



Data to Inform Research on Integrated Care for Dual Eligibles

Summary Report

Lacey Hartman and Elizabeth Lukanen

Prepared for: Arnold Ventures



Introduction

The importance of improving care for individuals enrolled in both Medicare and Medicaid (dual eligibles) has received considerable attention in recent years. Extensive research has shown that dual eligibles account for a disproportionate share of spending within both programs, and the lack of integration between the two programs contributes not only to excess costs but also lower quality care.^{1,2,3} As a result, states and the federal government have developed a variety of models aimed at improving integration of care for dual eligibles.

The need for better integration between Medicare and Medicaid has been explored in-depth in a wide range of publications.^{4,5,6} There is also literature describing the various models in place that seek to better integrate care and financing for this population and recommendations for accelerating positive outcomes under these programs.⁷ Missing from the discourse is a systematic assessment of data sources that can be used to study the dual eligible population. With support from the Arnold Ventures Foundation, we seek to fill this gap by summarizing the data that are available to researchers seeking to better understand the dual eligible population and the impacts of policies aimed at integrating care.

The purpose of this report is to describe the strengths and limitations of existing data sources that could be used to study the broad topic of integrated care for dual eligibles. A primary focus is whether these data can address research questions aimed at improving care for dual eligibles. The key questions of interest are built from Arnold Ventures' research agenda, which was developed in collaboration with a broad range of experts.⁸ They include:

- To what extent do available data allow for comparing outcomes for key subpopulations of dual eligibles, such as race/ethnicity, geography, those with chronic conditions, behavioral health care needs, and/or disabilities?
- To what extent do available data support understanding the utilization, cost, and quality of long-term services and supports (LTSS)?
- To what extent do available data support understanding dual eligibles' social needs, such as housing, food insecurity, and financial stability?

This report starts with a brief description of our approach to the data review and assessment. Next, we summarize findings by assessing the data sources across four areas critical to research aimed at dual eligibles: 1) enrollment in integrated care models; 2) analysis of priority subpopulations; 3) LTSS; and 4) enrollee social needs.

Within each section, we start by offering some high-level research questions relevant to the topic and summarize key findings from the data scan. We then provide an overview of important gaps and deficiencies within the data and their implications for researchers seeking to evaluate the impact of integrated care for dual eligibles. We conclude with a set of recommendations aimed at addressing these gaps and advancing the availability of comprehensive, high-quality data for research in this area.

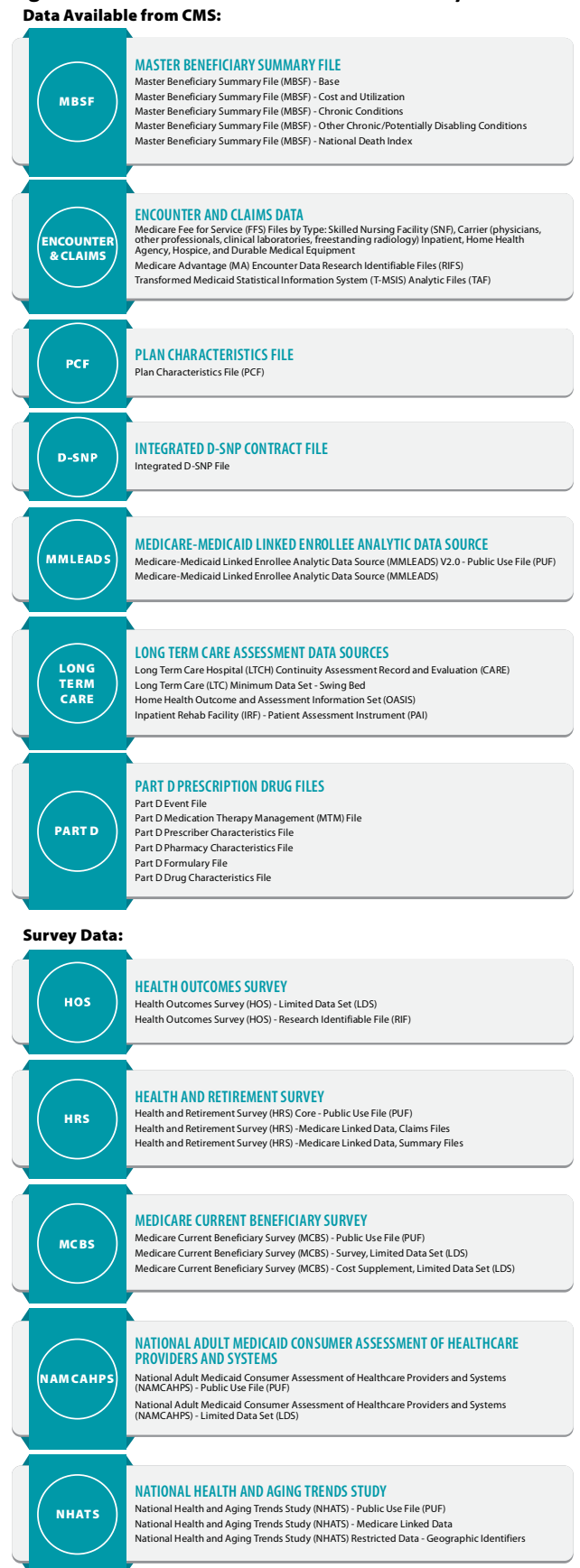
In addition to this report, SHADAC has produced a [companion Excel table](#) that contains the full abstraction details for each data source.

Overview of Data Assessment

This assessment included a systematic review of 34 data sources that can produce nationally representative estimates for dually eligible individuals (and/or estimates for all 50 states) and have the potential to inform research on integrated care models. Data sources were selected based on SHADAC’s extensive knowledge of a wide variety of data sources, our recent work developing an [inventory of evaluations](#) of integrated care models, and a scan of relevant and widely used data repositories such as the Research Data Assistance Center ([ResDAC](#)) and the Chronic Condition Data Warehouse (CCW). (See the sidebar for more information about ResDAC and the CCW). This review did not include broad population surveys such as the American Community Survey (ACS) or National Health Interview Survey (NHIS), because they provide limited value in studying dual eligibles due to both content and population focus. The review was also limited to data sources that had sufficient public-facing documentation to support our assessment. This review does not represent the full universe of data sources that could be used to evaluate integrated care for dual eligibles, but rather those that met our criteria for inclusion.

Figure 1 includes the full list of data sources included in the inventory. As the figure demonstrates, 22 of the sources are available from the Centers for Medicare & Medicaid Services (CMS) and 13 represent data from surveys. Most of the data sources in this inventory are associated with at least one other source and are grouped accordingly. In some cases, the grouped data sources contain associated, but distinct, content. For example, all of the Master Beneficiary Summary File (MBSF) data sources contain information about the Medicare beneficiary, but the data in the MBSF Base is different from what is contained in the MBSF cost and utilization file, the MBSF death file, and so on. On the other hand, the surveys included in our review (National Health and Aging Trends Study [NHATS], Medicare Current Beneficiary Survey [MCBS], Health and Retirement Study [HRS], Nationwide Adult Medicaid Consumer Assessment of Healthcare Providers and Systems [NAMCAHPS], and the Health Outcomes Survey [HOS]) are all associated with multiple data sources, which include different levels of restriction (e.g., public use files [PUFs] vs. data that require special permissions to access) and/or files that link to other sources, such as Medicare claims. For example, the NHATS includes a PUF, a restricted file with geographic identifiers, and a restricted data set that links the survey data to Medicare claims.

Figure 1. Data Sources Included in Inventory



For each source of data, we compiled technical documentation and systematically abstracted information across several fields including: summary description, content focus, standardized data type, unit of analysis, and timing (years of available data and release frequency). We documented how the data are collected, by whom, and how the data are made available for researchers (e.g., release of PUFs, application process, costs of restricted files, etc.). Finally, we summarized available data linkages and the availability of information for our four focus areas: enrollment in integrated programs, the ability to identify key subpopulations, information related to beneficiary social needs, and LTSS. The appendix Figure A-1 on page 18 contains a summary of key information from the [companion Excel table](#), including a brief description, data type, unit of analysis, most recent data available, availability of a public use file (PUF), the lowest level of geography supported by each data source, and whether or not the data supports the following analysis: trend (3 years or more), enrollment in integrated care models, and priority subpopulations.

Throughout this report, we assess the ability to link data. Our summary of this capacity assumes that researchers have obtained relevant data from the CCW, which has already conducted preliminary work to facilitate linkage at the beneficiary level. The CCW is the most practical and efficient path for researchers to access these data. As we discuss in more detail in the following sections, however, obtaining and linking data through the CCW still requires a significant investment in time and resources.

It is also important to note that our work did not include a comprehensive review of the quality of each data source. However, we do discuss some of the known, high-level limitations in our assessment of the gaps and strengths of available data in this document. In addition, we provide high-level information about limitations in the [companion Excel tool](#), along with links to more in-depth information about known data quality issues for each data source, where available.

