Using Administrative Data Proactively to Aid in Data Collection for New Respondent Panels for the Medicare Current Beneficiary Survey

Bradley Parsell[1], Felicia LeClere[1], Christopher Ward[1], Samantha Rosner[1], Rachel Carnahan[1], Kari Carris[1], Sai Loganathan[1], Andrea Mayfield[1], Tricia McCarthy[1], Becky Reimer[1]
[1]NORC at the University of Chicago, 55 E. Monroe Street Suite 3000, Chicago, IL 60640

Abstract
The Medicare Current Beneficiary Survey (MCBS) is a continuous, multipurpose 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. A new panel of survey participants is recruited from current Medicare enrollees each fall. The task of locating, contacting, and interviewing new panel members is an important challenge for any survey. For the MCBS, these activities are particularly challenging because the elderly population is more likely to be found in health care institutions such as hospitals and short- and long-term care facilities. Also for those persons in poor health or facing mobility or cognitive challenges, participation in a four year survey may be an activity they are unwilling to undertake. In this analysis, we use administrative data from the Master Beneficiary Summary Files (MBSF) and Minimum Data Set (MDS), commercially available data for address and telephone matching, and MCBS case management data for panels entering in 2015 and 2016. We assess the value of the administrative and commercial data in predicting the location of respondents and the likelihood of completing an interview.

Key Words: Administrative Data, Healthcare, Longitudinal

1. Introduction

Longitudinal studies are uniquely suited to answer a range of questions in fields as diverse as public health and healthcare, economics, education, psychology, and criminology. Whereas cross-sectional studies observe opinions, behaviors, and characteristics at one point in time, longitudinal studies allow researchers to examine dynamic relationships. Much attention is paid to panel attrition in longitudinal studies, and for good reason. To draw inference from such studies, it is important that cohorts remain representative of their underlying population. Loss at the outset of data collection – whether a respondent cannot be located, cannot be contacted, or does not complete the first interview – deserves examination as it can also introduce substantial nonresponse bias. Some studies replenish the sample after attrition to rebalance the representativeness of the sample; nothing, however, can replace the data that would have otherwise been collected from sampled persons.

Within each wave of a longitudinal study, nonresponse can occur at one of three points: locating the sample, contacting the sample person to obtain the interview, or completing the interview (Maitland 2012; Lepkowski and Couper 2002). The effect of nonresponse at each stage is multiplicative, so sample loss at any stage exacerbates overall nonresponse. We are never able to attempt to gain the cooperation of or complete an interview with someone we are never able to locate. A number of factors contribute to difficulty with locating and contacting sample persons, such as migration and labor force patterns, household tenure, race, ethnicity, widowhood, and other covariates of geographic mobility (Couper and Ofstedal 2009; Zabel 1998).

Survey operations use strategies during each step of the data collection process—locating, contacting, and completion of the interview—to combat nonresponse. One challenge of devising strategies for the first wave of data collection is a lack of access to a priori information about sampled persons. For later waves of a longitudinal study, previously collected data may be used to inform retention efforts in subsequent waves of data collection. Previous research has demonstrated that merging external information to the sampling frame may help reduce nonresponse in the first round of data collection (Couper 2013). Data about sampled units are often available via commercial data vendors, but a limited amount of research has been conducted to examine their effectiveness in designing and improving strategies to locate and contact sampled persons (Kennickell 2009; Amaya, Skalland, and Wooten, 2013). Similarly, little is known about the use of administrative data to enhance initial panel recruitment. Administrative data have often been employed to evaluate the quality of data collection or for initial sampling (Calderwood and Lessof, 2009; Kreuter, Müller, and Trappmann, 2010; Sakshaug and Kreuter, 2012). The use of administrative data in survey operations may also introduce bias from record linkage methods, such as model-based linkage, which can introduce another source of error (Harron, Goldstein, and Dibben, 2016).

Previous research on the value of supplementing sample frame data with administrative records to aid in survey data collection is small but growing. Adding information on sampled units can help inform data collection strategies, including those used to identify and locate sampled persons. For example, data on sampled persons' migration patterns or health status may shed light on the likelihood of locating the respondent or gaining their cooperation. In one example, Durrant, Arrigo, and Steele (2011) showed that using supplementary data on household-level and neighborhood-level characteristics from the 2001 Census of the United Kingdom could improve response rates in a face-to-face survey by targeting call times based on information known about household and families.

