 132,140 survey reports. A match was recorded for 40,427 event records, which is 31 percent of total Medicare bill records events and 69 percent of survey-reported events. The percentage of dollars matched was considerably higher.

Some small part of the unmatched 84,648 Medicare records are undoubtedly represented in the 91,713 survey-reported events that could not be matched under the matching criteria used. However, the 91,713 unmatched survey events would be expected to include a substantial share of events that are not covered by Medicare, and therefore would not be expected to match a Medicare paid claim. In addition, only 7,942 of the 91,713 unmatched survey-reported events have a Medicare payment amount. The 86,648 unmatched Medicare billing records strongly suggest that the survey reports seriously understate the number of Medicare services when compared to CMS billing records.

The under-reporting problem was more serious for event counts than for Medicare payments. The 86,648 unmatched Medicare bill events represent 69 percent of events, but only 52 percent of total payments. That is, 69 percent of total dollars on the Medicare bill side were successfully matched with survey reports. Unmatched Medicare events (\$268) were a little more than half as expensive on average as matched events (\$456). This is consistent with general household survey experience, which has shown that more expensive medical events are more likely to be remembered and reported at the interview. This trend is consistent with the hypothesis that survey respondents tend to remember major health events/more expensive events better than minor health treatments.

In addition to correcting for unreported events, the match also helped to fill in missing Medicare payment amounts and correct Medicare payment amounts that had been reported incorrectly. Of the 40,427 survey events matched to Medicare bill records, Medicare was reported as a payer on 56 percent of these events, and a Medicare payment amount was reported on 39 percent of these events. This means that the match and reconciliations generated corrections that made Medicare a payer of record on the 44 percent of cases where this information was originally omitted in the survey reports; and made it possible to determine the correct Medicare payment amount in the 61 percent of survey records where this information was omitted.

Not all services could be cleanly and easily matched from the two sources. The match employed "strength of evidence" criteria and "hierarchical algorithms" in order to identify matches, survey reports only, bill file reports only, and a small number of similar events for which it was not clear whether there was duplicate survey and bill reports or not. The methods and criteria used in the match are discussed in more detail in the Event-Level Matching discussion in section of this manual. In addition, Eppig and Edwards' paper, "Computer Matching of MCBS Data with Medicare Claims," presents a full discussion of methods and criteria.¹⁵

¹⁵ F. J. Eppig and Brad Edwards, "Computer matching of Medicare current beneficiary survey data with Medicare claims," in *Library of Congress Cataloging-in-Publication Data* (1996), 191.

C.2.3.5 Evidence supporting improved accuracy

On the 40,427 matched events, Medicare should have been reported as a payer on 100 percent of the survey reported events. However, Medicare was only reported as a payer for 22,644 or 56 percent of events. Consequently, the match corrected 44 percent of the records to make Medicare a payer of record.

On the 40,427 matched events, the Medicare payment amount was only reported on 39 percent of survey reports. The match filled in the correct Medicare payment for the 61 percent of survey reports.

C.2.3.6 Evidence of survey under-reporting

The 84,648 unmatched Medicare paid bill events strongly suggest a high level of under-reporting on the survey. While there are 91,713 unmatched survey reports on the other side, many of these events could not be reasonably expected to be undiscovered matches. For example:

Unmatched survey events unlikely to match an unmatched Medicare bill:

1. 9,929 unmatched survey events were for dental services that are rarely covered by Medicare.
2. 2,257 unmatched survey events had total payments equal to zero. (These were very likely parts of bundles of services that were covered in one global payment on the Medicare side, for example, post-operative services that were covered by a global surgery fee.)
3. Another 38,039 unmatched survey events were for MA enrollees. Virtually all of the Medicare services for these persons are paid through a capitated payment amount and the likelihood is very small that their events ever match a fee for service Medicare paid bill record.
4. There were 323 unmatched survey events where the beneficiary was only entitled to Part A or Part B of Medicare, but not both. Therefore, a survey reported service could reasonably not be expected to match a Medicare paid bill record.
5. Another 2,147 unmatched events were provided by the Veteran's Administration (VA) or in a military installation where no Medicare bill would be expected.
6. 12,470 unmatched survey events were for other medical services. While Medicare covers durable medical equipment such as wheelchairs, and supplies such as oxygen, it does not cover many items in the broad other medical services category such as eyeglasses, hearing aids, heating pads, incontinence supplies, etc. Average payments for unmatched survey reports of other medical events (\$207) was less than the average survey reported payments for matched events of this type (\$626) and less than total payments for unmatched Medicare claims in the same category (\$392). This suggests that most unmatched survey events for other medical services are probably not undiscovered matches.
7. Taken together, over 65,165 of the 91,713 unmatched survey events either definitely could not, or very likely would not, match a Medicare bill event record. This leaves 26,548 unmatched survey events to be explained.
8. This means 58,100 medical events, or 43 percent of Medicare bill records for beneficiaries living in the community, were not reported in survey interviews. (Calculated using 84,648 unmatched Medicare events minus 26,548 possible undiscovered matches among the unmatched survey events)

Unmatched survey events likely to be undiscovered matches:

9. On the other side, 7,942 unmatched survey-reported events reported a dollar amount that Medicare paid for the event. These unmatched survey events are very likely to be undiscovered matches.

Ambiguous events:

10. This leaves 18,606 unmatched survey events to be explained. There are many medical services and supplies that Medicare does not cover such as most alternative medicine services, over the counter remedies, etc.

C.2.3.7 Building a Complete File

Medicare covered services:

- A complete file would include all 40,427 matched events. These events, which were reported on both the survey and in Medicare bill event records, will combine the most accurate and complete information possible from both sources.
- All Medicare bill record unmatched events (84,648) should also be included. These event records are official records of Medicare program payments and will partially correct for survey under-reporting.
- It is more debatable which of the unmatched 91,713 survey records to include. The data include type of service, adjusted file summaries, and all unmatched survey reports except the 7,942 records with a Medicare payment. For the reasons discussed above, these 7,942 records are likely undiscovered matches that would duplicate some of the 84,648 unmatched Medicare bill event records if they were included. For official MCBS reporting purposes, CMS uses all unmatched survey reports except the 7,912 records with a Medicare payment.
- Hospice records, which were not entered in the event-level match, should be added into the file.

Total medical services including Medicare covered and non-covered services:

- In addition to the Medicare covered services listed above, Prescription Drug and Long-Term Facility records should be added to the file.

C.3 Imputation Information

There are several adjustments made to fill in payment amounts that are missing because the respondent did not know how much an event cost or did not know how the event was paid for (by whom, and how much for each payer). An abbreviated description of the process and procedures used to impute these missing values is available in the *2020 MCBS Methodology Report* located on the CMS MCBS website at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks>.

First, a target reimbursement or total payment for the event was established, all possible sources of payment were identified, and then the total payment was distributed across all payers. Missing amounts and payers were filled in using either analytic editing or statistical imputation. This process relied heavily on Medicare administrative records. The guiding principle of retaining as much survey data as possible, and filling in around it, was maintained throughout the process. Where feasible, information about the payers for a specific event, known payment amounts, and target reimbursement were used to determine unknown payment amounts by analytic edits. When insufficient information was available and analytic editing was impossible, unknown payment amounts were completed by statistical imputation.

