Integrating Administrative Data with Survey-Collected Data to Reduce Burden in Establishment Data Collection

Andrea Mayfield[1], Rachel Carnahan[1], Felicia LeClere[1]
[1]NORC at the University of Chicago, 55 E. Monroe St., Chicago, IL 60603

Abstract

The Medicare Current Beneficiary Survey (MCBS) is a longitudinal panel, multi-purpose survey of a nationally representative sample of the Medicare population, conducted by the Centers for Medicare & Medicaid Services (CMS) through a contract with NORC at the University of Chicago. The MCBS collects detailed data about community- and facility-dwelling Medicare beneficiaries, including health care use and expenditures, health status, and other factors that affect health care utilization. Facility data are collected via in-person interviews with facility staff in both Medicare-certified facilities and other types of facility settings. The current Facility data collection approach is unnecessarily burdensome to facility staff because it includes certain items that are already reported to CMS for Medicare-certified facilities through the Long-Term Care Minimum Data Set (MDS) and the Certification and Survey Provider Enhanced Reports (CASPER) process. This paper examines the feasibility of reducing respondent burden by incorporating data from the appropriate MDS and CASPER administrative records in lieu of interview data for a sub-set of overlapping items for those beneficiaries who are located in Medicare-certified facilities. Our analyses assess the comparability of survey-collected data with parallel data points from MDS and CASPER administrative records for select beneficiaries. We also evaluate how MDS and CASPER administrative data can be seamlessly integrated with MCBS survey-collected data for beneficiaries living in other facility settings and for items not found in administrative records. We share how our results have informed plans to modify the MCBS Facility instrument and integrate MDS and CASPER administrative records, where available, to create blended data products and streamline MCBS Facility data collection.

Key words: Establishment data collection; administrative data; integration of administrative data and survey-collected data; respondent burden

1. Introduction

There is significant opportunity for government agencies to use administrative data to support their survey programs. Given the increasing difficulty and cost associated with conducting surveys, potential threats to data quality, and rising pressure to produce more relevant, timely, and cost-effective estimates, continued reliance on sample surveys as the sole means of supporting evidence-based policy-making is not sustainable (Citro 2014; National Academies of Sciences, Engineering, and Medicine 2017). Though not originally collected for statistical purposes, data collected by government entities to support program administration and regulatory functions may be used in combination with survey data to enhance the quality and cost-efficiency of statistical products and reduce respondent burden (Commission on Evidenced-Based Policymaking, 2017; National Academies of Sciences, Engineering, and Medicine 2017). The potential for administrative data to be used in concert with survey data varies by the quality of the administrative data source, characterized by its accuracy, timeliness, and accessibility, as well its comparability to the survey data source (Seeskin et al., 2018). Of particular interest is the ability to reduce respondent burden and associated survey costs by taking advantage of extant data sources, especially administrative data that are captured by the sponsoring agency.

Government agencies use administrative data in a variety of ways to support the production of national statistics, including the construction of survey sampling frames, validation or imputation of survey responses, and full replacement of surveys. The potential to link administrative data and survey data holds particular analytic promise, as these linkages allow for the production of blended statistics in which administrative data provides either ancillary or alternative measures to enhance survey data (Citro 2014; Lohr and Raghunathan, 2017; Seeskin et al., 2018). Two notable examples of successful efforts to integrate administrative records into survey data collection in the Federal Statistical System include the National Health Interview Survey (NHIS) and the Medicare Current Beneficiary Survey (MCBS). Both surveys produce information about health, health care use, and barriers to care by combining in-person survey questionnaires and linked Medicare enrollment and claims data (Seeskin et al., 2018).

The MCBS is especially suited to integrate administrative data. At enrollment in Medicare, each beneficiary is assigned a Medicare beneficiary identification number, which is shared across administrative data sets and allows for direct linkage between them. Since the MCBS uses administrative data from Medicare beneficiaries' records as the sampling frame for the survey, the sampling frame for the MCBS contains this identification number, which allows direct linkage of all sampled beneficiaries to other administrative data sources produced by the Centers for Medicare & Medicaid Services (CMS) (Parsell et al., 2017). The MCBS also links survey data directly to claims data files and uses these sources to reconcile health care costs for Medicare and non-Medicare covered services with imputation. Records that include a Medicare claim number are matched directly on the claim number, while the remaining records are matched based on an iterative method that aligns service date, event type, and provider. The resulting file contains data for medical event types and services and contains fields for survey only, claims only, and survey and claims combined. The final payment amounts and source are generated from a combination of the available data (Seeskin et al., 2018).

Despite the potential benefits of leveraging administrative data, they are often characterized by limitations in accuracy, timeliness, accessibility, and comparability. This can lead to a lack of comparability between records, which can delay the production of analytically sound data (Seeskin et al., 2018).