A concern about surveys of the elderly and those who are more likely to be ill is the potential inability to complete interviews with less healthy sampled persons. A meta-analysis of longitudinal studies of the elderly population found that persons who were less healthy were more likely to drop out or not participate initially in a longitudinal study (Chatfield, Brayne, and Matthews 2005). This type of nonresponse can be difficult both to assess and to adjust for without knowledge about the health status of nonresponders. A recent study on attrition of panel respondents found that persons who reported a lower health rating were more likely to drop out of the MCBS (Ward et al., 2017). However, this study only included extant panel members and did not include non-responders from the original recruitment panel. Self-reported health status may be used to assess respondents' health in future interviews in a longitudinal study, but it is a subjective measure of health and, thus, may have limited utility. More objective measures of a person's health status may exist in administrative data that can be merged on to the sample frame and used to assess overall well-being and health.

Additional research is necessary to understand whether additional data merged to sample frame information can be used to reduce nonresponse through improved locating and cooperation at the outset of longitudinal data collection. In this research we use commercial and administrative data for the 2015 and 2016 incoming panels of the MCBS to investigate the utility of these types of data on initial panel contact and interview completion. Our first research question is whether augmenting the original sampling frame with commercial and administrative data will improve our ability to locate sample persons. List frames such as the MCBS often contain contact information, but it may be flawed or inaccurate. Thus, we test the value of commercial address matching and an administrative data source that helps locate individuals in facilities for locating sampled individuals. Secondly, we are interested in whether persons with objectively poorer health during their recruitment into the MCBS are less likely to cooperate and participate in the interview. To that end, we use administrative data to characterize the health of beneficiaries who are sampled for participation in the MCBS.

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 persons aged 65 and older, persons with certain disabilities, and persons with end-stage renal disease (ESRD). The MCBS uses a rotating panel design and collects data from Medicare beneficiaries up to twelve times over a span of four years. Incoming panels are sampled and recruited in the autumn of each year to replace the panel that rotates out in the summer. The survey covers topics including health care utilization and expenditures, sources of health insurance coverage, and health status and functioning, among others. Data are collected for sampled beneficiaries living in both noninstitutionalized (e.g., households, henceforth referred to as "community") and institutionalized (e.g., nursing homes, henceforth referred to as "facility") settings.

Different data collection protocols and instruments are used for community and facility interviews. While both instruments cover approximately the same topics, the MCBS community interviews are conducted with the sampled beneficiary or a designated proxy respondent, whereas facility interviews are conducted with facility staff rather than the sampled beneficiary. Due to the relatively small number of institutionalized persons, all newly-sampled beneficiaries are assumed to be in a community setting until a field interviewer identifies the location of the sampled person and uses case management protocols to move the sampled beneficiary to a facility interview. Standard locating protocols for sample members, then, are complicated in the MCBS by these additional steps.

The 2015 and 2016 incoming MCBS panels are the focus of our research. We use two incoming panels to understand how linked administrative data may help efforts to locate beneficiaries and complete interviews. Recruitment and data collection for these panels occurred during the autumn of 2015 and 2016, respectively. Administrative data from beneficiaries' records are used as the sampling frame for the MCBS. Each year, we use administrative enrollment data to select a clustered, stratified sample for the incoming panel. Table 1 below includes the sample sizes of the 2015 and 2016 incoming panels and the number of community and facility interviews attempted for each panel.

Table 1: 2015 and 2016 MCBS Incoming Panel Sample Sizes and Interview Types

Interview Type	*2015 Sample Size*	*2016 Sample Size*
Community	7,858	11,143
Facility	313	453
Total	8,171	11,596

While administrative enrollment data are a rich source of data for sampling and include an address for each beneficiary, that information may be inaccurate or not useful for contacting the sampled beneficiary. A beneficiary may have moved since the sample draw or the address on file may be the one that the beneficiary uses only for Medicare correspondence. These administrative enrollment data also do not contain any telephone information for sampled beneficiaries, which is one of the most useful items for locating a sampled beneficiary. To mitigate this operational problem, we supplement the sampling frame data with additional information. In what follows, we describe the data sources used to augment the sampling frame and define how they are used in the context of locating and interviewing respondents.