Different approaches were used with different categories of utilization to define payers and determine payment amounts. Records submitted to the survey/administrative match (which was discussed in the Technical Appendix C.2.3, Event Level Matching) were handled differently from those not matched. Survey-reported records for dental, vision, hearing, medical practitioner, inpatient, outpatient, institutional (other than long-term care), and medical equipment and supplies (survey utilization categories DU, VU, HU, MP, SD, SL, IP, OP, IU, and OM) were entered into the match with Medicare claims data. After the match, these events were individually assigned target reimbursement amounts, and then source of payment variables and separate payment amounts were calculated for each payer. Other procedures, usually some adaptation of the procedures sketched above, were used to determine payers, target reimbursements, and payments for other

categories of utilization. The next section discusses how target reimbursements were established, explains the procedures used for matched utilization (the largest category of utilization), and then discusses the smaller and more specific non-matched categories.

C.3.1 Determining target reimbursement

A primary rule was to establish the target reimbursement for an event with a missing total payment prior to determining or imputing the payment distribution. This was done to establish a target reimbursement that was consistent with payments shown for other similar services in the file. In this way, a credible target reimbursement can be used to inform and control the payment distribution. For Medicare covered services, target reimbursements were developed from Medicare claims; this method is more accurate than determining the amounts paid by individual sources of payment and summing them.

Another primary rule was to retain survey-reported payment data, even when it was only partial data, wherever possible. There are situations where retaining the reported payment amounts and establishing the target reimbursement amount without regard to individual payment amounts are mutually exclusive. On a few occasions, the target reimbursement had to be adjusted in order to retain reported payment data.

The rules for establishing target reimbursements depend first on whether or not Medicare claims data are available. If the survey-reported data match a Medicare claim record, or if the Medicare claim record was the only source of information about the service (i.e., nothing about the service was reported in the survey), the Medicare claims data were used to establish a target reimbursement. The target reimbursements for 75 percent of the events in this file were established using Medicare administrative bill payment data.

If the utilization was only reported in the survey (i.e., matching to Medicare claims was not successful in identifying a corresponding claims record), the survey data were used to create the target reimbursement. This occurred for about 25 percent of events in this file.

For a small subset of the survey reported events without a matching Medicare claim, but where Medicare was reported as a payer, a different approach was used to create a target reimbursement. A set of regression models, one for each type of event, was developed to predict the target reimbursement from the total charges reported in the survey.¹⁶

When the respondent did not report a total charge for the event but indicated that Medicaid was a payer, an imputed target reimbursement was created which was consistent with the generally lower payments made by Medicaid.

C.3.2 Filling in Missing Payments and Payers for Matched and Not Matched Utilization Records

The following procedures were used to determine who paid for each event, how much an event cost in total, and how much each payer paid. These procedures were applied to events in the 2020 Cost Supplement File designated: IPE (inpatient), OPE (outpatient), IUE (institutional), DUE (dental), VUE (vision), HUE (hearing), PME (prescribed medicine), and MPE (medical and surgical services, equipment and supplies).

¹⁶ Note that these records with a target reimbursement amount estimated by the regression models are the same records described in section 3.1: Analytic Notes for Non PM Event Segments. Most of these records were not included in the summary files (SOWMP = 1) because they were largely events with FFS Medicare payments. Some were retained in the summary files if the beneficiary had MA coverage (i.e., they were considered MA payments and not expected to have matched FFS administrative claims).

See section C.3.2.11 for information on how missing payments and payers are filled in for the FAE (long-term care facility) events.

C.3.2.1 Determining potential payers

Regardless of the method used for imputation, payment amounts were only imputed for potential payers. The total reimbursement for an event was distributed among 10 sources of payment (SOP):

- Medicare FFS
- Medicaid
- Medicare managed care
- Private insurance managed care
- Employment-based private health insurance
- Individually purchased private health insurance
- Private insurance, source unknown
- Out-of-pocket
- Uncollected liability
- Other public insurance

It was determined that payments made by the VA could not be estimated with sufficient accuracy. Therefore, in 2016 and beyond, payments from the VA are combined into the “other public insurance” source of payment.

Out-of-pocket payments are those payments made by the beneficiary or their family, either as cash or through Social Security or SSI checks to a nursing home. Medicare MCOs coverage (i.e., Medicare Part C/MA) is different enough from FFS coverage to merit its being reported separately. Non-MCO private insurance is characterized as individually purchased or employment-based because there are differences in cost and coverage depending on type. As this information is not known for residents of nursing homes (the nursing home staff are not likely to know, and thus are not asked, how the insurance was purchased), a third category of private, non-MCO insurance was created for private insurers when the source is not known. Uncollected liability refers to unpaid amounts where there is a legal obligation to pay. If there is an agreement between the provider and a payment source, which reduces the amount that the provider can collect for a service, there is no uncollected liability. On the other hand, if the respondent reports a total amount payable and specific payment amounts for all known sources of payment, and the sum of those payments is less than the total amount payable, the difference is considered an uncollected liability. Other public insurance includes payments made by the VA, as well as federal or state programs not included in the other categories, such as state drug programs like PACE in Pennsylvania.

An individual's insurance coverage can change during the course of a year. A health insurance timeline,¹⁷ created for each person in the 2020 Survey File, provided the basis for determining the potential payers for each event. The timeline contained complete insurance information, including Medicare entitlement, Medicaid eligibility, and enrollment in Medicare MCOs (Medicare Part C/MA), for every day of the beneficiary's Medicare eligibility during the year. Medicare entitlement, Medicaid eligibility, and enrollment in Medicare MCOs were captured from CMS administrative data, while information about private insurance was collected in the insurance portion of the interview and then supplemented by information obtained from statements and Medicare claims. In 1996, the methodology was refined to determine whether Medicaid was a possible payer for an event. It is now possible to distinguish whether a beneficiary has full Medicaid benefits or only Qualified Medicare Beneficiary (QMB) or Special Low-income Medicare Beneficiary (SLMB) status. If an individual has full Medicaid benefits, then Medicaid is a potential payer for all medical events. If an individual has QMB-only

¹⁷ The Health Insurance Timeline segment (HITLINE) containing the types of insurances, the coverage eligibility timeline, and the source information for the coverage is included in the Survey File LDS.

status, then Medicaid is a potential payer for cost-sharing amounts on Medicare-covered services, but not on medical events not covered by Medicare. If an individual has SLMB-only status, then Medicaid is not a potential payer for any medical events.

C.3.2.2 Payer indicators

A payer indicator code was used to identify definite and potential payers of the total charge for an event. Source of Payment (SOP) flags were used to initialize the payer indicator. Each SOP flag corresponded to one component of the payer indicator and could have a value ranging from 0 to 4 as shown below.