Integrated Care Models

One of the critical aspects of leveraging data sources to improve care for dual eligibles is the ability to identify beneficiaries enrolled in integrated care models. It is also essential to understand which data provide access or linkage to Medicare and/or Medicaid cost and utilization information. The federal government and states are using a variety of programs to better integrate care for dual eligibles; Figure 2 describes the models relevant to this data review. Although evaluations are underway, the body of evidence about their effects is limited and there are important gaps in the research, such as model comparisons to assess which were the most effective at integrating care and reducing spending.⁹ By leveraging data sources that identify enrollment in integrated care models and provide information on Medicare and Medicaid cost and utilization, researchers can seek to address a variety of research questions, including:

Critical Resources for Data Access and Use

Research Data Assistance Center (ResDAC)

Established in 1996, the Research Data Assistance Center (ResDAC) is a Centers for Medicare & Medicaid Services (CMS) contractor that provides free assistance to researchers interested in the CMS data. ResDAC provides researchers with technical assistance in obtaining and analyzing data, including request procedures, supporting documentation (such as record layouts and SAS input statements) and workshops on how to use the data.

CMS Chronic Conditions Data Warehouse (CCW)

The CMS Chronic Conditions Data Warehouse (CCW) provides researchers with Medicare and Medicaid beneficiary, claims, and assessment data that can be linked by beneficiary across the continuum of care. In the past, researchers analyzing data files were required to perform extensive analysis related to beneficiary matching and de-duplication in preparation for their study analysis. With the CCW data, this preliminary work to facilitate linkage is already accomplished and delivered as part of the data files sent to researchers. To request CCW data files, researchers should start by contacting ResDAC.

Virtual Research Data Center (VRDC): The CCW provides researchers with access to the VRDC, which is a secure environment that allows researchers to access and perform their own analysis and manipulation of CMS data virtually from their own workstation. Information about the VRDC is available [here](#).

- What impacts do integrated care models have on spending and utilization for those enrolled? Do impacts diminish or increase over time?
- How do outcomes related to functional status, cognitive limitations, and limitations in activities of daily living compare between people enrolled in integrated care models and those receiving standard Medicare and Medicaid benefits?
- How do integrated programs compare to one another? Is one type of integrated model more effective than another?

Figure 2. Evaluation of Integrated Care Models and Descriptions

Integrated Model	Year of Implementation	Description
Financial Alignment Initiative (FAI) Medicare-Medicaid Plans	2012	The goal of the FAI is for CMS and States to better coordinate the financing of Medicare and Medicaid and to integrate primary, acute, behavioral, and long-term services and supports for individuals who are dually eligible for Medicare and Medicaid.
Program for All Inclusive Care for the Elderly (PACE)	1990	PACE utilizes an interdisciplinary team to integrate medical and social services to individuals who are frail, elderly, and live in a community setting. Most of these individuals are dually eligible beneficiaries. PACE enrollees must be age 55 or older, eligible for nursing home services, live within a PACE organization’s service area, and have the ability to live safely in a community setting. PACE utilizes a capped financing model that permits health care providers to administer all needed services to beneficiaries as opposed to only services that are reimbursed by Medicare and Medicaid.
Dual Eligible Special Needs Plans (D-SNPs) aligned with MLTSS	2006	Medicare Advantage Dual Eligible Special Needs Plans (D-SNPs) are plans that enroll individuals who are dually eligible for Medicare and Medicaid, with specific eligibility requirements differing slightly by state. These programs integrate care by combining Medicare and Medicaid services. Some states required D-SNPs to be aligned with MLTSS even prior to requirements put in place at the federal level through the Bipartisan Budget Act (BBA) of 2018 (see more on the BBA below).
Fully Integrated Dual Eligible Special Needs Plans (FIDE-SNP) prior to Balanced Budget Act(BBA) of 2018	2010	FIDE-SNPs are D-SNPs that meet several criteria for integrating care across the two programs. Prior to the BBA of 2018, FIDE-SNPs needed to meet the following criteria: use a single Managed Care Organization to provide Medicare and Medicaid services to dually eligible beneficiaries; have a contract with a State Medicaid Agency that provides coverage for specific primary, acute, and long-term care services; integrate Medicare and Medicaid services with care management and specialty care network models.
FIDE-SNPs and Highly Integrated Dual Eligible (HIDE)-SNPs post-Bipartisan Budget Act (BBA) of 2018	2021	The Bipartisan Budget Act (BBA) of 2018 permanently authorized D-SNPs and set new requirements for the way they must coordinate with the Medicaid program. CMS then via regulation outlined three levels of D-SNPs from most to least integrated: FIDE-SNPs, HIDE-SNPs, and D-SNPs. ¹⁰ The main distinction between FIDE-SNPs and HIDE-SNPs is the existence of a capitated Medicaid contract for long-term services and supports by the same entity that contracts with CMS to operate as an MA plan. ¹¹ D-SNPs that do not qualify as FIDE-SNPs or HIDE-SNPs must notify states of hospital or skilled nursing facility (SNF) admissions for some high-risk individuals.
State-Specific Programs	Varies	Includes programs that pre-date the FAI and/or are run concurrently as separate programs under the FAI: MN Senior Health Options, MA Senior Care Options, and the WI Partnership Program.

As Figure 2 describes, the definitions and requirements for FIDE-SNP plans were altered under the BBA of 2018, with regulations enacting these changes going into place in 2019, and full implementation in state contracts in January of 2021. Prior to 2019, FIDE-SNPs can be clearly identified using the Plan Characteristics file (discussed below).

Figure 3 summarizes the data sources included in our review and whether they can be used to identify enrollment in integrated models. Also flagged is the availability of Medicare and/or Medicaid cost and utilization data.

Figure 3. Data Sources that Can Identify Enrollment in Integrated Care Models

Data Source	KEY: ● - Yes ● - Not beneficiary level data		Requires Link to MBFS Base to Identify Enrollment in Integrated Care Model				
	Medicare Cost or Utilization	Medicaid Cost or Utilization	FAI	PACE	FIDE-SNP pre-BBA ¹	FIDE-SNP/HIDE-SNPs post BBA ²	State-specific Programs
MBSF							
MBSF - Cost and Utilization	●		●	●	●		●
MBSF - Chronic Conditions			●	●	●		●
MBSF - Other Chronic/Potentially Disabling Conditions			●	●	●		●
MBSF - National Death Index			●	●	●	●	●
ENCOUNTER & CLAIMS							
Medicare FFS Claims Files by Type	●		●	●	●	●	●
MA Encounter Data RIFS	●		●	●	●	●	●
T-MSIS TAF		●	●	●	●	●	●
PLAN CHARACTERISTICS FILE							
Plan Characteristics File			●	●	●		●
INTEGRATED D-SNP CONTRACT FILE							
Integrated D-SNP File						●	
MMLEADS							
MMLEADS V2.0 - PUF	●	●					
MMLEADS ³	●	●		●	●	●	●
LONG TERM CARE							
LTCH – CARE	●		●	●	●	●	●
LTC Minimum Data Set - Swing Bed	●		●	●	●	●	●
Home Health OASIS	●		●	●	●	●	●
IRF-PAI	●		●	●	●	●	●
PART D							
Part D Event File	●		●	●	●	●	●
Part D MTM file	●		●	●	●	●	●
Part D Prescriber Characteristics File	●		●	●	●	●	●
Part D Pharmacy Characteristics File	●		●	●	●	●	●
Part D Formulary File	●		●	●	●	●	●
Part D Drug Characteristics File	●		●	●	●	●	●
HOS							
HOS – LDS ^{4,5}			●	●	●	●	●
HOS – RIF ^{4,5}			●	●	●	●	●
HRS							
HRS – PUF							
HRS - Medicare Linked Data, Claims Files ⁵	●		●	●	●	●	●
HRS - Medicare Linked Data, Summary Files ⁵	●		●	●	●	●	●
MCBS							
MCBS – PUF							
MCBS – Survey, LDS	●		●	●	●	●	●
MCBS - Cost Supplement, LDS	●	●	●	●	●	●	●
NAMCAHPS							
NAMCAHPS – PUF							
NAMCAHPS – LDS							
NHATS							
NHATS – PUF							
NHATS - Medicare Linked data ⁴	●		●	●	●	●	●
NHATS Restricted Data - Geographic identifiers							

1 - To identify enrollment in Fully Integrated Dual Eligible Special Needs Plans (FIDE-SNP), the data source must also be linked to the Plan Characteristics file. The Plan Characteristics file is not beneficiary level, but can be used to link back to the MBSF to obtain beneficiary level enrollment in the other integrated models. While MLTSS is not directly identified through linking to the MBSF or Plan Characteristics file, researchers could review state contracts to determine which plans meet the relevant criteria, and potentially link the plan data back to the Plan Characteristics and MBSF Base files to identify beneficiaries.

2-Beginning in 2021, researchers can use the D-SNP Integrated Status file to identify HIDE-SNPs and D-SNPs that must meet requirements set out in regulation in 2019 for integration between Medicare, Medicaid, behavioral health, and LTSS. We anticipate that linkage to identify these plan types will be possible as future years of data are available (none of the data released at the time of publishing would include information for 2021 activity). However, there may be issues with FIDE-SNP designations in 2019 and 2020 in other files.

3 - Includes a variable that identifies enrollment in PACE (no linkage required).

4 - Includes a variable that identifies enrollment in FIDE-SNP (no linkage required).

5- Select integrated care models can be identified, but sufficiency of sample size to evaluate specific programs is unknown.