2.2 Data Sources

We used data from three main sources to address our research questions: (1) commercially-available data, (2) administrative data from CMS, and (3) MCBS survey operations paradata.

2.2.1 Commercial data sources

We used commercially-available telephone and address information to provide additional information on our sample frame for locating sampled beneficiaries. Each year, prior to the beginning of data collection, we attempt to match all sampled beneficiaries to commercial database records. A successful match returns telephone and address information for the beneficiaries. Because the sample frame does not have telephone information, all successful matches that contain a telephone number are provided. Address information is provided only when a matched address differs from the address provided on the sample frame. These matched telephone numbers and updated addresses are made available to field interviewers in the case management system. Table 2 shows the match and update rates for telephone numbers and addresses.

Table 2: Match and Update Rates for Commercial Telephone and Address Information in the 2015 and 2016 MCBS Samples

Match Type	*2015*	*2016*
Address update	2,566 (31.4%)	1,569 (13.5%)
Telephone number	5,651 (69.1%)	2,810 (24.2%)

2.2.2 Administrative data

The MCBS is uniquely suited to this analysis as there are a number of administrative data sources that can be directly linked to the sampling frame. At enrollment in Medicare, each beneficiary is assigned a Medicare Health Insurance Claim (HIC) number. This HIC number is shared across administrative data sets and allows for direct linkage between them. The sampling frame of the MCBS contains the HIC number, which allows direct linkage of all sampled beneficiaries to other CMS administrative data sources. Because the files are directly linked through a unique identifier, record linkage methods and their associated potential bias do not apply. Among the supplemental administrative files, data are available on sampled persons' address, race and ethnicity, and likely chronic conditions, among other types of data. Our research focused on identifying administrative data that may be useful for locating beneficiaries and assessing their health status. We identified three sources available to address our research questions. They include: (1) the long-term care Minimum Data Set (MDS), (2) the Master Beneficiary Summary File (MBSF), and the (3) internal CMS Hierarchical Conditions Category (HCC) data. All of these datasets are able to be directly linked to the sample frame for all sampled beneficiaries, giving us information on sampled beneficiaries who did and did not respond to the interview request.

The MDS is used in our initial analysis concerning locating respondents. The MDS is a federally-mandated health assessment of residents living in Medicare- or Medicaid-certified nursing homes. CMS uses the MDS to assess and identify facility residents' health care problems, document individualized care plans, collect data for Medicare and Medicaid reimbursement systems, and monitor the quality of nursing home care. Because the MDS is conducted both upon admission to a facility and on an ongoing basis, MDS records may contain timely information about the beneficiary's physical location. This may help us ascertain the location of a sampled beneficiary prior to data collection. It may also result in the elimination of the extra step in field operations where initial contacts occur in the community (i.e., an interview can be immediately fielded as a facility interview). This would save valuable fielding time that can be used instead to gain cooperation and complete the interview.

We used data from the MBSF for our second analysis of the effect of health on the number of contacts needed to complete an interview. More specifically, we used the chronic conditions segment and Part A and B claims files, to provide constructs to assess beneficiary health status. We used the Part A (hospital) and Part B (medical insurance) claims files to create an indicator of whether a sampled beneficiary had Medicare claims for hospice care. For the 2015 and 2016 panels the percentage of beneficiaries with a hospice claim was 0.99% and 1.00%, respectively.

The chronic conditions segment of the MBSF contains information on 27 chronic conditions for each sampled beneficiary[1]. We used the chronic conditions segment to create a discrete count of a beneficiary's conditions. Table 3 contains the percent of sampled beneficiaries by the number of chronic conditions indicated on the MBSF.

[1] For more information on the chronic conditions segment, see https://ccwdata.org/web/guest/condition-categories

Table 3: Weighted Percentages (and Standard Errors) of Count of Chronic Conditions in the 2015 and 2016 MCBS Samples

Number of Chronic Conditions	*2015*	*2016*
0	50.1% (0.81)	54.2% (0.71)
1	6.8% (0.33)	6.6% (0.31)
2	7.6% (0.35)	6.7% (0.27)
3	8.1% (0.37)	7.4% (0.29)
4	7.7% (0.36)	6.9% (0.26)
5	5.9% (0.31)	5.5% (0.25)
6	4.4% (0.26)	4.2% (0.21)
7	3.3% (0.22)	3.0% (0.17)
8	2.4% (0.18)	2.0% (0.15)
9	1.3% (0.13)	1.4% (0.11)
10 or more	1.8% (0.15)	2.2% (0.15)