Source of payment (SOP) flag values:

- 0 - Source definitely did not pay
- 1 - Source definitely did pay, known amount
- 2 - Source definitely did pay, unknown amount
- 3 - Source possibly paid, beneficiary was covered at time of event
- 4 - Source possibly paid, beneficiary may have been covered at time of event

SOP values were set by using survey information about reported events, about the type of provider for the event (that is, whether the service was delivered by a managed care provider or a VA facility), and about the type of insurance coverage and/or program participation. SOP values also depended on Medicare claims data when a survey-reported event corresponded to a Medicare claim (a "matched" event). Based on all of this information, each source was determined to be a payer, a potential payer, or not a payer of charges for the event.

- Payers - A source was a definite payer if the SOP for that source had a value of 1 or 2. An SOP value of 1 indicates that the respondent reported that the payer had paid a portion of the charges and also reported a payment amount, or that Medicare claims data provided that information. An SOP value of 2 means that the respondent reported that a payer paid a portion of the charges, but did not know the exact amount, and no matching Medicare claim was found to provide this information.
- Potential payers - A source was a potential payer if the corresponding SOP had a value of 3 or 4. An SOP value of 3 meant that either the beneficiary definitely had that type of insurance coverage at the time of the event and the payer may have paid some amount, or the beneficiary received the service from that type of payer (i.e., a managed care provider or a VA facility), but did not report it as a payment source. An SOP value of 4 was used when there was doubt about the beneficiary's insurance coverage during the event or about the event date itself.
- Non-payers - If neither the respondent nor the Medicare claims data indicated that a payer had been a source of payment for an event, the SOP was set to 0.

A more comprehensive discussion of the rules used for setting the SOP flags is included in Technical Appendices C.3.4.

C.3.2.3 Translating payer indicators into sources of payment for matched utilization records only

A value of 1 for a particular payer indicator meant that the payer paid a portion of the total charge for the event, and a value of 0 meant that the payer did not contribute. Final payer indicator values were determined in one of three ways: 1) directly from the corresponding SOP values; 2) through analytic edits; or 3) through statistical imputation.

Different rules applied when payer indicator values were set directly from the corresponding SOP values, depending on whether the SOP was determined to be a definite payer, a potential payer, or a non-payer. The payer indicators were initialized as follows:

- If the source was a definite payer and the payment amount was known (SOP=1), the corresponding payer indicator was set to 1.
- If the source was a definite payer but the payment amount was not known (SOP=2), the payer indicator value was set to 1 with one exception.
 - ▶ If the event was for dental, vision, or hearing care or for durable or nondurable medical equipment not usually covered by Medicare, the Medicare payer component was set to 0. The rationale was that if the respondent was not able to report the Medicare payment, then it was more likely that Medicare had not actually paid for the ordinarily non-covered services.
- When the SOP was a potential payer (SOP=3 or 4), the corresponding payer indicator was set to missing and imputed (as 0 or 1) in a later step. However, the general rule for setting a payer indicator value based on a corresponding SOP value of 3 or 4 was sometimes modified by analytic edits.
 - ▶ For example, the Medicare payer indicator value was never set to missing. It was always equal to 0 or 1. When the SOP for Medicare was listed as a potential payer (SOP=3 or 4), which was rare, the Medicare payer indicator was set to 0 (Medicare did not contribute).
- When the SOP was not a payer (SOP=0), the corresponding payer indicator was set to 0, with exceptions for out-of-pocket payments and uncollected liability.
 - ▶ If the SOP was out-of-pocket or uncollected liability and the SOP value was equal to zero, the payer indicator was set to missing, to be imputed as 0 or 1 in a later step.

C.3.2.4 Analytic edits

Analytic editing of charge and source of payment data at the event-level also determined some payer indicator values. The general goal of the analytic edits was to resolve as many events as possible (i.e., to fully allocate total charges to payers) and to set as many payer indicator values as possible based on logic and knowledge of payer policies. The edits resolved some events without using a hot deck procedure to impute payment sources or amounts.

The analytic edits relied on having both unambiguous SOP values and external information about interaction among the insurance or payment sources. Edits for two of the payment sources (Medicaid and MCOs) depended on information specific to those payers, but payer indicators for other payment sources were also affected. The analytic edits are discussed fully in Technical Appendix C.3.4, as they apply to each source of payment.

Medicaid: Analytic edits were used extensively when Medicaid was a potential or actual source of payment for an event. One set of edits, designed to reflect the role of Medicaid as the payer of last resort, ensured that Medicaid could not be a payer if payments were reported or imputed for another third-party insurer (except Medicare), or if the provider was an MCO or VA facility. Another set of edits was developed for dually eligible beneficiaries whose cost-sharing liability is covered by Medicaid. For additional information, see Technical Appendix C.3.3.

Private and Medicare MCOs (Medicare Part C/MA): MCOs (especially Medicare-contracting MCOs) often operate differently from other third-party payers and tend to have unique payment patterns. For instance, risk and (to a lesser extent) cost Medicare MCOs are paid a set fee per enrolled Medicare beneficiary (called a capitated amount) designed to compensate the MCO for the expected costs of delivering Medicare's package of benefits. There are no Medicare claims or Medicare or insurance statements indicating the total charge for events covered by the capitated amount. Often the respondent only knows the copay amount, if there was one. Also, MCOs often provide "Medigap"-type coverage by paying for most of the member's deductibles and copays for Medicare-covered benefits. A beneficiary who belongs to an MCO does not need private Medigap insurance or

Medicaid coverage for these amounts. Thus, payment patterns for MCO beneficiaries tend to be simpler than those for FFS beneficiaries. The set of analytic edits for MCOs attempted to account for these simplified patterns and for the respondent's usual inability to report charges and payments for events. The MCO edits also attempted to avoid creating "illogical" payment patterns.

General Edits: At the beginning of the analytic editing, and after each main section of edits, an attempt was made to resolve events through addition or subtraction. Events without a known total charge but with a complete payer indicator vector (i.e., each payer was identified as either having paid or not paid for an event and each payer's amount was known) were completed by summing across all payment sources to derive the total charge. Events with a known total charge and complete except for one missing payment amount or payment source were completed by subtraction. The excess of charges over known payment amounts was attributed to the known payer, or the one missing payer indicator was set to 1 and the excess allocated to that payer.

If a service was provided free of charge, all payer indicators and payment amounts were set to 0. However, if the respondent reported an event as free, but also reported that a source other than Medicare or Medicaid paid something for the event, the total charge was reset to "missing," and imputed.

If a source was a potential payer for an event, or if the respondent reported that the payer had contributed to an event but did not know the amount, it was assumed that the payer was not actually a source if the current sum of reported payments equaled the reported total charge.

Payer Indicator Imputation: For information on this process, please see the *2020 MCBS Methodology Report* located on CMS' MCBS website at <https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks>.

If Medicaid was a payer, a Medicaid payment amount was calculated as a percentage of coinsurance and deductible for the Medicare service.