As shown in the figure above, the vast majority of the data sources we reviewed include information about Medicare cost or utilization, while relatively few have information on Medicaid. Only the Medicare-Medicaid Linked Enrollee Analytic Data Source (MMLEADS), Medicare Current Beneficiary Cost Supplement (MCBS-Cost Supplement, LDS) and Transformed Medicaid Statistical Information System Analytic Files (T-MSIS-TAF) capture information about Medicaid, and only MMLEADS provides comprehensive information about both Medicare and Medicaid. With the exception of the NAM CAHPS, all of the restricted data sources included in our review can be linked to identify enrollment in integrated care models. (None of the PUFs can be used to identify enrollment in these models.)

As the table demonstrates, linkage to the Medicare Beneficiary Summary File-Base File (MBSF Base) is required to identify enrollment in all of the listed integrated care models. In order to identify a particularly integrated form of Special Needs Plan—the Fully Integrated Dual Eligible Special Needs Plan (FIDE-SNP) prior to 2021—data must be linked to both the MBSF and the Plan Characteristics File (see sidebar for information about these files). Beginning in 2021, researchers can use the integrated D-SNP Contract file to link information about the level of integration of D-SNPs by contract number back to the MBSF (see sidebar for more information). The shift in definitional requirements starting in 2019 (described above) may complicate the ability for researchers to clearly identify FIDE-SNP plans during the transition years of 2019 and 2020.

An important consideration for researchers seeking to evaluate the impact of integrated models on care for dual eligibles is the timing and frequency of available data in relation to the date of model implementation. Figure 4 provides an overview of the years of data available for all sources included in our review, the frequency of data release, and the implementation dates of the four key integrated care programs considered in our assessment. As the figure demonstrates, the majority of data sources have at least three years of data following the implementation of all integrated care models, except the recently enacted changes to D-SNP definitions. There are three important exceptions: the MMLEADS, which has not been updated since 2012, the NAM CAHPS, which was fielded once from 2014–2015, and the D-SNP contract file, which just became available. We have also provided information about the anticipated timing (first or second half of 2021) of the next data release, where available.

Data Resources for Identifying Enrollment in Integrated Care Models

Identifying enrollment in integrated care models requires linkage to one or more data sources: the MBSF Base, the Plan Characteristics File, and the Integrated D-SNP Contract File. Below we provide more information about these data sources and the specific fields used for identifying integrated models and linkage.

Master Beneficiary Summary File - Base (MBSF Base): This file includes information for all Medicare beneficiaries, including monthly enrollment by plan type, and details about the plan, such as whether it is participating in one of the integrated care models for dual eligibles. Information about enrollment in PACE, a Medicare-Medicaid Plan, and state-specific integrated care models is contained in the monthly indicators for Part C enrollment (PTC_PLAN_TYPE_CD_01-12). Considerable detail related to entitlement and demographics are also available. This file can be linked to a wide range of data sources by the CCW-generated beneficiary identification number (BENE_ID), and to the Plan Characteristics and Integrated D-SNP Contract File (discussed below) by contract identification number (CONTRACT_ID). Because of the broad ability to link to other sources and the range of critical information contained in the file, the MBSF Base is key to researching integrated care for dual eligibles. The data are available from 1999–2019, and updated annually in conjunction with the claim files. The MBSF files are restricted, and must be approved for use by CMS. More detail about the MBSF Base and associated MBSF is available in the [companion Excel table](#).

Plan Characteristics File: The Plan Characteristics file contains Medicare Advantage plan and Prescription Drug Plan information and is separated into six subfiles: the plan "base" file, premium file, cost sharing tier file, service area file, special needs plans file, and multi-year crosswalk file. Both the fields needed to link to the MBSF Base (CONTRACT_ID) and the field that identifies whether a plan qualifies as a FIDE-SNP (FIDE_SNP) are included in the base file. More detail about the Plan Characteristics File is available in the [companion Excel table](#).

Integrated D-SNP Contract File: Beginning in 2021, researchers can use the Integrated D-SNP Contract file to link information about the level of integration of D-SNPs by contract number back to the MBSF. The file contains information for each plan, the state of operation, and whether the plan qualifies as a FIDE-SNP, HIDE-SNP, or "coordination-only" plan. The data are based on CMS review of State Medicaid Agency Contracts (SMACs) submitted in 2020.

Priority Subpopulations

A critical consideration for evaluating the impact of integrated care is how and whether enrollment and key outcomes differ for various priority subpopulations of dual eligibles. SHADAC’s recent review of evaluations of integrated care models for dually eligible populations found that more research is needed in order to assess the impact of models on specific subpopulations.¹² Key research questions that could be addressed with data sources that identify enrollment in integrated models and information for priority subpopulations include:

- Do integrated models work better for some types of beneficiaries than others?
- Are there differential impacts of integrated care for beneficiaries that are high cost? For beneficiaries with chronic conditions and/or behavioral health needs?
- Is there evidence of disparities by race/ethnicity and/or urban/rural status in enrollment and benefits of integrated models?
- Do integrated programs have any impact on partial dual eligibles (i.e., those only eligible to get some assistance with paying for premiums and cost sharing, but not eligible for additional Medicaid benefits)?
- Is there evidence of differential impacts of integrated care models across states?

In our scan, we reviewed documentation such as code books to assess the research viability for the following populations of interest: people age 65 and older, specific racial and ethnic groups, beneficiaries with full versus partial dual eligible benefits, high-cost beneficiaries, rural and urban beneficiaries, and individuals with chronic conditions, such as a behavioral health condition, or a disability. We did not conduct a sample size analysis to determine whether sample was sufficient to support all forms of analysis, because that was outside the scope of this project.

Figure 5 crosswalks the data sources included in our review with the ability to identify these key subpopulations, including the ability to produce estimates at the state level. For context, the figure also indicates whether the data source can identify enrollment in integrated care models. Also, as noted, the majority of data sources require linkage to identify some or all of the listed subpopulations.

Figure 5. Data Sources that Identify Priority Subpopulations¹

Data Sources	Link to Identify Integrated Models	Geography		Over/Under 64 Years	Race/Ethnicity	Full /Partial Dual Status	High Cost	Chronic Conditions	Behavioral Health	Disability
		State	Urban/Rural							
MBSF										
MBSF – Base	●	●	●	●	●	●	●	●	●	●
MBSF - Cost and Utilization	●	●	●	●	●	●	●	●	●	●
MBSF - Chronic Conditions	●	●	●	●	●	●	●	●	●	●
MBSF - Other Chronic/Potentially Disabling Conditions	●	●	●	●	●	●	●	●	●	●
MBSF - National Death Index	●	●	●	●	●	●	●	●	●	●
ENCOUNTER & CLAIMS										
Medicare FFS Claims Files by Type	●	●	●	●	●	●	●	●	●	●
MA Encounter Data RIFS	●	●	●	●	●	●	●	●	●	●
T-MSIS TAF ¹	●	●	●	●	●	●	●	●	●	●
PLAN CHARACTERISTICS FILE										
Plan Characteristics File	●	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
INTEGRATED D-SNP CONTRACT FILE										
Integrated D-SNP File	●	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MMLEADS										
MMLEADS V2.0 – PUF		●								
MMLEADS ²	●	●	●	●	●	●	●	●	●	●
LONG TERM CARE										
LTCH – CARE	●	●	●	●	●	●	●	●	●	●
LTC Minimum Data Set - Swing Bed	●	●	●	●	●	●	●	●	●	●
Home Health OASIS	●	●	●	●	●	●	●	●	●	●
IRF-PAI	●	●	●	●	●	●	●	●	●	●

Data Sources	KEY: ● - Yes ● - Yes but requires linkage ● - Not beneficiary level data										
	Link to Identify Integrated Models	Geography		Over/Under 64 Years	Race/Ethnicity	Full /Partial Dual Status	High Cost	Chronic Conditions	Behavioral Health	Disability	
		State	Urban/Rural								
PART D											
Part D Event File	●	●	●	●	●	●	●	●	●	●	●
Part D MTM file	●	●	●	●	●	●	●	●	●	●	●
Part D Prescriber Characteristics File	●	●	●	●	●	●	●	●	●	●	●
Part D Pharmacy Characteristics File	●	●	●	●	●	●	●	●	●	●	●
Part D Formulary File	●	●	●	●	●	●	●	●	●	●	●
Part D Drug Characteristics File	●	●	●	●	●	●	●	●	●	●	●
HOS											
HOS – LDS ³	●	●		●	●			●	●	●	
HOS – RIF ³	●	●		●	●			●	●	●	
HRS											
HRS – PUF		●		●	●			●	●	●	
HRS - Medicare Linked Data, Claims Files ²	●	●	●	●	●	●	●	●	●	●	●
HRS - Medicare Linked Data, Summary Files ²	●	●	●	●	●	●	●	●	●	●	●
MCBS											
MCBS – PUF ⁴			●	●	●	●		●	●	●	
MCBS - Survey, LDS ^{3,4}	●		●	●	●	●		●	●	●	
MCBS - Cost Supplement, LDS ^{3,4}	●		●	●	●	●	●	●	●	●	
NAMCAHPS											
NAMCAHPS – PUF		●		●	●	●		●	●	●	
NAMCAHPS – LDS		●		●	●	●		●	●	●	
NHATS											
NHATS – PUF				●	●			●		●	
NHATS - Medicare Linked data ³	●	●	●	●	●	●	●	●	●	●	●
NHATS Restricted Data - Geographic Identifiers		●	●	●	●			●		●	

N/A - The Plan Characteristics file is used to identify FIDE-SNPs, but would not be used for subpopulation analysis because it is not beneficiary level data.