Lastly, we used internal CMS HCC risk score data as an additional health status construct for our second analysis. CMS uses the HCC risk score to adjust capitated payments for beneficiaries in the Medicare Advantage (MA) program, among others. The HCC risk score is calculated for all Medicare beneficiaries, not just those in an MA program. It is a standardized measure created by CMS based primarily on International Statistical Classification of Diseases and Related Health Problems (ICD-10) codes found in Medicare claims and beneficiary demographics[2]. For our analyses, we used the HCC score as a construct of beneficiary health, with higher values representing poorer health status. Table 4 contains univariate distribution statistics for the HCC risk score used in our analyses.

Table 4: Weighted Distributional Statistics of HCC Risk Score in the 2015 and 2016 MCBS Samples

Statistic	*2015*	*2016*
Minimum	0.12	0.12
25th Percentile	0.5	0.5
Median	0.75	0.71
75th Percentile	1.3	1.24
Maximum	15.21	13.5

2.2.3 Survey operations paradata

We use an electronic case management system to manage field operations for the MCBS. It captures hundreds of paradata elements about the data collection process, which can be used to understand field procedures. We use paradata elements in our analyses both as analytic variables and for constructing dependent variables that measure the success of locating and contacting a sampled beneficiary. These paradata variables include the number of contacts attempted for a beneficiary, the data collection component (i.e., community or facility), whether a sampled beneficiary requires special locating efforts, and an outcome code that categorizes the result of the contact attempt.

[2] For more information on the HCC risk score calculation see https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.html

3. Methods

3.1 Locating Sampled Beneficiaries

To address our first research question--whether augmenting available sample frame data with commercial and administrative data improves our ability to locate sampled persons—we conducted two analyses. The first analysis used two logistic regression models to determine (1) if a sampled beneficiary required special locating effort and (2) if data collection efforts ceased due to inability to locate the sampled beneficiary. Special locating effort is defined as a field interviewer taking specific actions in the case management system to indicate that they are having trouble finding the sampled beneficiary. When a beneficiary is flagged for special locating, further investigation is done by field managers to locate them. Table 5 lists the number of sampled beneficiaries from 2015 and 2016 who required special locating efforts and the number of instances where data collection stopped for a sampled beneficiary due to failure to locate. About a quarter of the sample beneficiaries in incoming panels required special locating efforts, and field work was suspended on approximately 5 percent of the sampled beneficiaries because they could not be located.

Table 5: Frequency of Sampled Beneficiaries that Required Special Locating Effort and Failed Locating Efforts in the 2015 and 2016 MCBS Samples

Locating Effort	*2015*	*2016*
Required special locating effort	1,676 (20.5%)	2,711 (23.4%)
Failure to locate	359 (4.4%)	572 (4.9%)

We included the following variables as controls in our analyses: age (categorical), race, Hispanic ethnicity, a flag for Puerto Rican address (a sampling frame variable used for sample stratification), sex, and a flag which indicates whether a sampled address is a P.O. Box. The other predictors included in the models are those of interest: a flag indicating whether an address was updated from commercial data and a flag indicating whether a telephone number was successfully matched from commercial data. With these models, we were interested in assessing if a telephone match and/or address update predicted if a sampled beneficiary requires special locating effort or cannot be located.

In the second analysis on the predictors of locating, we assessed the viability of using the MDS administrative data to determine if a sampled beneficiary resides in an institutional setting prior to the start of data collection. Unlike the address and telephone data described above, matching the sample frame to the MDS data is not routinely done on the MCBS. Thus, this investigation was conducted post hoc to assess whether the linked MDS would have predicted whether a sampled respondent resided in a nursing home at the time of first contact. The MDS contains data for assessments done not only in institutional settings such as nursing homes but also institutions that do not require residence such as rehabilitation facilities[3]. Generally, if we encounter a sampled beneficiary who is using a rehabilitation facility, they are not a resident at the facility, or if they are a resident, it is temporary. If a field interviewer encounters a sampled beneficiary who is living in or getting treatment at a rehabilitation facility, he/she will attempt to follow up and conduct an interview at the

[3] Generally, if we encounter a sampled beneficiary who is using a rehabilitation facility, they are not a resident at the facility, or if they are a resident, it is temporary.

home or facility where the beneficiary currently resides or will be residing after their stay at the rehabilitation facility.