This paper examines the feasibility of reducing respondent burden by incorporating administrative data produced by CMS in lieu of interview data for specific, overlapping variables in a survey about facility-dwelling Medicare beneficiaries. We evaluate how issues of accuracy, timeliness, accessibility, and comparability can be reconciled between CMS administrative data sources and MCBS survey-collected data and how this research has informed plans to expand the use of administrative data on the MCBS to create blended Facility data products and streamline MCBS Facility data collection.

2. Background

2.1 Survey Description
The MCBS was launched in 1991 and is a continuously fielded, face-to-face survey of a nationally representative sample of the Medicare population conducted by CMS through a contract with NORC at the University of Chicago. The Medicare population includes all Medicare eligible persons aged 65 and over, and persons under age 65 with certain

disabilities or with end-stage renal disease (ESRD). The MCBS uses a rotating panel design and collects data from Medicare beneficiaries up to eleven times over a span of four years. Incoming panels are sampled and recruited in the fall of each year to replace the panel that rotates out in the winter. The survey covers topics including health care utilization and expenditures, sources of health insurance coverage, and health status and functioning. Data are collected for sampled beneficiaries living in noninstitutionalized (e.g., households, henceforth referred as "community") and institutionalized (e.g., nursing homes, henceforth referred to as "facility") settings.

The current MCBS Facility instrument collects data about facility-dwelling beneficiaries who reside in long-term care settings. About half of MCBS Facility interviews are conducted on behalf of beneficiaries residing in Medicare- or Medicaid-certified nursing facilities. The remaining Facility interviews are conducted in other types of long-term care settings, such as assisted living facilities, domiciliary care homes, personal care homes, and group homes.

During the Facility interview, interviewers administer questionnaire sections to knowledgeable facility staff rather than interviewing the beneficiary directly. Interviewers also abstract information from medical documentation, including medical records, billing records, and the most recent Long-Term Care Minimum Dataset (MDS) form, when available.

In 1997, certain items within the Facility instrument were updated to mirror the MDS form, thereby reducing the potential for errors in the abstraction process and making the abstraction task as efficient as possible. Despite its careful design, the current approach is burdensome to interviewers, who must abstract information from medical documentation, with or without the assistance of facility staff. The interview also requires the cooperation, time, and availability of multiple facility staff, such as the administrator, nursing staff, MDS coordinator, and billing officer. Further, selected questions are redundant with existing administrative sources that facility staff in Medicare- and Medicaid-certified facilities regularly report to CMS, including the MDS and the Certification and Survey Provider Enhanced Reports (CASPER). Shortening the Facility instrument by using existing CMS administrative data has the potential to reduce the burden on interviewers and facility staff.

2.3 CMS Administrative Data Sources
We investigated two CMS administrative data sources for their potential to replace portions of MCBS survey-collected data: (1) the MDS and (2) the CASPER.
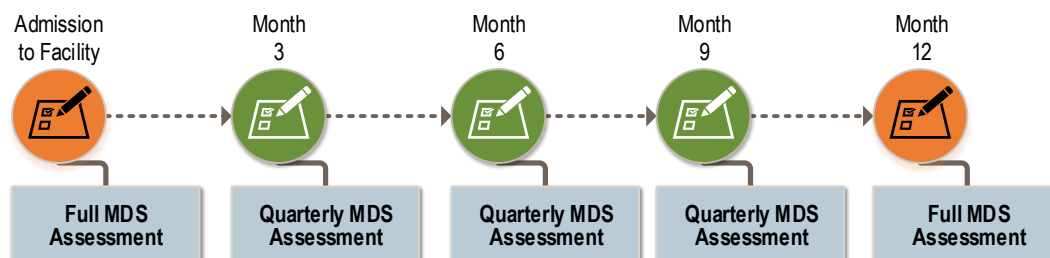
2.2.1 Long-Term Care Minimum Data Set (MDS)
The MDS is a federally-mandated health assessment of residents living in Medicare- or Medicaid-certified nursing homes. The purpose of the MDS is to assess and identify residents' health care problems, document individualized care plans, collect data for Medicare and Medicaid reimbursement systems, and monitor the quality of nursing home care. As such, the forms contain questions on numerous health-related topics, such as hearing, speech, and vision; cognitive patterns; mood; behavior; functional status; active diagnoses; health conditions; and medications (Centers for Medicare & Medicaid Services, 2016a).

There are two main types of MDS assessment forms – the Full MDS form and the Quarterly MDS form. Facility staff complete a Full MDS form for residents upon their admission to

a nursing home and, then, annually thereafter. The Full MDS assessment is also completed when a resident experiences a significant change in health status. In addition, a subset of items from the MDS is completed for each resident on a quarterly basis to monitor changes in the resident's health status between comprehensive assessments (known as the Quarterly MDS form). The MDS assessment schedule is illustrated in Exhibit 1.