1 - The majority of data sources require linkage to identify some or all subpopulations. See companion Excel table for detail.

2 - Select integrated care models can be identified.

3 - Select integrated care models can be identified, but sufficiency of sample size to evaluate specific programs is unknown.

4 - Survey does not support state estimates, but can support estimates of urban vs. rural as a national characteristic.

As Figure 5 demonstrates, all sources except the MMLEADS PUF, the Plan Characteristics file, the D-SNP contract file, and non-beneficiary level Part D data can be used to identify subpopulations over/under 65, by race/ethnicity, one or more chronic conditions, and disability status. All except the Plan Characteristics File, the D-SNP contract file, MMLEADS-PUF, and the NHATS-PUF and NHATS Restricted file support analysis for those with one or more behavioral health concern, and all except the NHATS-PUF, and the MCBS files support estimates at the state level. There are fewer data sets (22) that support analysis of urban/rural status among beneficiaries, beneficiaries with full vs. partial dual status (23), or those that allow for the identification of high cost beneficiaries (19). Notably, among the data sets that facilitate identifying enrollment in integrated models (21) all except the HOS support analysis of all other subpopulation groups as well.

While our review found that most of the data sources assessed can identify some priority subpopulations, it is important to note that those that can identify enrollment in integrated models and allow for analysis across all priority subpopulations are either claims, encounter, or assessment data sets from CMS, or survey data that are linked to those CMS data sources. Unsurprisingly then, while the PUFs assessed in this report allow for some subpopulation analysis, it is limited.

LTSS

Some of the most critical and expensive services that dual eligibles receive are long-term services and supports (LTSS).¹³ For many, the need for these services and an inability to pay for them is what triggers dual eligibility.¹⁴ In addition, one of the goals of integrated care for dual eligibles is to address care needs within the home or community where possible, known as “rebalancing.”¹⁵ Finally, broadly understanding the cost and quality of all types of LTSS is

critical for advancing policy to provide more integrated care for dual eligibles. Key research questions that could be addressed with data containing information about LTSS include:

- Are there differences in baseline and follow-up outcomes related to physical and cognitive functioning, etc., across different integrated care models?
- Are there differences in the cost, quality, and overall utilization of LTSS for beneficiaries enrolled in integrated programs?
- Are beneficiaries enrolled in integrated programs more or less likely to receive LTSS in community based settings than those receiving traditional Medicare and Medicaid benefits?

Figure 6 identifies data sources that provide information about LTSS, including the population focus of the data as it relates to LTSS; whether the source provides information on quality, cost, and utilization in the Medicare and Medicaid programs; the setting (institutional, home health); whether the data contain or can be linked to claims; and the ability to identify enrollment in integrated models. As the table demonstrates, most of these sources contain some information about the quality of LTSS received by dual eligibles. Fewer provide information about cost or utilization, and the data sources are skewed toward providing information about LTSS delivered in institutional settings, such as nursing homes, rather than at home or in the community.

Figure 6. Data Sources that Include Information on LTSS

Data Source	LTSS Beneficiary Population	Quality and Outcomes	Medicaid Cost or Utilization	Medicare Cost or Utilization	Setting		Claims or Link to Claims	Link to Identify Integrated Models
					Institutional	Home Health		
ENCOUNTER & CLAIMS								
Medicare FFS Claims Files by Type ¹	Medicare FFS paid LTSS	●		●	●	●	●	●
MA Encounter Data RIFS ¹	MA paid LTSS	●		●	●	●	●	●
T-MSIS TAF ¹	Medicaid paid LTSS	●		●	●	●	●	●
MMLEADS								
MMLEADS	Any Medicaid or Medicare FFS paid LTSS		●	●	●	●	●	●
LONG TERM CARE								
LTCH - CARE ²	Receiving care in LTCH facility	●	●	●	●		●	●
LTC Minimum Data Set - Swing Bed ²	Receiving care in a swing bed	●	●	●	●		●	●
Home Health OASIS ²	Receiving home health services from certified agency	●	●	●		●	●	●
IRF-PAI ²	Receiving care in an IRF-PAI	●	●	●	●		●	●
PART D								
Part D MTM file ⁴	In MTM program	●			●		●	●
HRS								
HRS - Medicare Linked Data, Claims Files ^{1,5}	Medicare paid LTSS	●		●	●	●	●	●
HRS - Medicare Linked Data, Summary Files ^{1,5}	Medicare paid LTSS	●		●	●	●	●	●
MCBS								
MCBS - Survey, LDS	Medicare paid LTSS			●	●	●	●	●
MCBS – Cost Supplement, LDS ¹	Medicare paid LTSS	●	●	●	●		●	●
NHATS								
NHATS PUF ¹	Institutional LTSS	●			●		●	
NHATS - Medicare Linked data ^{1,2,3,5}	Medicare paid LTSS	●		●	●	●	●	●
NHATS Restricted Data - Geographic identifiers ³	Institutional LTSS	●			●		●	

This table is limited to data sources in the review that contained information on LTSS.

1 - Quality and outcomes: Claims/encounter based quality measures.

2 - Quality and outcomes: Assessment instrument based outcomes.

3 - Quality and outcomes: Includes physical and cognitive functioning for those in institutional long-term care facilities.

4 - Quality and outcomes: Resolution of problems with medications.

5 - Select integrated care models can be identified, but sufficiency of sample size to evaluate specific programs is unknown.

Another important takeaway from our assessment is that several sources capture aspects of LTSS for specific subpopulations—for example, the NHATS has data on LTSS for respondents that end up institutionalized at some point during the data collection and follow-up period, and the long-term care data sets capture outcomes and utilization for beneficiaries that receive care in the relevant settings. In addition, several data sources provide information about cost and utilization of these services in the Medicare program, either directly or through linkage to claims data. However, only one data source (MMLEADS) provides comprehensive information on Medicare and Medicaid utilization and costs for LTSS, and this data source has not been refreshed since 2012, which limits its utility in providing ongoing data for researchers interested in studying the longer-term impacts of integrated programs.















Enrollee Social Needs

Individuals who are dually eligible for Medicare and Medicaid often have complex social needs. Addressing these needs is critical to improving care, and can help delay use of expensive acute or long-term care.¹⁶ For these reasons, understanding the social needs of the population is critical to assessing the effectiveness of interventions and models to integrate care. The definition of “social needs” is potentially quite broad; we focused our review on the following high-level topics that emerged from the data sources included in our review: food security, housing, and financial strain. Data sources with information about enrollee social needs could be used to address a variety of research questions, including:

- How do enrollee social needs vary by key subpopulations?
- Are enrollees in integrated programs more or less likely to face unmet social needs? Is there evidence of these programs mitigating needs for these enrollees over time?
- What is the relationship between unmet social needs among enrollees and the cost and quality of acute care services?

In our review, survey data were the strongest source of information about these social needs. No claims data contained information on this topic, but one of the assessment tools (Home Health OASIS) contained some relevant content. Figure 7 summarizes the four high-level data sources (the HRS, MCBS, and NHATS are associated with multiple files) with information on social needs, along with example items under each category of need. The table also indicates whether the data source can be linked to claims information (all can) and whether it is possible to identify enrollment in integrated models (all can support this linkage).

Figure 7. Data Sources that Include Beneficiary Level Information on Social Needs

Data Source	Link to Identify Integrated Models	Food Security	Housing Conditions	Financial Strain	Link to Claims Available
 Home Health OASIS	●		 - inadequate cooling/heating - safety hazards - pests - contaminated water	 - unable to pay rent/utilities - unable to pay uncovered medical expenses - unable to pay uncovered medication expenses	●
	●	 - didn't have enough money to buy needed food - eat less because can't afford food	 - condition of home - safety of neighborhood	 - foreclosure - have too much debt right now	●
	●	 - couldn't afford balanced meals - hungry because not enough money for food - cut size or skipped meals		 - medical bills being paid off over time	●
	●	 - skipped meals because of cost - how often skipped meals	 - tripping hazards - broken windows - roof problems - pests in home	 - unable to pay for housing - unable to pay for utilities	●

This table is limited to data sources in the review that contained information about social needs. For the HRS and NHATS - Select integrated care models can be identified, but sufficiency of sample size to evaluate specific programs is unknown.

Summary Assessment and Implications for Researchers

Our review did not identify one “perfect” data source for addressing key research questions related to integrated care for dual eligibles, but we found that some sources provided greater potential than others. In Figure 8, we summarize core aspects of each data source’s performance on the topics discussed above, including: most recent years of data, accessibility (whether there is a PUF), whether the data source provides information about costs or utilization in Medicare and/or Medicaid, and the ability of the data to identify enrollment in integrated care models, to support subpopulation analysis, and to support analysis of LTSS and issues related to enrollee social needs. Below, we describe these summary findings and the impact they have on research aimed at studying integrated care for dual eligibles.