For each beneficiary, the MDS contains a record for each assessment he or she had. For this analysis, we only used the most recent MDS assessment in the administrative data as we believe this record will be the most likely record to have accurate information on the physical location of the beneficiary. We categorized MDS matches into four types: (1) no match to MDS, (2) likely institutional match, (3) a discharge assessment match, and (4) a Prospective Payer Systems (PPS) assessment match. We defined likely institutional matches as those which were not discharge or PPS assessments. PPS assessments are generally done at rehabilitation facilities, which, as noted, we differentiated from institutions like nursing homes where residence is more likely permanent. Discharge assessments are done at the time of a discharge from a facility, which suggests the beneficiary no longer resides at that facility and thus using the address from a discharge MDS assessment for locating purposes does not seem beneficial. Table 6 shows the match rate for each of the match types in the 2015 and 2016 MCBS panels. Roughly 3.5 percent of the MCBS samples in 2015 and 2016 successfully matched to the MDS.

Table 6: Frequency (and percentage) of MDS Match Types in the 2015 and 2016 MCBS Panels

Match Type	*2015*	*2016*
No match to MDS	7,793 (96.4%)	11,079 (96.5%)
Institutional match	190 (1.7%)	258 (1.6%)
Discharge match	135 (1.3%)	191 (1.4%)
PPS Match	53 (0.5%)	68 (0.5%)

We built a logistic regression model to predict whether a sampled beneficiary was pursued as a facility interview during data collection. When a field interviewer finds a sampled beneficiary is institutionalized, they use data collection protocols in the case management system to switch the sampled beneficiary from a community interview to a facility interview. Paradata from the case management system were used to create this outcome variable. We created three dummy variables for the MDS match types to assess the value of the data to locating after controlling for age (categorical), race, Hispanic ethnicity, a flag for Puerto Rican address (a sampling frame variable used for sample stratification), sex, and a flag which indicates whether a sampled address is a P.O. Box.

3.2 Contacting and completing interviews with sampled beneficiaries

Analyses to address our second main research question – whether a beneficiary's health status predicts our ability to complete an MCBS interview – used administrative data that are available for all sampled beneficiaries. For these analyses, we included only sampled beneficiaries who were fielded in the community. The facility interview is conducted with facility personnel rather than the sampled person and, thus, we did not expect that a beneficiary's health status would impact our ability to complete an interview. However, a community interview is completed with the sampled person or designated proxy respondent, and completing an interview with either may be impacted by the beneficiary's health status. This analysis was conducted in a post hoc manner; these health status constructs were not employed in any way during data collection for the 2015 or 2016 incoming panels.

We built two models, a logistic regression model and a linear regression model. The first model predicts the likelihood of completing an interview with the sampled beneficiary. Table 7 shows the number of sampled beneficiaries with completed community interviews in 2015 and 2016. Approximately 53 percent of sampled beneficiaries residing in the community completed an MCBS interview in the baseline panels for 2015 and 2016.

Table 7: Number (and Percentage) of Completed MCBS Community Interviews in 2015 and 2016 Samples

Completed MCBS Community Interview	*2015*	*2016*
Yes	4,098 (52.15%)	5,939 (53.3%)
No	3,760 (47.85%)	5,204 (46.7%)
Total	7,858	11,143

The second model predicts, for the beneficiaries who successfully completed an MCBS community interview, the number of contact attempts from a field interviewer required to complete an interview. Contact attempts are broadly defined as any effort made by the field interviewer to complete an interview. A contact attempt could be a visit to the beneficiary's home, a phone call to set up an appointment, or any other contact attempt recorded in the case management system. Because we measured the effort required to complete an interview, the research population for the second model is limited to beneficiaries with a completed community interview. Table 8 shows the distribution of number of contact attempts required to complete an MCBS interview. The median number of contacts for the incoming panels in 2015 and 2016 are 5 and 4 respectively.