C.3.2.5 Other Utilization (Not Matched)

The following procedures were used to determine who paid for each event, how much an event cost altogether, and how much each payer paid, for events that were not matched to Medicare claims data on a service-by-service basis, specifically hospice services. (The procedures used for missing payments or payers for prescription drugs and facility utilization are described separately below. For information on the editing and creation of these types of utilization, refer to the Prescription Drugs and Long-Term Facility segments). Long-term facility and prescription drug utilization are presented in the 2020 Cost Supplement File as event-level records designated FAE (facility) and PME (prescription medicines). Hospice records are presented as summary records only.

C.3.2.6 Hospice Services

Hospice utilization is unusual in terms of Medicare administrative records because it is the only utilization that is recorded in two different ways, in two different files. The beginning and ending dates of the hospice benefit periods are recorded in Medicare Administrative enrollment data, while the bill records are part of the NCH repository. This dual reporting served as an internal check on the dates of service on the billing records.

C.3.2.7 Determining and imputing payment amounts for Hospice Services

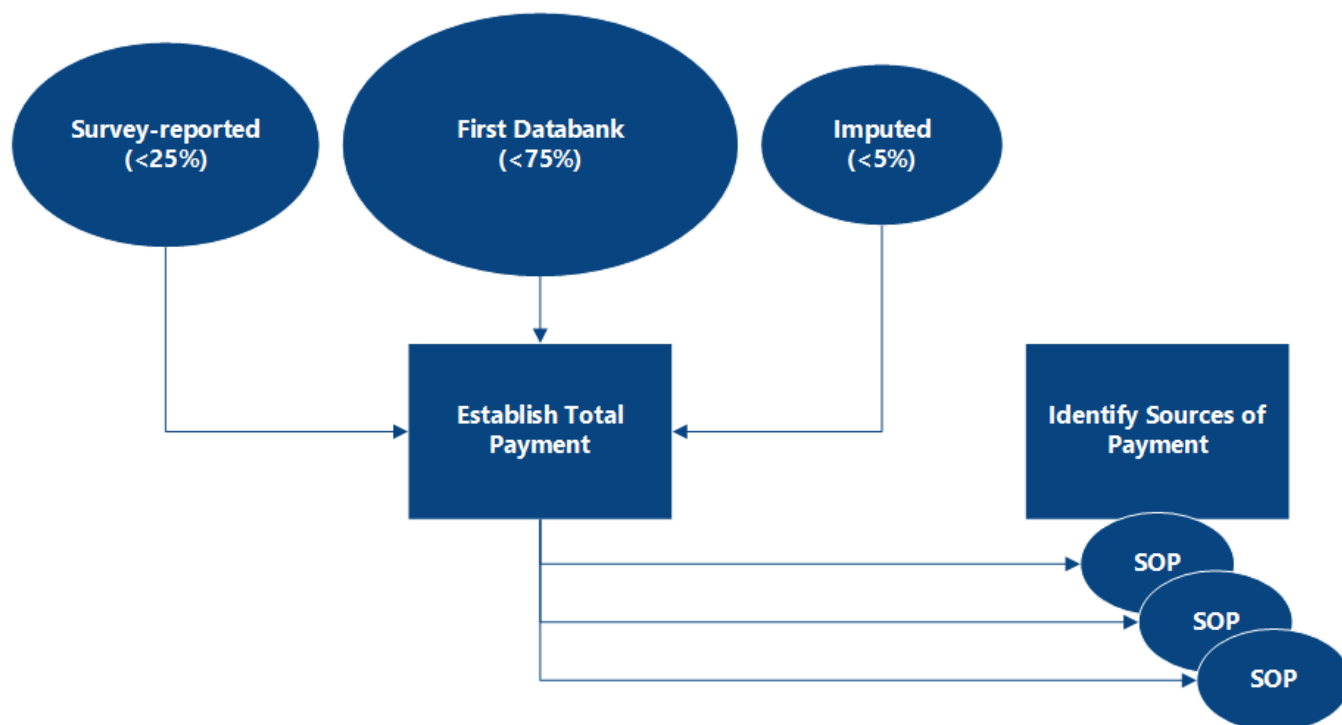
With a target reimbursement amount (representing the "total cost" of the event) and payer indicator values indicating which payers contributed some payment toward the total, the share "amounts" paid by the individual payers could be determined.

If Medicare payments were known to be incomplete, then utilization for the missing periods was completed by editing from the existing billing records. The hospice benefit is paid on a per-diem basis, and the missing data were completed with average per diem rates calculated from existing bills. Virtually all services provided to the hospice beneficiary are fully covered by Medicare, and as there are no copayments or deductibles, there is no cost sharing (Prescribed medicines are an exception, as there may be a small copayment for drugs, which are reported separately, and also inpatient respite care for which the patient pays 5 percent of the Medicare allowed rate – typically under \$5). Hence, the Medicare reimbursement is the target reimbursement, and Medicare is the sole payer of hospice bills. Hospice bills were not matched; as a result, there is some overlap between hospice utilization and events reported in the survey. The overlapping survey events are usually, but not always, home health events. When an unmatched survey event occurs at the same time as a Hospice event (i.e., is an overlapping event), the event is retained but the payer indicators and amounts are zeroed out to prevent double counting.

C.3.2.8 Prescription Drug Data

Prior to 2006, all prescription drug data was based on information collected in the survey. Due to the advent of Part D, beginning in 2006 Part D Events (PDEs) were available for Medicare beneficiaries enrolled in MA Part D Plans (MAPDs) or stand-alone Part D Plans (PDPs). For survey-reported events that matched to PDEs, information on the drug “claim” was used to calculate the total payment field as well as the other payer fields. The approach used to fill in missing drug payment data for unmatched events was similar to that used for other missing payment amounts described above. The first step was to establish a total payment amount for each drug event. First preference was given to using survey reports of the total payment for the drug. For approximately one-quarter of survey-only drug events on the file, the total payment was reported. For about three-quarters of the survey-only drug events, an administrative drug pricing source (FDB MedKnowledge (formerly NDDF Plus)) was used to impute prices. The administrative source was used only when no total payment was reported, and it was rarely used to supersede the survey reported payment. Finally, a small proportion (<5%) of survey-only drug events had total payments established using statistical imputation techniques.

After the total payment was established for each drug event, potential sources of payment were identified using a similar approach to that outlined earlier. In the last step, the total payment amount was distributed across the sources of payment (see Exhibit C.3.2.8). In cases where a total payment was available from either a survey report or the FDB, unknown payment amounts for a specific payer were handled by accounting techniques and analytic edits before employing statistical imputation. In a small percentage of cases where the total payment was derived by statistical imputation, the payer amounts were also derived through statistical imputation.

Exhibit C.3.2.8: Establishing Total Payment Amount and Allocating to Sources of Payers**Preparation of survey-reported data**

Prior to matching or imputation, the prescription drug data collected in the survey were edited for consistent spelling. MCBS staff edited the records to ensure that the same drug was always reported in the same way. All unique drug name spellings supplied in the survey for medications reported for all four survey panels through the end of 2020¹⁸, were gathered together in a single list. Using the FDB, MCBS staff manually assigned corrected spellings to each unique supplied spelling.