Exhibit 1.   MDS Assessment Schedule

Admission to Facility	Month 3	Month 6	Month 9	Month 12
Full MDS Assessment	Quarterly MDS Assessment	Quarterly MDS Assessment	Quarterly MDS Assessment	Full MDS Assessment

2.2.2 Certification and Survey Provider Enhanced Reports (CASPER)
The CASPER data set supports the certification and regulatory function of CMS. Agencies perform regular surveys of facilities to determine whether the facility meets the requirements for participation in the Medicare and/or Medicaid programs. Chief among these requirements is the facility's performance and effectiveness in rendering a safe and acceptable quality of care (Centers for Medicare & Medicaid Services, 2016b). Certification survey data for every nursing home in the United States that is qualified to provide services under Medicare, Medicaid, or both are included in CASPER. As a provider-level data source, CASPER includes information such as facility name, facility address, number of beds by certification type, types of services provided, and aggregate information about resident health status and conditions. Importantly, CASPER also includes the CMS Certification Number (CCN), a unique six-digit identification number assigned to each facility certified to participate in Medicare and/or Medicaid. Recognizing the value of the CCN in identifying facilities that should have CASPER and MDS administrative data, a research plan was developed to determine if these administrative data sources could be used in lieu of survey-reported data for sub-set of overlapping items for those beneficiaries who reside in Medicare- or Medicaid-certified facilities.

3.   Methods

3.1 Analytic Objectives
We investigated two questions to assess whether CASPER and MDS administrative data sources could be used to shorten the Facility instrument:

1.   How many MCBS cases have records available in CASPER and MDS administrative data sources?
2.   How many variables can be skipped in the Facility instrument when CASPER and MDS administrative data are available for a case?

To answer these questions, we obtained and linked CMS administrative data from 2015-2017 to MCBS Facility data from Fall 2015, Fall 2016, and Fall 2017. We then assessed the coverage of MCBS cases in CMS administrative records. Next, we evaluated the item-level redundancy between CMS administrative data sources and the MCBS Facility instrument to determine the level of overlap between the two sources. We then assessed the agreement between CMS administrative data items and corresponding MCBS Facility

variables to determine whether CMS administrative records would serve as a sufficient substitute for survey-collected data for overlapping items. Finally, we also assessed how, operationally, MDS and CASPER administrative records should be integrated into MCBS Facility data products.

3.1 Availability of CMS Administrative Data

CMS provided us with an extract of 2015, 2016, and 2017 MDS administrative data linked to MCBS participants. The extract included records with assessment dates between January 1, 2015 and December 31, 2017, matched to Medicare beneficiary identification numbers for MCBS panel members in the same time period. The resultant files contained approximately 16,000 observations and 738 variables per year. We used 2016 CASPER data purchased from a vendor, Cowles Research Group (CRG).

3.2 Matching Process

To link MCBS Facility data, MDS administrative data, and CASPER administrative data with a high degree of certainty, it is necessary to establish linkages at three levels: (1) the beneficiary-level, (2) the assessment-level, and (3) the facility-level. MCBS Facility data contains information about the beneficiary's health status (beneficiary-level data), any reported MDS assessments (assessment-level data), and the beneficiary's place of residence (facility-level data). CASPER is a facility-level data source while MDS administrative data is an assessment-level data source, containing individual records for each MDS assessment conducted for each beneficiary within a specified time period.

While a beneficiary can have several MDS assessments conducted per year, only the most recent MDS assessment is accessed during the MCBS Facility interview. Therefore, when linking MCBS Facility data to MDS administrative data, it is necessary to conduct the linkage in two steps. First, we identify the appropriate MCBS beneficiary in MDS administrative data by linking the MCBS case identification number to the Medicare Beneficiary identification number. Next, we locate data for the appropriate MDS assessment by matching the MDS assessment date reported in the MCBS Facility instrument to the MDS assessment date recorded in MDS administrative data.

On their own, MCBS Facility data and CASPER administrative data share no common identifiers except Facility name and address, which may vary in terms of spellings and abbreviations across the two sources. MDS and CASPER administrative data, however, share a common identifier, referred to as "CMS Certification Number" or "CCN" in MDS administrative data and "Provider Number" in CASPER administrative data. Therefore, after establishing a beneficiary-assessment link between MCBS Facility data and MDS administrative data, we used the CCN to locate the exact facility-level record in CASPER administrative data. Exhibit 2 illustrates each step in the linkage process.