Figure 8. Summary Assessment

Data Sources	Most Recent Year of Data	Public Use File	Medicare Cost or Utilization	Medicaid Cost or Utilization	Link to Identify Integrated Models	Supports analysis of:		
						Subpops. ¹	LTSS	Enrollee Social Needs
KEY: ● - Yes ● - Not beneficiary level data								
MBSF								
MBSF – Base	2019				●	●		
MBSF - Cost and Utilization	2018		●		●	●		
MBSF - Chronic Conditions	2018				●	●		
MBSF - Other Chronic/Potentially Disabling Conditions	2018				●	●		
MBSF - National Death Index	2016				●	●		
ENCOUNTER & CLAIMS								
Medicare FFS Claims Files by Type	2019		●		●	●	●	
MA Encounter Data RIFS	2017		●		●	●	●	
T-MSIS TAF	2018 ¹			●	●	●	●	
PLAN CHARACTERISTIC FILE								
Plan Characteristics File	2019				●	N/A		
INTEGRATED D-SNP CONTRACT FILE								
Integrated D-SNP File	2021	●			●	N/A		
MMLEADS								
MMLEADS V2.0 – PUF	2012	●	●	●				
MMLEADS ¹	2012		●	●	●	●	●	
LONG TERM CARE								
LTCH – CARE	2018		●		●	●	●	
LTC Minimum Data Set - Swing Bed	2018		●		●	●	●	
Home Health OASIS	2018		●		●	●	●	●
IRF-PAI	2019		●		●	●	●	
PART D								
Part D Event File	2018		●		●	●		
Part D MTM file	2018		●		●	●	●	
Part D Prescriber Characteristics File	2019		●		●	●		
Part D Pharmacy Characteristics File	2019		●		●	●		
Part D Formulary File	2019		●		●	●		
Part D Drug Characteristics File	2019		●		●	●		
HOS								
HOS - LDS ³	2019 ²				●	●		
HOS - RIF ³	2019				●	●		
HRS								
HRS – PUF	2018	●				●		●
HRS - Medicare Linked Data, Claims Files ³	Varies ²		●		●	●	●	●
HRS - Medicare Linked Data, Summary Files ³	2015		●		●	●	●	●
MCBS								
MCBS – PUF	2017	●				●		●
MCBS - Survey, LDS	2018		●		●	●	●	●
MCBS - Cost Supplement, LDS	2018		●	●	●	●	●	●

Data Sources	Most Recent Year of Data	Public Use File	Medicare Cost or Utilization	Medicaid Cost or Utilization	Link to Identify Integrated Models	Supports analysis of:		
						Subpops. ¹	LTSS	Enrollee Social Needs
NAMCAHPS								
NAMCAHPS – PUF	2015	●				●		
NAMCAHPS – LDS	2015					●		
NHATS								
NHATS – PUF	2019	●				●	●	●
NHATS - Medicare Linked data ³	2017 ²		●		●	●	●	●
NHATS Restricted Data - Geographic identifiers	2019					●	●	●

N/A - The Plan Characteristics file is used to identify FIDE-SNPs, but would not be used for subpopulation analysis because it is not beneficiary level data.

1 - Data source identifies characteristics of key subpopulations including rural/urban, over/under 65, race/ethnicity, full/partial dual status, those with high cost, chronic conditions, behavior health diagnosis or a disability. Most data sources require linkage to identify some or all priority subpopulations, see full data inventory for details

2 - See full assessment table for details.

3 - Select integrated care models can be identified, but sufficiency of sample size to evaluate specific programs is unknown.

Data Linkage is Unavoidable

As demonstrated in the Figure 8, it is possible to identify enrollment in integrated care models with the majority of data sources included in our review; however, doing so requires linkage across multiple data sets. Such linkage is also required to support the bulk of the subpopulation analysis afforded by the data sources in our review. Although the CCW has made linkage much more straightforward, researchers must still identify the sources to be linked to address their research questions, gain access to and approval for multiple sources, and, if necessary, individually clean each data source. All of these steps must precede any actual analysis focused on research questions of interest.

In addition, none of the data sources included in our review could be used to directly identify beneficiaries in D-SNPs aligned with MLTSS prior to 2021 (starting in 2021, researchers can identify FIDE-SNPs that integrate MLTSS through the Integrated D-SNP file). Data users with research questions specific to earlier use of D-SNPs aligned with MLTSS could review state contracts to determine which plans meet the relevant criteria, and potentially link the plan data back to the plan characteristics and MBSF Base files to identify beneficiaries, but this approach would also be quite time consuming.

Researchers Cannot Rely on Public Use Data

There were only six PUFs in our review and in general these data files provide very limited utility for research on this topic. With the exception of the Integrated D-SNP File, none of the public use data files can be used to identify enrollment in integrated care models, which thereby limits these sources'

Leveraging Existing Data Resources Requires Time, Money and Expertise and Gaps Persist

It can be helpful to think about the cost and complexity of acquiring relevant data in the context of a single research question and the potential data sources available to answer it. For example, researchers may begin with the NHATS to answer the question: *How do integrated models differ in the total cost and cognitive outcomes for dual beneficiaries who spend at least part of the year in an institutional long term care setting?*

The first part of this question (cognitive outcomes) could potentially be answered with the NHATS-PUF. However, getting information about enrollment in integrated care models and total cost would require obtaining NHATS data linked to Medicare claims, which requires submitting application to both NHATS and CMS. If researchers were interested in studying a broader population of dual eligibles than those included in a survey such as the NHATS (which is likely, given that the sample of NHATS enrolled in integrated programs may not be sufficient for in-depth analysis), then the cost and time associated with obtaining a range of data files from CMS is likely to be much more substantial. For example, in order to obtain the full picture of Medicare and Medicaid utilization and costs for relevant beneficiaries, researchers would need to request the T-MSIS TAF, the MBSF Base, Part D files, and Medicare FFS and/or MA Encounter Data RIFs. All of these data files require submitting requests to CMS, and must either be obtained as separate, physical data sets to be stored in a secure IT environment provided by the researcher, or by purchasing access through the Virtual Research Data Center (VRDC). (See the full abstraction table for detailed information about the application process and costs by data source.) The time and costs associated with such a request would depend in part on the size of the relevant cohort and the number of researchers needing access (if conducting analysis in the VRDC), but regardless is considerably more resource intensive than requesting the NHATS files. Finally, by shifting the data resources beyond the NHATS, researchers would lose the panel-based information on cognitive outcomes.

overall value for research on this topic. Public use files are also much more limited in their ability to produce analysis for priority subpopulations of interest. So, while these data are easier to obtain and typically easier to use, they offer limited application to core research questions. A lack of serviceable public use data creates higher bar for entry into this research area, thus discouraging less experienced and/or under-resourced researchers from entering this field of study.

The Most Robust Data Sources are Also the Most Complex

While the restricted data sources that allow identification of enrollment in integrated care also support analysis for the majority of key subpopulations, the sources that allow for analysis across all priority subpopulations are either claims, encounter, or assessment data sets from CMS, or survey data that are linked to those CMS data sources. All of these sources or combinations of linked data sources are extremely complex. They pose the most burden in terms of the cost and time needed to access the data and the most skill and experience in terms of coding, analysis, and interpretation. All of this has implications not only for cost and time of relevant research studies, but on the type of researchers able to study the dual eligible population.

Issues with Data Quality, Power, and Comprehensiveness

Our data scan included a high-level assessment of documented data strengths and weaknesses, including quality, completeness, and other limitations. (See the [companion Excel table](#) for a detailed overview by source.) Based on our review, we found strengths and limitations differed by source and varied in importance depending on the research questions at hand. Examples of important weaknesses included a limited focus (e.g., MMLEADS lacks most information on managed care service utilization and spending), small sample size for those enrolled in integrated models (survey data sources), and insufficient years of data (NAM CAHPS). Also, as noted earlier, there are concerns that the identification of FIDE-SNPs may be difficult in 2019 and 2020, due to the transition in definitions called for in the BBA of 2018.

Two data sources that form a critical base of information for the dual eligible population—the TMSIS TAF and the Medicare Advantage RIFs—face a wide range of quality and comprehensiveness issues. For example, in the TMSIS TAF, the majority of states have incomplete managed care claims and are missing a considerable amount of race/ethnicity data.^{17,18} In the Medicare Advantage RIFs, service-level detail can vary depending on a variety of factors, including the extent to which a Medicare Advantage Organization (MAO) captures FFS-level detail in their interactions with providers.¹⁹ In addition, differences in payment structures and data collection by MAOs may contribute to some variables being populated less consistently compared with FFS claims data, and information about payment is not included in the encounter data because it is considered proprietary.²⁰ There are ongoing efforts to address the data quality and comprehensiveness issues in both data sets but, to date, significant issues such as these remain.

When taken together, many of the weakness contributed to a similar problem—they lead to an incomplete picture of a complex population that requires nuanced study. A limited focus prevents robust comparisons, small sample sizes makes it challenging to study priority subpopulations, and limited years of data prevents the verification of results over time. Many of these data quality issues limit the complexity and robustness of the research that could be conducted.

For a complete summary of the strengths and weaknesses by data source, please see the [companion Excel table](#).

Research Potential is Limited by Significant Data Gaps

Medicaid: Existing data are heavily skewed toward providing information about Medicare utilization, costs, and quality (see Figure 8). Only the MMLEADS and T-MSIS capture information about Medicaid, and only the MMLEADS provides comprehensive information about both Medicare and Medicaid, and that data source has not been updated since 2012, which significantly limits its utility given the scope of changes in integrated care for dual eligibles since that time. Timely, integrated information about Medicare and Medicaid is critical for evaluating the impact of integrated care models on dual eligibles. The programs cover different but critical aspects of care for dual eligibles; for example, Medicare covers the majority of inpatient, outpatient, and post-acute care, but Medicaid funds the majority of LTSS.²¹ A full picture of care is needed to evaluate efforts to integrate care across the programs.