Table 8: Distributional Statistics of Number of Contact Attempts Required to Complete an MCBS Interview in the 2015 and 2016 MCBS Samples

Statistic	*2015*	*2016*
Minimum	1	1
25th Percentile	3	2
Median	5	4
75th Percentile	8	8
Maximum	40	43

Both models included the same control predictors: age (categorical), race, Hispanic ethnicity, a flag for Puerto Rican address (a sampling frame variable used for sample stratification), sex, and a flag which indicates whether a sampled address is a P.O. Box. Predictors of interest in both models are constructs of beneficiary health status created from administrative data: a flag for the presence of a hospice claim, the HCC risk score, and the number of chronic conditions[4].

[4] Collinearity between HCC score and count of chronic conditions was a concern when building the model, as one might expect these measures to be highly correlated. However, the correlation between the two were rather low, Pearson's ρ of .3 in 2015 and .23 in 2016.

4. Results and Findings

4.1 Locating Sampled Beneficiaries

Table 9 includes the results of the first model, which predicts whether a sampled beneficiary required special locating effort during data collection.

Table 9: Whether a Sampled Beneficiary Required Locating Effort: Log Odds (and Standard Errors) for Commercial Data Matches

Predictor	*2015*	*2016*
Intercept	-1.42*** (.08)	-1.27*** (.04)
Age[1]		
<45 years	0.36** (.09)	0.35*** (.07)
45-64	0.24* (.11)	0.03 (.09)
70-74	-0.11 (.09)	0.02 (.07)
75-79	-0.15 (.09)	-0.1 (.07)
80-84	-0.29** (.09)	-0.23** (.07)
85+	0.02 (.09)	-0.13 (.07)
Gender[2]		
Male	0.05 (.07)	0.02 (.05)
Race[3]		
Black, Non-Hispanic	0.19 (.12)	0.29** (.09)
Asian	-0.13 (.22)	0.29* (.14)
Hispanic	0.05 (.13)	0.19 (.1)
American Indian or Alaskan Native	-0.04 (.36)	0.08 (.39)
Other	-0.01 (.36)	0.18 (.26)
Puerto Rican address	-0.95* (.39)	-0.06 (.31)
Address is P.O. Box	2.48*** (.13)	1.78*** (.17)
Address update	0.69*** (.08)	0.69*** (.08)
Telephone match	-0.73*** (.08)	-0.69*** (.07)

Note.—*significant at p<.05 level, **significant at p<.01 level, ***significant at p<.0001 level.
1: Reference category is age 65-69.
2: Reference category is female
3: Reference category is white, Non-Hispanic

We found that both an address update and telephone match from commercial data were significant predictors of the likelihood a sampled beneficiary required special locating efforts during data collection. Receiving a telephone number match for a sampled beneficiary reduced the likelihood a beneficiary required special locating effort. Receiving an address update from commercial data increased the likelihood a sampled beneficiary required special locating.

The second model has the same predictors as the first model, but the outcome variable is whether data collection efforts stopped due to a failure to locate the sampled beneficiary. Results from the second model appear in Table 10.

Table 10: Whether Data Collection Efforts Stopped due to Failure to Locate: Log Odds (and Standard Errors) for Commercial Data Matches

Predictor	*2015*	*2016*
Intercept	-3.42*** (.17)	-3.31*** (.09)
Age[1]		
<45 years	0.43** (.14)	0.66*** (.11)
45-64	0.42* (.17)	0.13 (.14)
70-74	-0.11 (.14)	0.15 (.11)
75-79	-0.13 (.17)	-0.47** (.13)
80-84	-0.47** (.16)	-0.44** (.16)
85+	0.02 (.16)	-0.21 (.12)
Gender[2]		
Male	0.12 (.13)	0.18 (.1)
Race[3]		
Black, Non-Hispanic	0.49* (.2)	0.63** (.17)
Asian	0.23 (.34)	0.8** (.21)
Hispanic	0.92*** (.19)	0.53** (.19)
American Indian or Alaskan Native	-0.52 (.73)	1.06* (.49)
Other	0.66 (.63)	0.28 (.45)
Puerto Rican address	-0.68 (.39)	0.84* (.38)
Address is P.O. Box	1.28*** (.18)	0.67** (.26)
Address update	0.66*** (.14)	0.67*** (.13)
Telephone match	-0.59*** (.15)	-0.54*** (.13)

Note.—*significant at p<.05 level, **significant at p<.01 level, ***significant at p<.0001 level.
1: Reference category is age 65-69.
2: Reference category is female
3: Reference category is white, Non-Hispanic

The second model shows results similar to those of the first model. Having matched telephone data for a sampled beneficiary significantly decreased the likelihood that we fail to locate a sampled beneficiary. Receiving an address update from commercial data was significantly predictive of being unable to locate a beneficiary.