Exhibit 2. Process for Linking MCBS Facility Data to CMS Administrative Sources
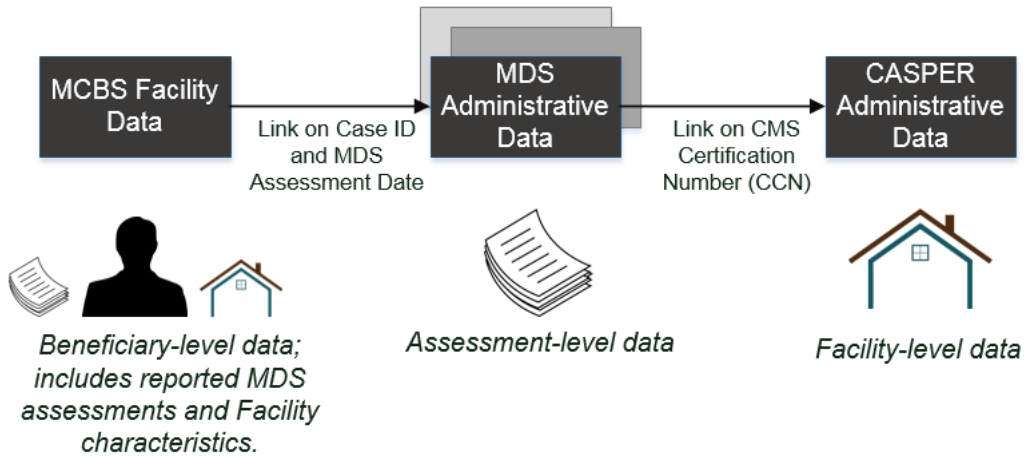
4. Results

4.1 Coverage of MCBS Facility Cases in CMS Administrative Data (Research Question 1)

Because MDS and CASPER administrative data are required for Medicare- or Medicaid-certified nursing homes, skilled nursing facilities (SNFs), and rehabilitation facilities, we would expect to see CMS administrative data only for MCBS Facility-dwelling beneficiaries living in these types of settings. In the fall rounds in 2015-2017 approximately 52-53% of MCBS Facility-dwelling beneficiaries lived in a Medicare- or Medicaid-certified nursing home, SNF, or rehabilitation facility that would be required to complete and report a MDS assessment to CMS (Exhibit 3). Approximately 46-47% of MCBS Facility-dwelling beneficiaries lived in a non-certified nursing home or some other facility setting, such as an assisted living facility or personal care home, where MDS assessments are not mandated.

Exhibit 3. Facilities Certified by Medicaid and Medicare in MCBS Facility Data, 2015-2017

Certified by Medicaid, Medicare, or Both	2015 N	2015 %	2016 N	2016 %	2017 N	2017 %
Nursing home	531	51.8	590	51.9	556	50.1
Hospital-based SNF unit or Rehabilitation Facility	14	1.4	23	2.0	27	2.5
Not Certified by Medicaid or Medicare	**2015 N**	**2015 %**	**2016 N**	**2016 %**	**2017 N**	**2017 %**
Nursing home	17	1.7	27	2.4	22	2.0
Assisted living facility, Board and care home, Personal care home, or some other type of facility setting	462	45.1	495	43.6	496	45.0

After linking MCBS Facility data to CMS administrative data, we investigated the degree to which administrative data are available for MCBS beneficiaries. In Fall 2015, Fall 2016, and Fall 2017, for about half of cases, facility staff reported an available MDS assessment during the MCBS Facility interview. Out of all MCBS Facility-dwelling cases, approximately 40% matched exactly to CMS administrative data, meaning that the

beneficiary's identification number was located in MDS administrative data, the MDS assessment date reported in MCBS Facility data matched to the assessment date recorded in MDS administrative data, and the beneficiary's place of residence reported in MCBS Facility data matched to the Facility's certification number in MDS and CASPER administrative data. Another six to eight percent of cases had a partial match in CMS administrative data, meaning that MDS assessments could be identified for a given beneficiary in MDS administrative data, but the exact assessment date reported in the MCBS Faculty instrument could not be located in MDS administrative data. For a small number of cases (1.8 to 4 percent), a MDS assessment was reported in MCBS Facility data but no assessment was found for the beneficiary in MDS administrative data (Exhibit 4).

Exhibit 4. Coverage of MCBS Facility Cases in CMS Administrative Data, 2015-2017

Type of Match	2015[a] N (%)	2016[a,b] N (%)	2017[b] N (%)
Total Facility Cases	1,024	1,112	1,102
Eligible for Matching: Reported MDS Assessment During MCBS Facility Interview	494 (48.2)	568 (51.1)	580 (52.6)
Exact Match: ✓ MDS assessment date reported in MCBS data matched to assessment date in MDS admin data ✓ Beneficiary's place of residence reported in MCBS data matched to Facility's certification number (CCN) in MDS and CASPER admin data	398 (39.6)	468 (43.6)	471 (44.1)
Partial Match: ✓ MDS assessment reported in MCBS data ✓ Assessments for MCBS beneficiary exist in MDS admin data ✗ MDS assessment date reported in MCBS data not found MDS admin data	78 (7.8)	71 (6.6)	66 (6.2)
No Match: ✗ MDS assessment reported in MCBS data but no assessment found for beneficiary in MDS admin data	18 (1.8)	29 (2.7)	43 (4.0)

[a,b] Match results for 2015 data are not significantly different from those for 2016 data but are significantly different from those for 2017 data (χ^2=10.6, p<0.01). Results from 2016 and 2017 data are not significantly different.