Full Spectrum of LTSS: As demonstrated in the summary assessment table (Figure 8), approximately half of the 33 data sources (16) contain some information about LTSS, but the data are fragmented. Additionally, as we discuss in more detail in the LTSS section above, survey data sources only provide information for respondents who end up receiving institutional care, and the assessment data sources are limited to each specific setting (e.g., home health, swing bed, etc.) Also noted in the LTSS section above is the fact Medicaid is the primary payer for a range of institutional and community-based LTSS for people needing assistance with daily self-care, and only the TMSIS-TAF and MMLEADS contain this information. More systematic, accessible data is needed to study the impact of models that seek to integrate acute care and LTSS.

Socials Needs: As shown in Figure 8, only seven of the sources of data provide information on enrollee social needs, and only the Home Health OASIS and fully linked versions of the NHATS and HRS facilitate identification of enrollment in integrated care models. In addition, the data sources that provide information on social needs (surveys) are likely to have sample size constraints when it comes to analyzing integrated care models, or are limited in their scope (e.g., the OASIS data set only captures information for beneficiaries that use home health).

Recommendations to Improve and Expand Data Resources

The summary assessment and the implications of data gaps, quality concerns, and accessibility for researchers inform the recommendations outlined below.

Recommendation: Leverage existing data sources to make a beneficiary-level source of information on enrollment in integrated care models available to more researchers.

As we have discussed above, it is possible to identify integrated enrollment in the restricted data sources included in our review, but doing so requires linkage across multiple files. The MBSF and Integrated D-SNP files provide the bulk of the information needed, but currently the MBSF is only available through special request to CMS, and the integrated care data file must be linked back to the MBSF to identify beneficiaries enrolled in integrated D-SNPs. A potentially helpful addition to the existing data infrastructure would be a beneficiary-level file that contains all of the information about enrollment in integrated models in one place, including start and end dates of enrollment. It would also be helpful if CMS could include information about integration with MLTSS and behavioral health at the plan level in this file (currently, information about MLTSS and behavioral health is summarized at the state level only). Ideally, this data source would also provide researchers with the information needed to identify enrollment in D-SNPs that were integrated with MLTSS prior to 2019. Finally, it would be extremely helpful, and expand the number and type of researchers that could engage in this type of research, if a version of this file could be made available at limited cost.

This type of beneficiary-level, linkable data source has the potential to enhance the ability of researchers to leverage multiple data sources to understand the impacts of integrated care on dual eligibles. It would also address some of the issues mentioned above; for example, the ability to link enrollment in integrated models to all survey data sources included in the review would expand the number of data sources that provide information on social needs that could be used in research about integrated care models.

Recommendation: Maintain an integrated, accessible source of information capturing both Medicare and Medicaid utilization, costs, and quality, similar to MMLEADS.

One of the key advantages of the MMLEADS is that it contains summarized information for dual eligibles' costs and utilization in both the Medicare and Medicaid program (see sidebar for more detailed information about the MMLEADS). Although the MMLEADS data sets are large and require considerable storage space and analytic expertise to leverage effectively, acquiring and using this type of summarized data is less time and cost intensive than relying on disparate survey and claims-based files to obtain a full picture of Medicare and Medicaid utilization for dual eligibles. Ideally, this revised data source would also include information about LTSS paid by Medicaid (as is the case for the existing MMLEADS data) because this is a critical data gap for researchers seeking to understand the impact of integrated models without accessing the TMSIS TAF data.

Medicare-Medicaid Linked Enrollee Analytic Data Source (MMLEADS)

The Medicare-Medicaid Linked Enrollee Analytic Data Source (MMLEADS) integrates Medicare and Medicaid data to create a data source that includes information on enrollment, demographics, diagnostic conditions, service utilization, and spending information for three groups of beneficiaries: Medicare-Medicaid dually enrolled, Medicare-only, and Medicaid-only blind and disabled beneficiaries.

The restricted use data file consists of four linkable data files: two person-level files (Beneficiary/Enrollee File and Chronic Conditions File Medicaid Service), and two service-level files (Medicare Service-Level File and Medicaid Service-Level File).

We also recommend improving on the MMLEADS by incorporating information about managed care. The inclusion of more comprehensive data on managed care utilization is essential, given that the share of dual eligibles in managed care continues to increase and the majority of integrate care models relevant to dual eligibles rely on managed care to some extent.²²

Recommendation: Enhance sample size for dual eligibles in survey data to provide more comprehensive information on social needs and key subpopulations.

Although we did not conduct a sample size analysis of the surveys included in our review, it is unlikely that these data contain sufficient samples necessary to analyze social needs across integrated care models, or for key subpopulations within these models. For example, the 2018 MCBS (which is the only survey with a public facing codebook that provides counts of records in integrated care models) included 118 people enrolled in a Medicare-Medicaid Plan (MMP) and ten people in PACE. Enhancing the sample size of dual eligibles in these data sources could provide researchers with much more powerful information to understand the social needs that dual eligibles face, how they vary across key subpopulations, and, over time, whether integrated care models mitigate challenges with housing, food security, and financial stability.

Conclusion

Our review indicates that there is a range of rich, diverse data sources that can be used to address research questions related to the impacts of integrated care models for dual eligibles. However, our review also highlighted critical gaps in the availability, comprehensiveness, quality, and accessibility of these data sources.

The recommendations we outline above represent our assessment of the most efficient, high-value strategies for addressing these gaps and limitations. However, we recognize that implementing these recommendations would be complex and also require significant investments in time, expertise, and resources. As a result, it may be helpful to consider prioritizing their implementation, as well as considering phased approaches where feasible.

For example, the creation of a beneficiary-level source of information on enrollment in integrated care models may be the most feasible in the short term because there is already a strong data infrastructure that contains much of this information. Creating a summary data set similar to the MMLEADS is likely to be more time consuming and expensive, and it may be most effective to focus that investment on creating summarized information for Medicaid, both because this information is missing in the majority of other sources included in our review, and because the data source that contains it—the T-MSIS TAF—is relatively new and so the bench of researchers with experience using that complex data source is more limited than those with experience using the Medicare data, for example, which have been available for many years. In terms of investing in enhanced survey sample, it would be wise to make this investment in one data source first, and then take the lessons learned in terms of sampling, weighting, and analysis and apply them to additional sources going forward.

Enhancing the data available to understand the impact of integrated care models on dual eligibles is critical to improving health and well-being for this population, and for ensuring the financial sustainability of the public programs that support them.

Table A-1. Summary of Key Information from Data Assessment

KEY: ● - Yes ● - Not beneficiary level data	Most Recent Year of Data	Public Use File	Lowest Level of Geography	Trend Analyses	Link to Identify Integrated Models	Subpops. ²
Data Source Name and Description						
MBSF						
MBSF - Base <i>Data Type: Enrollment Unit of Analysis: Beneficiary</i> The MBSF base segment includes Medicare beneficiary enrollment information, (A/B/C/D) along with demographic information.	2019		ZIP Code	●	●	●
MBSF - Cost and Utilization <i>Data Type: Other Unit of Analysis: Beneficiary</i> The MBSF Cost and Utilization file includes one record per beneficiary, with summary utilization and annual payments for range of fee for service (FFS) inpatient, outpatient, DME, and prescription services.	2018		ZIP Code	●	●	●
MBSF - Chronic Conditions <i>Data Type: Other Unit of Analysis: Beneficiary</i> The MBSF Chronic Conditions segment includes one record per beneficiary with flags for each 27 chronic conditions.	2018		ZIP Code	●	●	●
MBSF - Other Chronic or Potentially Disabling Conditions <i>Data Type: Other Unit of Analysis: Beneficiary</i> The MBSF Other Chronic or Potentially Disabling Conditions file includes one record per beneficiary with flags for the presence of 35 chronic or potentially disabling conditions not included in the chronic condition file including behavioral health, tobacco and alcohol use, developmental disorders, and disability related conditions.	2018		ZIP Code	●	●	●
MBSF - National Death Index <i>Data Type: Other Unit of Analysis: Beneficiary</i> The MBSF National Death Index file includes one record per deceased beneficiary with information about cause of death.	2016		ZIP Code	●	●	●
ENCOUNTER & CLAIMS						
Medicare FFS Claims Files by Type <i>Data Type: Claims Unit of Analysis: Claim</i> The Medicare FFS RIFs are annual files drawn from several service types including: Skilled Nursing Facility (SNF), carrier (physicians, other professionals, clinical laboratories, freestanding radiology, etc.) inpatient, outpatient, home health agency, and Durable Medical Equipment. Enrollment and demographic information for beneficiaries associated with these claims is in the MBSF base file.	2019		ZIP Code	●	●	●
MA Encounter Data RIFs <i>Data Type: Claims (encounters) Unit of Analysis: Claim (encounter)</i> The MA encounter data RIFs are annual files drawn from several service types including: Skilled Nursing Facility (SNF), carrier (physicians, other professionals, clinical laboratories, freestanding radiology, etc.) inpatient, outpatient, home health agency, and Durable Medical Equipment. Enrollment and demographic information for beneficiaries associated with these encounters is contained in the MBSF base file.	2018		ZIP Code	●	●	●
T-MSIS TAF <i>Data Type: Survey, enrollment, claims, assessment Unit of Analysis: Multiple depending on file type (beneficiary; claim, enrollment time period)</i> The Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF) Research Identifiable Files (RIF) are a research-optimized version of T-MSIS data. The annual demographic enrollment (DE) file has information on demographics, eligibility and enrollment for all beneficiaries. The four monthly claims files have data related to inpatient hospital services, long-term care, pharmacy, and other services which includes: physician, outpatient hospital, dental, clinic, laboratory, X-Rays, home health, personal support, and managed care capitation payments. There are also several supplemental files: the eligibility dates file where each record represents an enrollment spell, the managed care enrollment file that contains information about the specific managed care plans beneficiaries have been enrolled in, a file related to waiver enrollment, a similar file related to Money Follows the Person, and a Health Home and State Plan Options file that summarizes enrollment in health homes, SPO, and presence of a Health Home Chronic Condition.	2018 ¹		ZIP Code	●	●	●
PLAN CHARACTERISTICS FILE						
Plan Characteristics File <i>Data Type: Other Unit of Analysis: Plan</i> The Plan Characteristics file contains Medicare Advantage plan and Prescription Drug Plan information separated into six subfiles: plan "base" file, premium file, cost sharing tier file, service area file, special needs plans file and multi-year crosswalk file.	2019			●	●	●
Integrated D-SNP Contract File Integrated D-SNP File <i>Data Type: Other Unit of Analysis: Plan</i> The Integrated D-SNP file contains information at the plan-level about whether a D-SNP qualifies as a HIDE-SNP or FIDE-SNP.	2021	●		●		●