Preparation of FDB (First Databank) data

The 2020 FDB served as a pricing reference and as a source of therapeutic class for prescription medicines. However, survey reports of total payments were given preference over a FDB price because MCBS records and FDB records could not be matched exactly on all fields. The FDB generally identifies the name, form, strength, and packaging size of the drug in a single entry. The MCBS collected prescription size in the survey but could not collect the packaging size of the drug prescribed. In the survey, form and strength are also collected, but as separate items, not as part of the name. In the initial match, therefore, a FDB name "Septra DS Tab 800 mg" had to be changed to "Septra DS," to increase the likelihood of a match between the two sources on name.

Assignment of wholesale prices

In the FDB, a wholesale price is assigned to each NDC entry. The NDC is an 11-digit code; the first five digits define the manufacturer or labeler, the next four digits identify the drug product, and the last two digits identify the packaging size. As noted above, the MCBS does not collect packaging size, but instead collects prescription size, and unit average wholesale prices can differ substantially by packaging size. Using a relative frequency distribution of packaging sizes within each drug type, weighted by utilization rates from Part D

¹⁸ The prescription drug data are cleaned for all panels included in the current LDS release (i.e., 2017 through 2020 Panels). Winter 2018 (Round 80) is the first round prescription drug data are collected for the 2017 Incoming Panel. Winter 2021 (Round 89) is the last round prescription drug data are collected for all panels given that prescription drug data for 2020 through 12/31/2020 are not reported until the Winter 2021 interview.

prescription drug event data, MCBS staff developed a composite price for drugs that come in multiple packaging sizes.

After both survey data and FDB data were cleaned as described, survey prescription data were matched to the modified FDB information by drug name, form, strength, and packaging size, in that order, to develop a wholesale acquisition cost. Often, it was not possible to match on all four variables. If the survey drug name was not known or could not be matched, a wholesale acquisition cost was imputed. If the drug name was known but form or strength was not known, the missing characteristic was imputed and the wholesale acquisition cost was then obtained through a match to the FDB. For example, if the respondent reported a prescription of Diamox but did not know the strength, a wholesale acquisition cost was imputed using the weighted average price of all Diamox prescriptions. This weighted average price was developed using Part D prescription drug event data.

A small number of survey entries could not be translated to any drug listed in the FDB. In general, these entries were either misspellings that made it impossible to determine the drug name or not a specific drug (e.g., "little green pills"). These entries were classed as "untranslatable," and an average price was imputed based on frequency distributions of drug claims from Part D prescription drug event data. In some cases, the size of the prescription was known but the price was not. Average unit costs (per pill, per milliliter, etc.) were then multiplied by the prescription size, to derive a whole prescription cost. In other cases, prescription size was estimated through the respondent's answers to a series of probe questions, which were asked during the interview when the respondent did not know the size of the prescription.

Converting wholesale acquisition cost into event price

Establishing a price for prescription drug records with no survey reported price began with the assignment of wholesale acquisition cost. Event prices that were less than \$.50 were reset to missing and imputed statistically. Non-missing wholesale prices were multiplied by a pricing factor that varied depending on the likely payer(s) of the event. Six pricing factors were developed: retail, MCO, VA, Medicaid, employer sponsored and other public insurer. The retail pricing factor was actually a series of factors which reflected empirical evidence of the relationship between the wholesale acquisition cost and what the respondents reported paying. The retail factor was 39 percent. The managed care pricing factor estimates that MCOs pay approximately 151 percent of the wholesale acquisition cost of prescription medicines. The VA factor was developed using VA drug cost data that was provided by the Department of Veterans Affairs. The Medicaid pricing factor was developed using composite data from CMS' Medicaid Drug Rebate System, and included a dispensing fee of \$4.77, a discount off the wholesale acquisition cost (53 percent) and a rebate percentage of 24.1 percent. The employer-sponsored insurer is 104 percent of the wholesale acquisition cost.

Determining target reimbursement

Target reimbursements were developed differently for prescription medicines than for other services (Target reimbursements for other types of services are described above in Technical Appendix C.3.1). If Medicare claims data were unavailable, adjusted "event prices" were used to develop target reimbursements.

The target reimbursement is defined as the price that the beneficiary paid for a single purchase of a single drug. For single purchases (one unique medicine, purchased only once and not refilled), the price reported by the respondent was the target reimbursement.

If the respondent could not give a price, the event price, adjusted by the appropriate pricing factor (discussed below) was the target reimbursement. For multiple purchases (a single prescription, filled multiple times, or multiple prescriptions), the target reimbursement was developed as for single purchases and then divided by the number of purchases to yield a target reimbursement for each purchase.

If several drugs were reported together (“bundled”), but the total cost was not known, a target reimbursement was developed for each drug in the bundle, based upon the event price adjusted by the appropriate pricing factor. If several drugs were bundled together and a total cost was reported, that total cost was used to control the imputation of the individual drug prices. A relative percentage of the total cost was developed for each drug, using the event price adjusted by the appropriate pricing factor; those percentages were applied to the reported total cost, and the result became the target reimbursement for each drug. If the event price for one or more of the drugs in the bundle was missing, an average price for all strengths and forms of the drug was used in the computation, unless the drug name was not known; in those cases, an average event price was used. These averages were then used to calculate relative percentages, which were then applied to the amount reported in the survey for the bundle.

Determining potential payers

Potential payers for prescription medicines were determined in essentially the same way that potential payers were identified for matched utilization.

C.3.2.9 Adjustments for Missing Days and Undated Use

This section describes the adjustments made (at the person and service level, but not at the event-level) to:

1. Compensate for data that are missing because some periods of the beneficiary’s Medicare entitlement were not covered by interviews. CMS administrative records are used to establish the exact period of Medicare entitlement during 2020 and calculating the number of Medicare days;
2. Allocate undated survey events, primarily prescription drugs and some other medical equipment, between years where interview reference periods spanned two years.

Calculating Medicare covered days and residence history

The periods of Medicare entitlement and living situations are established in order to validate and supplement utilization reported in the survey with information reported on claims and bills from CMS’ national claims history database. This is accomplished by matching survey-reported utilization to the CMS records that was described earlier in the section on Filling the Gaps (Appendix C.2.2).

For most beneficiaries, the period covered by the survey and the period of the beneficiary’s Medicare entitlement are identical: they both cover all 366 days of 2020. However, the Medicare entitlement period may be longer than the period covered by the survey for beneficiaries who left the survey before the end of 2020, or died, and no proxy information is available. The most common reason for incomplete data is the beneficiary’s refusal to participate further in the survey. Only beneficiaries who participated in the survey for at least 60 percent of the period they were eligible for Medicare during the year were retained for the 2020 Cost Supplement File.¹⁹

To identify Medicare entitlement and where the beneficiary was residing during the survey period (in the community or in a facility), three variables are provided in the Residence Timeline segment: total number of days entitled to Medicare (D_T DAYS); number of days where the beneficiary was living in the community (D_C DAYS); and number of days where the beneficiary was living in a facility (D_F DAYS).