4.2 Number of Variables that Can be Skipped when CMS Administrative Data are Available (Research Question 2)

To determine the number of variables that can be skipped in the Facility instrument when CMS administrative data are available, we conducted two analytic steps. First, we ascertained the redundancy between CMS administrative data and MCBS Facility data to determine which variables would be candidates to skip when administrative data are available. Next, we assessed the comparability of the CMS administrative data and MCBS Facility data to determine if MDS administrative data could serve as a sufficient substitute for survey-collected data.

4.2.1 Overlap between the MCBS Facility Instrument and CMS Administrative Data

After comparing the contents of the MCBS Facility data files to the MDS administrative data files, we found that 73 out of 682 variables in the Facility instrument specifications matched exactly to the Full MDS form (which translates to 118 analytic variables in MCBS

Facility data files)[1]. This overlap occurred within the Health Status (HS) section of the MCBS Facility instrument, which was expected since HS was designed to mirror the MDS form to facilitate abstraction of data. An additional 19 variables from the Background (BQ), Health Insurance (IN), and Expenditures (EX) sections matched partially to items in the Full MDS form.

As CASPER is a provider-level data source, we limited our comparison of CASPER administrative data and MCBS Facility data to the Facility Questionnaire (FQ) section, which is the provider-level section in the instrument that collects information about the facility's contact information, certifications and licenses, types of services provided, bed counts by certification or license type, and billing rates. After comparing CASPER administrative data to MCBS Facility data, we found that 28 variables in the FQ section matched exactly to CASPER administrative data. Examples of matching items included facility name, address, and telephone number. Questions about the number of beds certified as Nursing Facility, Skilled Nursing Facility, or Intermediate Care Facility for Individuals with Intellectual Disabilities also matched exactly between the Facility instrument specifications and CASPER.

An additional 31 variables matched partially between the FQ specifications and CASPER forms and administrative data, which means there were differences in question wording, response options, question scope, or reference period. Partially matching items included questions about facility type, facility ownership, and services provided. Finally, an additional 55 variables in the FQ section were not found in CASPER forms or administrative data, including operational questions, questions about the number of beds licensed as personal care or assisted living facility beds, and questions about facility billing rates. Exhibit 5 illustrates the MCBS Facility instrument flow; the orange border on the FQ and HS sections highlight the sections of the Facility instrument that contain the most significant overlap with MDS and CASPER administrative data sources.

Despite being exact and partial matches with administrative data, questions about facility contact information, facility type, and facility ownership are considered essential to the data collection operation. As these items are therefore not candidates for removal from the survey, we excluded them from additional analyses.

Exhibit 5. MCBS Facility Instrument Flow and Overlap with CMS Administrative Data

---
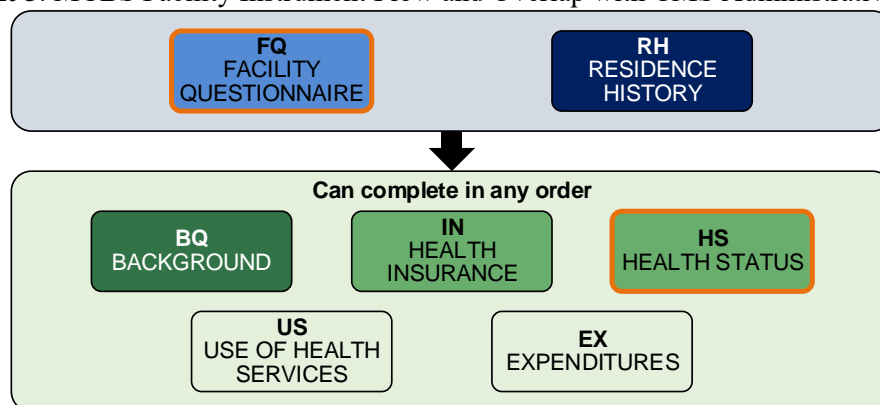
[1] To be considered an exact match, a variable on the MDS form and the Facility instrument specifications had to match in terms of (1) question text, (2) response options, (3) scope of question, and (4) reference period.