KEY: ● - Yes ● - Not beneficiary level data	Most Recent Year of Data	Public Use File	Lowest Level of Geography	Trend Analyses	Link to Identify Integrated Models	Subpops. ²
Data Source Name and Description						
MMLEADS						
MMLEADS V2.0 - PUF <i>Data Type: Other Unit of Analysis: State</i> The MMLEADS Public Use Data File is an Excel Table that includes aggregate information on enrollment, demographics, diagnostic conditions, service utilization and spending information at the state level for three groups of enrollees: Medicare-Medicaid dually enrolled, Medicare-only, and Medicaid-only blind and disabled beneficiaries.	2012	●	State	●		
MMLEADS <i>Data Type: Enrollment, claims, linked Unit of Analysis: Beneficiary</i> The MMLEADS integrates Medicare and Medicaid data to create a data source that includes information on enrollment, demographics, diagnostic conditions, service utilization and spending information for three groups of enrollees: Medicare-Medicaid dually enrolled, Medicare-only, and Medicaid-only blind and disabled beneficiaries. The restricted use data file consists of four linkable data files, two person-level files (Beneficiary/Enrollee File, Chronic Conditions File Medicaid Service) and two service-level files (Medicare Service-Level File and Medicaid Service-Level File).	2012		ZIP Code	●	●	●
LONG TERM CARE						
LTCH - CARE <i>Data Type: Assessment Unit of Analysis: Assessment</i> Assessment instrument used for all patients receiving inpatient services in a facility certified as a hospital and designated as LTCH under the Medicare program. All-payer data set including Medicare FFS and Medicare Advantage beneficiaries, among other payers.	2018		State	●	●	●
LTC Minimum Data Set - Swing Bed <i>Data Type: Assessment Unit of Analysis: Assessment</i> The Long Term Care Minimum Data Set - Swing Bed is a health status screening and assessment tool. Swing bed providers are hospitals that can use their beds to provide either acute or post-acute skilled nursing care as needed.	2018		State	●	●	●
Home Health OASIS <i>Data Type: Assessment Unit of Analysis: Assessment</i> The Home Health Outcome and Assessment Information Set (OASIS) contains data items developed to measure patient outcomes and for improve home health care. The OASIS assessments are required of all home health agencies certified to accept Medicare and Medicaid payments.	2018		State	●	●	●
IRF-PAI <i>Data Type: Claims Unit of Analysis: Claim/event</i> The Inpatient Rehabilitation Facility - Patient Assessment Instrument (IRF-PAI) is collected on all Medicare patients who receive services from an inpatient rehabilitation unit or hospital. The data collected for IRF-PAI are used to measure quality of care.	2019		ZIP Code	●	●	●
PART D						
Part D Event File <i>Data Type: Assessment Unit of Analysis: Assessment</i> There are several linkable files associated with the Part D benefit. The event file includes information about each transaction/fill covered by both Prescription Drug Plans (PDPs) and Medicare Advantage Prescription Drug plans (MA-PDs).	2018		State	●	●	●
Part D MTM file <i>Data Type: Other Unit of Analysis: Beneficiary</i> Beneficiary-level information for participants in the MTM program. Eligible beneficiaries include those with multiple chronic diseases, taking multiple Part D drugs, and those likely to have high expenditures.	2018		State	●	●	●
Part D Prescriber Characteristics File <i>Data Type: Other Unit of Analysis: Provider</i> Descriptive information for the prescribers identified in the event file.	2019		State	●	●	●
Part D Pharmacy Characteristics File <i>Data Type: Other Unit of Analysis: Pharmacy</i> Descriptive information for the pharmacies identified in the event file.	2019		State	●	●	●
Part D Formulary File <i>Data Type: Other Unit of Analysis: Formulary</i> The Medicare Part D Formulary file is a suite of three sub-files: formulary, excluded drug and Over the Counter Drug that contain information on how the plan covers the prescription drugs filled.	2019			●	●	●
Part D Drug Characteristics File <i>Data Type: Other Unit of Analysis: Drug</i> The Medicare Part D Drug Characteristics file contains information that can be appended to the event file for researchers that do not have access to a proprietary NDC database.	2019			●	●	●

KEY: ● - Yes ● - Not beneficiary level data		Most Recent Year of Data	Public Use File	Lowest Level of Geography	Trend Analyses	Link to Identify Integrated Models	Subpops. ²
Data Source Name and Description							
HOS							
HOS - LDS³ <i>Data Type: Survey Unit of Analysis: Survey respondent</i> The Health Outcomes Survey (HOS) is a repeated cross-section nationwide survey of adult beneficiaries enrolled in Medicare Advantage plans. It is designed to measure the physical and mental health of a random selection of beneficiaries from each participating MA organization over two years (it includes an annual baseline survey and two year follow-up for each cohort). The survey includes a variety of patient reported physical and mental health outcomes and effectiveness of care measure. Two types of LDS files are available, baseline (completed data collected during an annual baseline) and analytic (merged cohort files containing baseline and follow-up data).		2019 ¹		County	●	●	●
HOS - RIF³ <i>Data Type: Survey Unit of Analysis: Survey respondent</i> The HOS RIFs contain all of the variables in the LDSs with the addition of specific direct person identifiers and plan identifiers that are excluded or aggregated in the LDSs. This includes information to identify the same individuals across multiple cohorts, such as name, address, Medicare Health Insurance Claim number, and Social Security Number. It also includes plan identifiers and characteristics such as contract number, enrollment at sampling, and plan name.		2019		ZIP Code	●	●	●
HRS							
HRS - PUF <i>Data Type: Survey Unit of Analysis: Survey respondent</i> The Health and Retirement Study (HRS) a national longitudinal study of the economic, health, marital, and family status, as well as public and private support systems, of older Americans (people age 50 and older).		2018	●	Nation ¹	●		●
HRS - Medicare Linked Data, Claims Files³ <i>Data Type: Survey, enrollment, claims, assessment Unit of Analysis: Multiple depending on file type (survey respondent, claim, assessment)</i> The Health and Retirement Study (HRS)-Medicare linked dataset links HRS survey information linked to CMS claims and assessment data for the HRS study population. Includes MBSF, MBSF chronic conditions, MBSF cost and utilization, inpatient hospital, skilled nursing facility, carrier, outpatient, hospice, home health, DME, Part D, and assessment data files (OASIS, Minimum Data Set, and inpatient rehabilitation facilities). Each of these data types are abstracted separately elsewhere in this document.		Varies ¹		ZIP code	●	●	●
HRS - Medicare Linked Data, Summary Files³ <i>Data Type: Survey, enrollment, summary claims Unit of Analysis: Beneficiary year, quarter, or interview gap (time between interviews)</i> Three summary files condense Medicare Part A and B claims and enrollment information into one observation per beneficiary in a specific period of time--annual (BASf), quarterly (BQSF), and based on the time period between HRS interview dates (BISf).		2015		ZIP code	●	●	●
MCBS							
MCBS - PUF <i>Data Type: Survey Unit of Analysis: Survey respondent</i> The MCBS is a continuous, in-person, multi-purpose longitudinal survey, representing the population of beneficiaries aged 65 and over and beneficiaries aged 64 and below with disabilities, residing in the United States. The MCBS primarily focuses on economic and beneficiary topics including health care use and health care access barriers, health care expenditures, and factors that affect health care utilization. Respondents are followed for up to four years. Hispanics, the disabled, and those over 85 are oversampled to support estimates for these subpopulations.		2017	●	Nation ¹	●		●
MCBS - Survey, LDS <i>Data Type: Survey Unit of Analysis: Survey respondent</i> The Survey LDS contains the same information as the PUF, but with considerably more detail within each content area. The Survey LDS also contains data for beneficiaries living in nursing homes or other facilities. (separate interviews are conducted with facility staff to gather information, respondents in facilities are not interviewed directly)		2018		Nation ¹	●		●
MCBS - Cost Supplement, LDS <i>Data Type: Survey, enrollment, claims, assessment Unit of Analysis: Multiple depending on file type (survey respondent, claim, assessment)</i> The cost supplement can only be requested in conjunction with the survey LDS file. It contains data at three levels: The Event level reports all payers, costs, and utilization at the most detailed level available (one observation per event per person). The Service Summary level summarizes the same information at the person- service level (one observation per service type per person, e.g., summary data for all inpatient use by one beneficiary). The Person Summary level summarizes all payers and costs across service categories and summarizes type of service amounts (one observation per person).		2018		Nation ¹	●		●