The final locating model addressed the utility of the MDS for locating sampled beneficiaries residing in institutional settings. Table 11 contains the results of the logistic regression model which predicts whether a facility interview was attempted for a sampled beneficiary.

Table 11: Whether a Facility Interview was Attempted for a Sampled Beneficiary, Log Odds (and Standard Errors) for MDS Match Types

Predictor	*2015*	*2016*
Intercept	-1.9*** (.18)	-1.76*** (.18)
Age[1]		
<45 years	0.79** (.21)	0.67** (.22)
45-64	0.31 (.28)	0.29 (.29)
70-74	-0.44 (.23)	-0.63* (.28)
75-79	-0.52** (.19)	-0.42* (.2)
80-84	-0.02 (.19)	0.36* (.18)
85+	1.28*** (.16)	1.49*** (.16)
Gender[2]		
Male	-0.06 (.17)	-0.26 (.18)
Race[3]		
Black, Non-Hispanic	0.03 (.33)	-0.42 (.3)
Asian	-1.64* (.69)	0.11 (.51)
Hispanic	-0.16 (.36)	-0.47 (.37)
American Indian or Alaskan Native	0.12 (.6)	-0.49 (.5)
Other	-12.73*** (.22)	1.03 (.72)
Puerto Rican address	-0.33 (.89)	0.58 (.69)
Address is P.O. Box	0.29 (.38)	1.01* (.41)
MDS Match Type		
Institutional match	2.6*** (.21)	3.59*** (.23)
Discharge match	-0.71* (.27)	-0.93** (.25)
PPS match	0.42 (.31)	-0.01 (.32)

Note.—*significant at p<.05 level, **significant at p<.01 level, ***significant at p<.0001 level.
1: Reference category is age 65-69.
2: Reference category is female
3: Reference category is white, Non-Hispanic

For the sampled beneficiaries with an institutional MDS match, the likelihood of locating that person in a facility setting increases significantly. For sampled beneficiaries who are successfully matched to a discharge MDS record, the likelihood of locating them in an intuitional setting significantly decreased. Finally, a match to a PPS MDS record did not covary with locating the sampled beneficiary in a facility setting.

4.2 The Impact of Health Status on Contacting Sampled Beneficiaries and Completing Interviews

Table 12 contains the results of a model which predicts the likelihood of completing a community interview. In this model, we assessed whether a beneficiary's health status impacts our ability to complete an MCBS interview.

Table 12: Whether an MCBS Interview was Completed for a Sampled Beneficiary: Log Odds (and Standard Errors) for Beneficiary Health Status Indicators

Predictor	*2015*	*2016*
Intercept	-0.03 (.05)	0.08 (.05)
Age[1]		
<45 years	-0.07 (.08)	0.01 (.06)
45-64	0.3** (.1)	0.25** (.07)
70-74	-0.11 (.06)	-0.06 (.06)
75-79	0.08 (.06)	-0.01 (.06)
80-84	-0.04 (.07)	0.06 (.06)
85+	-0.07 (0.07)	-0.11 (.06)
Gender[2]		
Male	0.14* (.06)	-0.03 (.05)
Race[3]		
Black, Non-Hispanic	0.23* (.1)	-0.06 (.09)
Asian	-0.96*** (.21)	-0.75*** (.14)
Hispanic	0.02 (.1)	0.03 (.1)
American Indian or Alaskan Native	1.19** (.44)	0.37 (.36)
Other	-0.48 (.28)	-0.18 (.28)
Puerto Rican address	0.91** (.31)	0.4 (.3)
Address is P.O. Box	-0.37** (.11)	-0.15 (.15)
Presence of a hospice claim	0.11 (.28)	0.21 (.22)
HCC risk score	0.06* (.03)	0.1** (.03)
Chronic conditions count	0.01 (.01)	0.01 (.01)

Note.—*significant at p<.05 level, **significant at p<.01 level, ***significant at p<.0001 level.
1: Reference category is age 65-69.
2: Reference category is female
3: Reference category is white, Non-Hispanic

The presence of a hospice claim and the sampled beneficiary's number of chronic conditions were not significantly related to the likelihood of completing an MCBS interview net of other covariates. The HCC risk score was positively and significantly related to the likelihood of completing an MCBS interview.