4.2.2 Agreement between the MCBS Facility Instrument and CMS Administrative Data
After determining overlap between MCBS Facility data and CMS administrative sources, we examined agreement for parallel data points. To assess comparability between MCBS Facility data and MDS administrative data, we calculated the total number of discrepant values across the 118 analytic variables that matched exactly between the two sources. Our calculation was limited to the MCBS cases for which we could determine an exact linkage to the CMS administrative data files.

Approximately 27-30% of cases had no discrepancies between MDS administrative data and MCBS Facility analytic variable values for each year of linked data, meaning that values for each parallel data point in the 2015-2017 files matched exactly (Exhibit 6). Nearly 82-84% of cases had only up to five discrepant values between parallel MCBS analytic and MDS administrative data points, meaning that values for up to five parallel variables did not match exactly. Fewer than 10% of cases had greater than 10 discrepant values for each year's linked data. After reviewing cases with high levels of discrepant variables, we have concluded that these discrepancies are likely the result of interviewer error. It is also worth noting that in some cases, one discrepant value on a filter question led to discrepant values on follow-up questions or check-all items. When discrepant values occurred, there were no patterns of divergence.

Exhibit 6. Agreement between MCBS Facility Data and MDS Administrative Data

	# MCBS/MDS Variables with Discrepant Values*	2015 Cumulative %	2016 Cumulative %	2017 Cumulative %
	0	29.2	27.1	26.9
	1	51.8	49.8	49.5
Median	2	68.6	63.9	62.9
Upper Quartile	3	77.4	73.9	71.1
	4	82.4	74.4	77.7
	5	83.9	83.1	81.9
	⋮	⋮	⋮	⋮
	10	93.7	89.7	91.5

*No significant differences found when comparing agreement results for 2015-2017 data via Chi-square testing.

To analyze agreement between the MDS administrative data and MCBS analytic data for partially matching variables, we attempted to recode the variables to match more closely. After conducting recodes, we compared values of MDS administrative data variables and MCBS analytic variables to determine how well they matched. Using these guidelines, we determined that 10 partially matching analytic variables are candidates for removal when MDS administrative data are available for a case, including items collecting the beneficiary's Medicaid number, mental health conditions, and other verbatim medical conditions.

To assess agreement between MCBS Facility data and CASPER administrative data, we examined comparability of items related to bed counts by certification type and personal care services questions. As part of confirming that a Facility is eligible for the MCBS

Facility instrument, the FQ section establishes whether the facility has any beds with the following types of certification or licenses:

- Certified by Medicaid as Nursing Facility;

- Certified by Medicare as Skilled Nursing Facility (SNF);

- Certified by Medicaid as Intermediate Care Facility for Individuals with Intellectual Disabilities (ICF-IID);

- Licensed as nursing home beds by the State Health Department; or

- Licensed as personal care, board and care, assisted living, or domiciliary care beds by the State Health Department.

The Facility instrument also asks for the number of dually certified beds; beds not licensed or certified as nursing beds; and the total number of facility beds (totaling eight analytic variables). As questions about Medicare-certified, Medicaid-certified, and total beds match exactly between CASPER administrative data and the Facility instrument, we compared values for these variables. As shown in Exhibit 7, number of Medicaid-only and Medicare-only beds agree closely between the two sources (>90% agreement) while the number of dually-certified and total beds have satisfactory agreement (62% to 76% across the three years). Differences in the number of dually-certified and total beds between CASPER administrative data and MCBS analytic data may be due to a point-in-time difference between the two data sources used for this analysis. While MCBS data were collected in Fall 2015, Fall 2016, and Fall 2017, respectively, the CASPER administrative data were from March 2016.

Another key part of establishing whether a facility is eligible for the MCBS Facility interview is to determine whether a facility provides one or more of the following personal care services:

- Nursing or medical care,

- Supervision over medications,

- Help with bathing,

- Help with dressing, or

- Help with eating.

Some information about personal care services is also available in CASPER forms and administrative data, though these variables match partially between the two sources due to question wording and code frame discrepancies. Whereas the Facility instrument uses broad terms to define each personal care service (e.g., help with walking), CASPER forms use more specific terminology (e.g., ambulation with assistance or assistive device). Further, while the Facility instrument asks whether services are available to residents via a "Yes/No" question, CASPER measures services provided to an aggregate number of residents using the Resident Census and Conditions of Residents form.

CASPER administrative data can be recoded to match MCBS analytic data. For example, if more than one resident is dependent or requires staff assistance for bathing, then we can assume that the facility offers assistance with bathing. There are some limitations to this approach. Even if a nursing home is not currently providing personal care assistance to

residents, it does not mean that they are incapable of providing such care or do not offer them. Despite this potential limitation, we were able to find highly comparable information (nearly 100%) for all services (Exhibit 7).