KEY: ● - Yes ● - Not beneficiary level data	Most Recent Year of Data	Public Use File	Lowest Level of Geography	Trend Analyses	Link to Identify Integrated Models	Subpops. ²
Data Source Name and Description						
NAMCAHPS						
NAMCAHPS - PUF <i>Data Type: Survey Unit of Analysis: Survey respondent</i> Survey of adults in Medicaid focused on patient experience and utilization.	2015	●	State			●
NAMCAHPS - LDS <i>Data Type: Survey Unit of Analysis: Survey respondent</i> The LDS provides additional variables related to age, race, ethnicity, and language not available in the PUF. It also does not suppress any values.	2015		State			●
NHATS						
NHATS – PUF <i>Data Type: Survey Unit of Analysis: Survey respondent</i> The NHATS gathers information in person from a nationally representative sample of Medicare beneficiaries ages 65 and older about a range of topics related to health, disability, physical and cognitive functioning in later life. Data include information from performance-based tests related to physical and cognitive capacity in addition to self-reported information. For respondents living in nursing homes/other facilities, separate interviews are conducted with facility staff to gather information about the residential setting. Respondents are interviewed annually to document change over time.	2019	●	Nation	●		●
NHATS - Medicare Linked data³ <i>Data Type: Survey, enrollment, claims, assessment Unit of Analysis: Multiple depending on file type (survey respondent, claim, assessment)</i> This file contains all of the information from the NHATS PUF and is linked to a full suite of Medicare claims and assessment data for linked respondents. Includes MBSF, MBSF chronic conditions, MBSF cost and utilization, inpatient hospital, skilled nursing facility, carrier, outpatient, hospice, home health, DME, Part D, and assessment data files (OASIS, Minimum Data Set, and IRF). Each of these data types are abstracted separately elsewhere in this document.	2017 ¹		ZIP Code	●	●	●
NHATS Restricted Data - Geographic identifiers <i>Data Type: Survey Unit of Analysis: Survey respondent</i> The restricted NHATS data with geographic identifiers includes city, state, county, and census tract for each NHATS Participant. City and state of residence for spouse/partners, children, and household members are available at each round. Files linking residence to the Dartmouth Health Atlas's Hospital Referral Region (HRR) to Rural-Urban Continuum Codes (RUCC) are also available.	2019		Sub-state ¹	●		●

1 - See full data inventory for details.

2 - Data source identifies characteristics of key subpopulations including rural/urban, over/under 65, race/ethnicity, full/partial dual status, those with high cost, chronic conditions, behavior health diagnosis or a disability. The majority of data sources require linkage to identify some or all subpopulations. See full assessment table for detail.

3 - Select integrated care models can be identified, but sufficiency of sample size to evaluate specific programs is unknown.

REFERENCES

- ¹ Medicaid and CHIP Payment and Access Commission (MACPAC). (2020). *Integrating care for dually eligible beneficiaries: Background and context* [Report to Congress on Medicaid and CHIP; Chapter 1]. MACPAC. <https://www.macpac.gov/wp-content/uploads/2020/06/Chapter-1-Integrating-Care-for-Dually-Eligible-Beneficiaries-Background-and-Context.pdf>
- ² Medicare Payment Advisory Commission (MedPAC). (2020). *Dual-eligible beneficiaries* [Data book; Section 4]. MedPAC. http://www.medpac.gov/docs/default-source/data-book/july2020_databook_sec4_sec.pdf?sfvrsn=0
- ³ Medicare-Medicaid Coordination Office. (2015). *Fiscal year 2015 report to Congress*. Centers for Medicare & Medicaid Services (CMS). https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/MMCO_2015_RTC.pdf
- ⁴ Rizer, A. COVID-19: If ever there was a time to care about Medicare-Medicaid integration, it's now. Arnold Ventures. <https://www.arnoldventures.org/stories/covid-19-if-ever-there-was-a-time-to-care-about-medicare-medicaid-integration-its-now/>
- ⁵ Center for Health Care Strategies (CHCS). (2021). *Promoting integrated care for dual eligibles (PRIDE)*. <https://www.chcs.org/project/promoting-integrated-care-for-dual-eligibles-pride/>
- ⁶ Integrated Care Resource Center (ICRC). (2021). *Home*. <https://www.integratedcareresourcecenter.com/>
- ⁷ Hayes, K., Hoagland, G. W., Harootunian, L., Salyers, E., & Wu, K. (2020, July 27). A pathway to full integration of care for Medicare-Medicaid beneficiaries. Bipartisan Policy Center (BPC). <https://bipartisanpolicy.org/report/a-pathway-to-full-integration-of-care-for-medicare-medicaid-beneficiaries/>
- ⁸ Arnold Ventures. (n.d.). *Complex care research agenda*. <https://craftmediabucket.s3.amazonaws.com/uploads/AVComplexCareResearchAgenda-v3.pdf>
- ⁹ State Health Access Data Assistance Center (SHADAC). (2020). Evaluations of integrated care models for dually eligible beneficiaries: Key findings and research gaps. Medicaid and CHIP Payment Access Commission (MACPAC). <https://www.macpac.gov/wp-content/uploads/2019/07/Evaluations-of-Integrated-Care-Models-for-Dually-Eligible-Beneficiaries-Key-Findings-and-Research-Gaps.pdf>
- ¹⁰ Medicare-Medicaid Coordination Office. (2020). *Guidance to Dual Eligible Special Needs Plans*. <https://www.cms.gov/files/document/cy2021dsnpsmedicaremedicaidintegrationrequirements.pdf>
- ¹¹ Center for Medicare and Medicaid Services. (2019). *Medicare-Medicaid Integration and Unified Appeals and Grievance Requirements for State Medicaid Agency Contracts with Medicare Advantage Dual Eligible Special Needs Plans (D-SNPs) for Contract Year 2021*. <https://www.medicare.gov/federal-policy-guidance/downloads/cib111419-2.pdf>
- ¹² State Health Access Data Assistance Center (SHADAC). (2020). Evaluations of integrated care models for dually eligible beneficiaries: Key findings and research gaps. Medicaid and CHIP Payment Access Commission (MACPAC). <https://www.macpac.gov/wp-content/uploads/2019/07/Evaluations-of-Integrated-Care-Models-for-Dually-Eligible-Beneficiaries-Key-Findings-and-Research-Gaps.pdf>
- ¹³ Medicare-Medicaid Coordination Office. (2020). *People dually eligible for Medicare and Medicaid*. Centers for Medicare & Medicaid Services (CMS). https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/MMCO_Factsheet.pdf
- ¹⁴ Office of Disability, Aging and Long-Term Care Policy (DALTCP). (2019). *Analysis of pathways to dual eligible status: Final report*. Office of the Assistant Secretary for Planning and Evaluation (ASPE). <https://aspe.hhs.gov/pdf-report/analysis-pathways-dual-eligible-status-final-report>
- ¹⁵ Medicaid.gov. (n.d.). *Balancing long term supports and services*. Centers for Medicare & Medicaid Services (CMS). <https://www.medicare.gov/medicaid/long-term-services-supports/balancing-long-term-services-supports/index.html>
- ¹⁶ Association for Community Affiliated Plans (ACAP). (2020). *Addressing social determinants of health through dual-eligible special needs plans: Gap analysis and policy development*. Center for Health Care Strategies (CHCS). <https://www.chcs.org/media/Arnold-Ventures-SDOH-Report-FINAL.pdf>
- ¹⁷ Medicaid.gov. (n.d.). *Data Quality Atlas: Exploring data quality (DQ) assessments by topic*. Centers for Medicare & Medicaid Services (CMS). <https://www.medicare.gov/dq-atlas/landing/topics/info>
- ¹⁸ Centers for Medicare & Medicaid Services (CMS). (2019). *Quantifying the distribution and completeness of select demographic variables in 2016* [Brief #4121]. Research Data Assistance Center (ResDAC). https://www.resdac.org/sites/resdac.umn.edu/files/4121_Demographic_Variables.pdf
- ¹⁹ Chronic Conditions Data Warehouse (CCW). (2021). *User documentation*. Centers for Medicare & Medicaid Services (CMS). <https://www2.ccwdata.org/web/guest/user-documentation>
- ²⁰ Chronic Conditions Data Warehouse (CCW). (2021). *User documentation*. Centers for Medicare & Medicaid Services (CMS). <https://www2.ccwdata.org/web/guest/user-documentation>
- ²¹ Medicaid and CHIP Payment Access Commission (MACPAC). (n.d.). *Dually eligible beneficiaries*. <https://www.macpac.gov/topics/dually-eligible-beneficiaries/>
- ²² Medicare-Medicaid Coordination Office. (2020). *People dually eligible for Medicare and Medicaid*. Centers for Medicare & Medicaid Services (CMS). https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/MMCO_Factsheet.pdf