Allocating services between years

The cost and utilization data collected during the 2020 interviews cover more than just that calendar year. Each interview serves as a boundary to the next interview - the respondent describes medical care that took place “since the last interview” - and those boundaries are generally not the beginning or ending of the

¹⁹ This population aligns with the Cost Supplement ever enrolled weights population. Please see section 3.5 of this document and the *2020 MCBS Methodology Report* for more information on the ever enrolled weights.

calendar year. As a result, the first (Fall 2019 Round 85) or last (Winter 2021 Round 89) interviews generally include utilization that covers part of two calendar years. To adjust the utilization in these cases, dated event records were edited to remove those that took place outside of 2020, and undated events (prescription medicines) were pro-rated according to the number of 2020 days in the interview reference period to total days in the reference period.

Simply pro-rating use between the two calendar years was considered but rejected. By assuming that use occurred in both years, this procedure could overstate the number and rate of persons using services in a year. In place of this, a random number generator was used to assign services to calendar years. The probability of an event being placed in 2020 was based upon the ratio of 2020 days in the reference period to total days in the reference period. For example, assume a reference period had 120 days, and 90 of these days were in 2020. For each event, a random number between 1 and 120 was generated. For all events where the random number was 90 or less, the service was allocated to 2020. For all events with random numbers between 91 and 120, the service was allocated to the other year.

Filling in Medicare covered days not surveyed: PM data

When there is a gap in survey data, that is, a period for which a beneficiary was enrolled in Medicare but was not covered by a survey interview, it is necessary to estimate the medical service usage during that gap period. For beneficiaries with gaps who were interviewed in 2020, reported services were simply prorated upward to cover the gap. For example, for prescription drugs, the number of prescriptions per day was calculated for the interview period and multiplied by the number of gap days. This assumes, in effect, that the beneficiary used prescriptions at the same rate in the interview and gap periods. Likewise, to get adjusted sums for all payers, the cost per prescription per payer per day was calculated and multiplied by the adjusted number of prescriptions for each payer.

If the beneficiary was not surveyed due to a missed interview, a different approach was used. To cover these non-interview gap periods, a donor was selected who was similar to the person in terms of personal and economic characteristics. The donor's use of prescription drugs (measured in prescriptions per day and cost per prescription per payer per day) was used to impute use and payment data.

Filling in Medicare covered days not surveyed: Non PM data

For non-prescription events with a gap in survey data, reported services were prorated upward to cover the gap. This proration was provided for the Non PM summary data. There was no adjustment made to the event-level data. For these gaps in survey data, no additional events were added, nor were the individual events for these beneficiaries adjusted; however, when the data were summarized to the person level or summarized to the service level, the proration was applied. The ratio applied was simply the ratio of the number of days in the year divided by the number of days for which survey data were available for the beneficiary. For example, if a beneficiary had a 100 day gap in a 365 day year, then survey data were available for 265 days, and the ratio would be $(365 / 265) = 1.377$.

For Incoming Panel beneficiaries, a different approach was taken to estimate data for gap periods. For most of these beneficiaries, the first survey interview does not cover medical expenses going back to the date of enrollment. Therefore, there is a need to estimate medical events from the beneficiary's date of enrollment through the earliest of either: the date of their next interview, their date of death if they have passed away, their date of lost entitlement if applicable, or December 31st. Since medical claims data that have been processed through FFS Medicare are available, these claims do not need to be estimated. However, there are many medical events that are either covered by MA or not covered by traditional FFS Medicare. Examples of these uncovered events may include dental visits or regular physician visits if the beneficiary does not have Medicare Part B. For these cases, new records are created to fill in Medicare covered days that have not been surveyed in a process known as unit-level imputation.

In the unit-level imputation process, a donor beneficiary is selected from a pool of beneficiaries with similar demographic attributes and insurance coverage to the Incoming Panel beneficiary. Selected data records are copied from the donor beneficiary and assigned to the Incoming Panel beneficiary. Only medical events that occurred within the Incoming Panel beneficiary's gap period are retained. If the donor beneficiary did not have any medical events that were not covered by FFS Medicare during the Incoming Panel beneficiary's gap period, then no additional data records were created. The new data records have payment information and amounts provided by the donor beneficiary's data record, but demographic and other variables associated with the Incoming Panel beneficiary are retained.

C.3.2.10 Determining and Imputing Payment Amounts for Long-Term Care Facility Events

The payments for time periods when a beneficiary lived in a long-term care facility contained on the FAE segment were determined from the billing period and payment information collected via the Facility Instrument. This information was not always available for collection, so imputation was needed to fill in payments for these missing periods.

To impute payments for the periods of missing data, information was gathered from time periods where payment information was reported. The reported payments for basic and ancillary services from the long-term care facilities were summed over the period of provided data and then divided by the number of days within that period. Medicare payments were excluded from this calculation because these payments were unlikely to be missing. The resulting daily rate was then applied to estimate payments for the periods of missing data by multiplying the daily rate by the number of days within the period of missing data. For some beneficiaries, either no payment information was available or the only known payments were from Medicare. For these beneficiaries, standard daily or monthly rates were gathered from the given facility, and these were used to estimate a reasonable daily rate, which was then used to estimate payments for the periods of missing payment information. A small number of facilities did not have the standard daily or monthly rates available. For beneficiaries living at these facilities, a median daily rate was calculated based on the rates provided by the facilities that had standard rates available.

Once the total payments for each stay at a facility with missing payment information were estimated, then amounts were allocated to the appropriate source of payment (private insurance, Medicaid, out of pocket, supplemental security, VA, or other) based on the allocations observed for periods where payment information was available. If there was no observed payment information available for a given beneficiary, averages were calculated based on data from beneficiaries that did have payment information available. Finally, the imputed payments were reviewed for extreme values and edited if necessary.

C.3.3 Analytic Edits of Sources of Payment (SOP) Values for Non PM Events

The general goal of the analytic edits is to resolve as many events as possible (i.e., to fully allocate total charges to payers) and to set as many payer indicator values as possible based on logic. The edits resolved some events without using a hot deck procedure to impute payment sources or amounts.

C.3.3.1 Medicaid

Analytic edits were used extensively when Medicaid was a potential or actual source of payment for an event. One set of edits—designed to reflect the role of Medicaid as the payer of last resort—ensured that Medicaid could not be a payer if payments were reported or imputed for another third-party insurer (except Medicare), or if the provider was a MCO or VA facility. Another set of edits was developed for dually eligible beneficiaries whose cost-sharing liability is covered by Medicaid.

Out-of-pocket payments were allowed when Medicaid was a payer only if the respondent was able to report the out-of-pocket amount(s). Medicaid may cover copays and deductibles for dually eligible beneficiaries and

Qualified Medicare Beneficiaries such that the respondent has no out-of-pocket costs for Medicare-covered services.