Exhibit 7. Agreement between MCBS Facility Data and CASPER Administrative Data

Item in MCBS and CASPER Data	2015 % Agreement	2016 % Agreement	2017 % Agreement
Bed Counts by Certification			
# Beds Medicaid Only	96	96	96
# Beds Medicare Only	93	89	90
# Dually Certified Beds	62	64	64
Total # Beds in Facility	76	75	74
Personal Care Services Offered			
Nursing or Medical Care	100	100	100
Supervision Over Medicines	100	100	99.6
Help with Bathing	100	100	100
Help with Dressing	100	100	100
Help with Eating	99.8	99.8	99.8

*No significant differences found when comparing agreement results for 2015-2017 data via Chi-square testing

5. Discussion

The MCBS is in an ideal position to leverage CMS administrative data because MDS and CASPER administrative data are available and accessible to the survey's sponsoring agency, CMS. In addition, MDS administrative data and the MCBS sampling frame both include the Medicare beneficiary identification number while MDS and CASPER administrative data both share the CCN, which facilitates linking between the three data sources. Further, given that parts of the MCBS Facility instrument were designed to mirror the MDS to aid in the abstraction process, this helps enhance the comparability of MCBS survey-collected data and MDS administrative data.

Our analyses assessed several potential challenges to incorporating CMS administrative data into MCBS Facility data collection and data file production, including potential issues surrounding the accessibility, timeliness, and comparability of the files. Our results demonstrated:

- ✓ CMS administrative data are **available** and **accessible** for MCBS data processing;
- ✓ We are able to **link** MCBS and CMS administrative records with a high degree of certainty;
- ✓ The **coverage** of MCBS Facility cases in CMS administrative data is on par with what we expected; and
- ✓ We found sufficient **agreement** between MCBS data and CMS administrative data for overlapping variables such that CMS administrative data could serve as a substitute for MCBS survey-collected data for beneficiaries living in certified facilities.

Based on these results, we will begin using CASPER and MDS administrative data sources to shorten the Facility instrument for beneficiaries living in Medicare- and/or Medicaid-certified facilities beginning in Fall 2019.

To determine that CMS administrative data are likely to be available for a beneficiary during post-interview data processing, we are modifying the MCBS Facility instrument to include a look-up tool, with CASPER administrative data as its data source. For any interview conducted at a nursing home, skilled nursing facility, or rehabilitation facility, the instrument will direct the interviewer to use the look-up tool to verify that the Facility's name, address, or CCN exists in CASPER administrative data (see Exhibit 8). Selection of a facility record from the look-up tool will trigger programming logic in the Facility Questionnaire section and Health Status section to skip select items redundant with the CASPER and MDS administrative data sources, resulting in a shortened interview for these cases. In total, we will skip 129 out of 682 variables in the Facility instrument (totaling 148 analytic variables). The variables to be skipped include 83 variables in the HS section that are redundant with MDS administrative data for Medicare- or Medicaid-certified facilities as well as variables in the FQ section about personal care services offered and bed counts by certification type that are redundant with CASPER administrative data.

The shortened Facility interview will still contain questions that ask the facility staff to provide information about whether there is a MDS assessment for the beneficiary and if so, the date of the most recent assessment. During data processing, the CCN collected via the lookup tool and the MDS assessment date collected via the interview will be used to aid in matching to the appropriate MDS and CASPER administrative data records and data for the variables skipped during the Facility interview will be incorporated into data processing.

The full Facility interview will continue to be administered on behalf of beneficiaries living in other facility settings or in nursing facilities for which a CCN is not located in the lookup tool. During data processing, survey-collected data from the full and shortened Facility interviews will be combined with MDS and CASPER administrative data to create blended data products.

Exhibit 8. Operational Process for Integrating CMS Administrative Data into MCBS Facility Data Collection and Processing
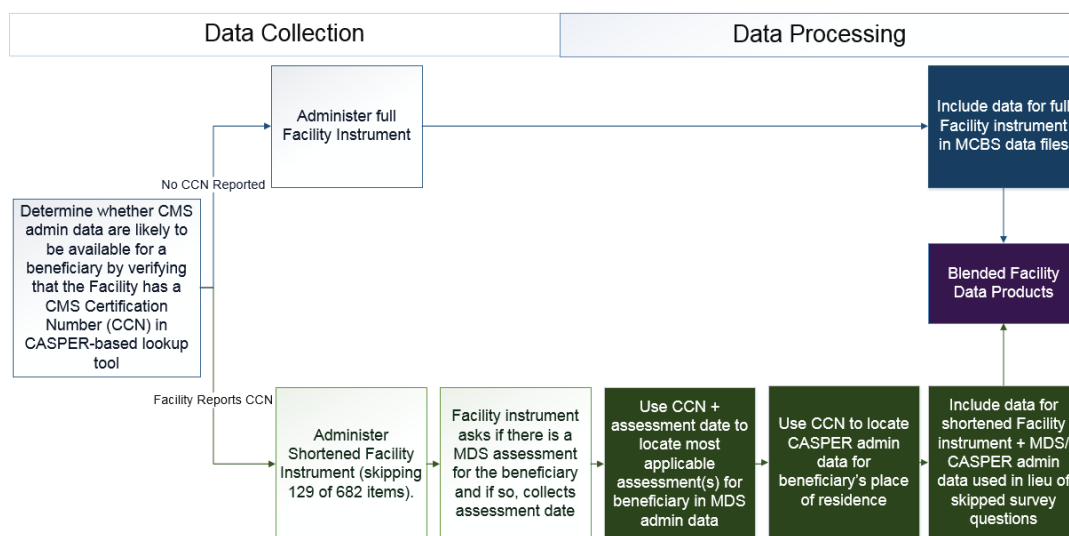
Using CASPER and MDS administrative data to streamline the Facility instrument offers several opportunities to the MCBS. First, shortening the Facility instrument and eliminating the need to abstract from MDS forms reduces burden for interviewers and facility respondents. The ability to skip questionnaire items and replace the information with administrative data will reduce burden for approximately 40-44% of annual facility cases. For cases with administrative data, changes to the Facility instrument flow are expected to reduce Facility interview administration time by approximately 15 minutes.

Reducing the length of this section for nursing home facilities should have positive impacts on gaining and maintaining cooperation. Further, reducing the amount of abstraction in the Facility instrument is a modernization that reflects the changing landscape of facility health records. Anecdotal feedback from interviewers indicates that with the adoption of electronic MDS assessments, it is increasingly difficult to gain access to electronic records for abstraction during the interview. Finally, removing collection of redundant information has the potential to enhance data quality.

We have conducted several phases of user acceptance testing on the changes to the MCBS Facility instrument. In addition, we have conducted small scale pilot testing on the associated changes to the data collection protocols and new items designed to collect the CCN from facility staff. Feedback from the pilot test has been implemented and plans are underway to implement the revised MCBS Facility instrument in 2019. Given that the accuracy, timeliness, accessibility, and comparability of MDS and CASPER administrative data sources can be reconciled with MCBS survey-collected data, MCBS Facility data products will reflect the successful integration of administrative and survey-collected sources starting with the 2019 Limited Data Set files.

References

Centers for Medicare & Medicaid Services. (October 2016). Long-term Care Facility Resident Assessment Instrument 3.0 User's Manual. Version 1.14. Retrieved from: https://downloads.cms.gov/files/MDS-30-RAI-Manual-V114-October-2016.pdf

Centers for Medicare & Medicaid Services. (2016, November 4). State Operations Manual. Chapter 2—The Certification Process. Retrieved from: https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/som107c02.pdf

Citro, C.F. (2014, December 19). From multiple modes for surveys to multiple data sources for estimates. *Survey Methodology*, 40 (2), 137-161.

Commission on Evidenced-Based Policymaking. (September 2017). *The Promise of Evidence-Based Policymaking*. Retrieved from https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf

Lohr, S.L., and Raghunathan, T.E. (2017). Combining survey data with other data sources. *Statistical Science, 32* (2), 293-312.

National Academies of Sciences, Engineering, and Medicine (NAS). (2017). *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*. Division of Behavioral and Social Sciences and Education; Committee on National Statistics; Panel on Improving Federal Statistics for Policy and Social Science Research Using Multiple Data

Sources and State-of-the-Art Estimation Methods. Washington, DC: National Academies Press.

National Center for Health Statistics, Office of Analysis and Epidemiology (NCHS). (2017). The linkage of National Center for Health Statistics surveys to Medicare enrollment and claims data - Methodology and analytic considerations. Hyattsville, Maryland. Retrieved from https://www.cdc.gov/nchs/data-linkage/cms/nchs_medicare_linkage_methodology_and_analytic_considerations.pdf

Parsell, B.W., LeClere, F.B., Ward, C.D., Rosner, S., Carnahan, R., Carris, K.L., Loganathan, S., Mayfield, A., McCarthy, T. and Reimer, R.L. (2017) "Using Administrative Data Proactively to Aid in Data Collection for New Respondent Panels for the Medicare Current Beneficiary Survey." JSM Proceedings, Administrative Data Section. Alexandria, VA: American Statistical Association.

Seeskin, Z.H., LeClere, F.C., Ahn, J., Williams, J.A. (2018). "Uses of Alternative Data Sources for Public Health Statistics and Policymaking: Challenges and Opportunities." JSM Proceedings, Government Statistics Section. Alexandria, VA: American Statistical Association. Pages 1822-1861.