C.3.3.2 Private and Medicare MCOs

MCOs (especially Medicare-contracting MCOs) often operate differently from other third-party payers and tend to have unique payment patterns. For instance, risk and (to a lesser extent) cost Medicare MCOs are paid a set fee per enrolled Medicare beneficiary (called a capitated amount) designed to compensate the MCO for the expected costs of delivering Medicare's package of benefits. There are no Medicare claims or insurance statements indicating the total charge for events covered by the capitated amount.

Often the respondent only knows the copay amount, if there was one. Also, MCOs often provide "Medigap"-type coverage by paying for most of the deductibles and copays for Medicare-covered benefits. A beneficiary who belongs to an MCO does not need private Medigap insurance or Medicaid coverage for these amounts. Thus, payment patterns for MCO beneficiaries tend to be simpler than those for FFS beneficiaries. The set of analytic edits for MCOs attempted to account for these simplified patterns and for the respondent's usual inability to report charges and payments for events. The MCO edits also attempted to avoid creating "illogical" payment patterns.

C.3.3.3 General Edits

At the beginning of the analytic editing, and after each main section of edits, an attempt was made to resolve events through addition or subtraction. Events without a known total charge but with a complete payment vector (i.e., each payer was identified as either having paid or not paid for an event and each payer's amount was known) were completed by summing across all payment sources to derive the total charge. Events with a known total charge and complete except for one missing payment amount or payment source were completed by subtraction. The excess of charges over known payment amounts was attributed to the known payer, or the one missing payer indicator was set to 1 and the excess allocated to that payer. If a service was provided free of charge, all payer indicators and payment amounts were set to 0.²⁰

If a source was a potential payer for an event, or if the respondent reported that the payer had contributed to an event but did not know the amount, it was assumed that the payer was not actually a source if the current sum of reported payments equaled the reported total charge.

C.3.4 Setting SOP Flags

Each beneficiary's health insurance timeline, survey-reported events and Medicare claims were used to establish an indicator variable (SOP flag) for each of the source of payment (SOP) categories. Information in the SOP flags was, in turn, used to determine the corresponding payer indicator variables, which were used in imputation to determine whether or not a possible source of payment actually paid something toward the cost of an event.

This section outlines the rules that applied to the process of setting the values of the SOP flags. SOP flags can have one of five possible values:

- 0 - Source definitely did not pay
- 1 - Source definitely did pay, known amount
- 2 - Source definitely did pay, unknown amount
- 3 - Source possibly paid, beneficiary was covered at time of event

²⁰ If the event was reported as free, but the respondent had also reported that a source other than Medicare or Medicaid had paid something for the event, the total charge was set to missing and imputed.

- 4 - Source possibly paid, beneficiary may have been covered at time of event

C.3.4.1 SOP Medicare

Medicare Part A and Part B entitlement dates established the period of Medicare coverage.

1. If the sample beneficiary was entitled to Medicare Part A benefits, Medicare was a potential source of payment for: Inpatient hospital – IP events and SNF – IU events. The initial value of the Medicare SOP flag was 3 (possible payer) for these event types.
2. If the sample beneficiary was entitled to Medicare Part B benefits, Medicare was a potential source of payment for: Outpatient hospital – OP events and Part B Physician/Supplier services – DU, VU, HU, ER, HP, HF, MP, SD, SL, and OM events. The initial value of the Medicare SOP flag was 3 (possible payer) for these event types.

C.3.4.2 SOP Medicaid

If either the respondent or CMS administrative data indicated that the sample beneficiary had Medicaid coverage, the Medicaid SOP flag was initially set to 3 for all events which occurred during the period of Medicaid coverage.

C.3.4.3 SOP Managed care

The managed care flag was set based on information in the beneficiary's health insurance timeline and CMS' administrative records of managed care enrollments.

1. If CMS administrative records indicated that the beneficiary was enrolled in a Medicare managed care plan, but the beneficiary did not report the enrollment, the Managed care SOP flag was initialized to a value of 4 for all events that occurred during the beneficiary's enrollment.
2. The MCO SOP flag was set to 4, for all events except DU, VU, HU, MP, and PM, if the health insurance section shows that the beneficiary was in an MCO, whether or not it is a Medicare MCO.
3. For DU, VU, HU, and MP events where HMOASSOC and HMOREFER are applicable, if either HMOASSOC or HMOREFER = 1, the MCO SOP flag was set to 3 (possible payer, coverage definite); otherwise, if the respondent answered don't know or did not respond to either HMOASSOC or HMOREFER, the MCO SOP flag was set to 4 (possible payer, coverage not definite); else the MCO SOP flag was set to 0 (MCO did not pay).
4. For PM events, the MCO SOP flag was initialized to 3 if the respondent indicated that the MCO covers dental services, otherwise the MCO SOP flag was initialized to 4.

C.3.4.4 SOP Veterans Administration

Beginning with the 2016 data year, the VA SOP flag was combined with the "other public insurance" source of payment flag. Please see section C.3.4.7 "Other public insurance" for a description of that source of payment.

C.3.4.5 SOP Private health insurance

Employer based information about private health insurance (PHI) as a payment source was provided in the insurance section of the interview, by the respondent, and through insurance statements. Information about the source of the policy (used to differentiate between employer-sponsored and individually purchased private health insurance) was also provided by the respondent in the insurance section of the interview.

1. The employer-sponsored PHI SOP flag was set to 3 for all types of services, except prescribed medicines, which occurred while the sample beneficiary was covered by employer-sponsored health insurance, based on the health insurance timeline and the date of the event.

2. For prescribed medicines, employer-sponsored health insurance was considered a possible source of payment (initial value SOP=3) if the respondent said that the plan covered drugs. If the respondent said that the plan did not cover drugs but reported a specific amount the plan paid for another PM event, the employer-sponsored PHI SOP flag for all PM events during the same round was set to 4.
3. If the event date was missing or ambiguous and the sample beneficiary's insurance coverage changed during the round, the employer-sponsored PHI SOP flag was set to 4 instead of 3 where applicable.

Information about individually purchased private health insurance as a payment source was provided in the insurance section of the interview, by the respondent, and through insurance statements. Information about the source of the policy (used to differentiate between employer-sponsored and individually purchased private health insurance) was also provided by the respondent in the insurance section of the interview.

1. The Individually Purchased PHI SOP flag was set to 3 for all event types, except prescription medicines, which occurred while the sample beneficiary was covered by individually purchased private health insurance, based on the beneficiary's health insurance timeline and the date of the event.
2. For prescription medicines, the Individually Purchased PHI SOP flag was set to 3 if the respondent reported that the individually purchased PHI plan covered drugs. If the respondent said the plan did not cover drugs but reported a specific amount the plan paid for another prescription medicine, the Individually Purchased PHI SOP flag was set to 4 for all prescription medicines reported in the same round.
3. If the event date was missing or ambiguous, and the sample beneficiary's insurance coverage changed during the round, the Individually Purchased PHI SOP flag was set to 4 instead of 3 where applicable.

C.3.4.6 SOP Out-of-pocket and SOP uncollected liability

The out-of-pocket and uncollected liability flags were not set based on health insurance timelines. In many cases, these two categories could not be ruled out as payers based on the health insurance timeline, or even after the claims match.

C.3.4.7 SOP Other public insurance

Beginning with the 2016 data year, the VA SOP flag was combined with the "other public insurance" source of payment. The following describes how the other public insurance source of payment is determined based on VA and other public insurance information provided in the interview by the respondent.

1. For all event types except prescription medicines, if the respondent indicated that the service was provided by a VA hospital or clinic, or if the respondent reported coverage by "other public insurance", then the other public insurance SOP flag was set to 3. If the respondent was not certain that the service was provided by the VA, or if the respondent was uncertain of having "other public insurance" coverage, then the other public insurance SOP flag was set to 4.
2. For prescription medicines, the other public insurance SOP flag was set to 4 if the VA or "other public insurance" paid a known amount for some other drug in the same round.

C.3.4.8 Updating SOP flags using survey-collected cost data

The initial values of the SOP flags were updated when survey-collected cost data provided more definitive information. If the respondent reported the amount the payer paid, the appropriate SOP flag was set to 1. If the respondent did not know how much the payer paid, the SOP flag was set to 2.

C.3.4.9 Updating SOP flags based upon matching Medicare claims data

The initial values of the SOP flags were also updated when the utilization could be linked to Medicare claims records.

The Medicare payment amount and the Medicare SOP flag were updated if the survey-reported utilization matched Medicare claims data, or if the Medicare claims data provided the only record of the utilization. If the Medicare claims record showed a positive, non-zero Medicare payment, the Medicare SOP flag was set to 1, to show that the payment amount was known and would not have to be imputed. If the claims record showed that the sample beneficiary's Medicare benefits were exhausted, the Medicare SOP flag was set to 1, and the Medicare payment amount was set to \$0.00. If the claims record indicated that the service was not a Medicare covered service or that the beneficiary did not have Medicare coverage for the service, both the Medicare payment amount and the Medicare SOP flag were set to zero.

If the claims record showed that Medicare was a secondary payer, the appropriate SOP flag for the primary payer was set to 1 (identifying the insurer as the primary payer), and the Medicare claim was used to develop the amount paid by the private insurer.

C.3.4.10 Updating SOP flags based on unmatched "survey only" utilization

The Medicare SOP flag was set to zero for all unmatched survey events unless the Medicare SOP flag already had a value of 1 or 2.

C.3.4.11 Resolving Conflict between Person Level Survey Reported Health Insurance information and Event Specific Survey Reported Source of Payment Data

For a very limited number of events (less than .5%) the reported source of payment data conflicted with the individuals' reported health insurance information. In these situations, the payment data was manipulated to conform to the health insurance data for the following payers: Medicaid, Employer Sponsored Health Insurance, managed care coverage, and Medigap insurance. Since Out-of-pocket and Uncollected Liability are always potential payers, there were no inconsistencies for these payer categories. Inconsistencies in Medicare SOP data were not resolved, but unmatched survey reported events with Medicare payment and no MA payment were excluded from the payment summaries' adjusted totals (see section 3.1.4).

C.4 Changes to Note in Prior Data Years

2019

Two new segments, Hearing Utilization Events (HUE) and Vision Utilization Events (VUE), were added in 2019 with data on survey-reported vision and hearing events.

2018

Variables PMCOND, PMKNWNM, TABSADAY, and TABTAKE were added to the Prescribed Medicine Events (PME) segment in 2018.

STRNNUM2 and STRNUNI2 are no longer included in the Community Questionnaire and do not appear in the 2018 Cost Supplement File.

2017

In order to provide consistency among event segments several changes were made to align common variables and eliminate redundant, information.

Exhibit C.4.1: Changes to Align Variables

Segment	Added	Dropped	Renamed
DUE	DVFLURID, DVSEALNT, DVINLAYS, DVRECALL, DVIMPLNT, DVABSESS, DVDENTRS, DVREPAIR, DVTMJ		EVBEgxx to D_BEgxx
IPE		COINDAY, LRDAYs	EVBEgxx to D_BEgxx EVENDxx to D_ENDxx
IUE		E1DGNSCD	EVBEgxx to D_BEgxx EVENDxx to D_ENDxx
MPE			EVBEgxx to D_BEgxx
OPE	D_BEgYY, D_BEgMM, D_ENDYY, PROV, STATUS	FROMDT, THRUdT EVBEgYY, EVBEgMM, EVBEgDD, E1DGNSCD	

2016

For the 2016 data year, it was determined that payments made by the VA could not be estimated with sufficient accuracy. Therefore, observed payments from the VA have been combined into the "other public insurance" source of payment beginning in 2016.

2013

For historical reference, please refer to the following changes noted in the 2013 data year.

Mandatory Payment Reductions in the Medicare FFS Program, Sequestration, took effect in 2013. Additional details can be found at: <https://www.cms.gov/outreach-and-education/outreach/ffsprovpartprog/downloads/2013-03-08-standalone.pdf>.

Medicare FFS claims records will show the reduction, which is reflected in the Medicare payment amount. The PDE claims do not have sequestration amounts on them. The reduction was taken off of the capitated payments to the providers and will not be reflected in the PDE claim records.

APPENDIX D: TABLE OF LINKS TO MCBS DOCUMENTATION

MCBS Resources	Links
CMS MCBS website	https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS
MCBS LDS file information	https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA - NewLDS
MCBS PUF	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/MCBS-Public-Use-File/index
CMS Chronic Conditions Warehouse (CCW)	https://www.ccwdata.org/web/guest/home/
Data User's Guides, Methodology Reports, Codebooks, and LDS Variable Crosswalks	https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Codebooks
Chartbook	https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Tables
Early Look, Data Briefs, Infographics, and Tutorials	https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Data-Briefs
Bibliography	https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Bibliography
Questionnaires and Questionnaire User Documentation	https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/MCBS/Questionnaires
MCBS Interactives – COVID-19 Data Tool and Survey File PUF Data Tool ²¹	https://mcbs-interactives.norc.org/

²¹ The MCBS Interactives consist of two data tools, the Survey File PUF Data Tool and the COVID-19 Data Tool. Each tool contains multiple interactive dashboards that allow users to sort and visualize data according to a variety of demographic and health-related factors.

APPENDIX E: 2020 MCBS Cost Supplement File Segments and Historic RIC Segments

Cost Supplement Segment	Abbrev	Historic RIC Segment
Dental Utilization Events	DUE	DUE
Facility Events	FAE	FAE
Hearing Utilization Events	HUE	N/A
Inpatient Hospital Events	IPE	IPE
Institutional Events	IUE	IUE
Medical Provider Events	MPE	MPE
Outpatient Hospital Events	OPE	OPE
Prescribed Medicine Events	PME	PME
Vision Utilization Events	VUE	N/A
Person Summary	PS	PS
Service Summary	SS	SS
Cost Supplement Ever Enrolled Weights	CSEVWGTS	X
Cost Supplement Longitudinal Weights	CSL2WGTS CSL3WGTS	N/A