The second model in our exploration of the respondent's health status on interview process and completion appear in Table 13. Here we assessed the impact that a beneficiary's health status has on the amount of effort required to complete the MCBS interview.

Table 13: Contact Attempts Required for a Completed Interview: Beta Coefficients (and Standard Errors) for Beneficiary Health Status Indicators

Predictor	*2015*	*2016*
Intercept	4.93*** (.19)	4.86*** (.21)
Age[1]		
<45 years	0.18 (.31)	-0.35 (.27)
45-64	-0.53 (.29)	-0.21 (.26)
70-74	-0.54* (.24)	-0.37 (.26)
75-79	-0.94*** (.23)	-0.52* (.23)
80-84	-1.02*** (.24)	-0.83** (.23)
85+	-0.97*** (.24)	-0.84** (.24)
Gender[2]		
Male	-0.09 (.14)	-0.1 (.11)
Race[3]		
Black, Non-Hispanic	0.55* (.26)	0.04 (.19)
Asian	1.57* (.75)	1.25** (.41)
Hispanic	1.03** (.27)	0.94** (.25)
American Indian or Alaskan Native	-1.85*** (.37)	-0.16 (.43)
Other	-0.01 (.6)	-0.06 (.59)
Puerto Rican address	-1.67*** (.34)	-1.64*** (.31)
Address is P.O. Box	-0.1 (.26)	-0.1 (.25)
Presence of a hospice claim	-1.22** (.42)	-0.36 (.46)
HCC risk score	0.08 (.06)	-0.07 (.05)
Chronic conditions count	-0.004 (.03)	-0.03 (.02)

Note.—*significant at p<.05 level, **significant at p<.01 level, ***significant at p<.0001 level.
1: Reference category is age 65-69.
2: Reference category is female
3: Reference category is white, Non-Hispanic

The presence of a hospice claim was significantly predictive of requiring one less contact attempt in the 2015 panel year. However, the presence of a hospice claim was not a significant predictor of the number of contact attempts required to complete an MCBS interview in 2016. Neither HCC risk score nor the number of chronic conditions were significant predictors of the effort required to complete an MCBS interview.

5. Discussion and Summary

Many researchers have used administrative data to conduct post hoc analyses and assess the accuracy of data collected in a survey. Further, most federal surveys that have address information do match to commercial sources to gather telephone numbers and update addresses where possible. Despite this, there is not an extensive literature on the value of administrative and commercial data to survey data collection. In this paper, we explored the value of adding commercial and administrative data sources available to the MCBS and assessed their potential usefulness to locating sampled beneficiaries and determining if we need to tailor data collection protocols for less healthy persons.

The analysis of matched commercial data returned promising results. First, providing field interviewers with a matched telephone number has a positive impact on our ability to locate a sampled beneficiary. Presence of a matched telephone number reduced the likelihood of needing special locating or failing to locate the sampled beneficiary. There are several reasons why this may be true. Persons with telephone data in commercial data banks may be less likely to have moved recently or the portability of their numbers allows us to find people who move. Interestingly, we also found the opposite result when receiving an address update for a sampled beneficiary; that is, an address update increased the probability that special locating efforts were needed. This may suggest that sampled beneficiaries for whom we receive a new address are more transient and difficult to find. From the second model, we also find beneficiaries with updated addresses are significantly more likely to be dropped from the survey because they are unlocatable. If there are systematic differences between the group of sampled beneficiaries for whom we do and do not get an updated address, we may have bias in our estimates. To combat this potential bias, more resources could be allocated to the data collection effort for sampled beneficiaries having an updated address. Given the relatively large number of updated addresses we receive for each sample, more research is needed in this area.

Our research which leverages the MDS administrative data to aid in locating institutionalized beneficiaries yielded positive results. While the MCBS frame contains address information, the sampled unit is a Medicare beneficiary; the address information may be for a person’s home, a relative’s home, or an institution such as a nursing home. Our hypothesis was that prior to data collection, we could use the MDS data to identify sampled beneficiaries who reside in institutionalized settings and field them as an MCBS facility interview. While data collection protocols already exist to switch a sampled beneficiary from a community interview to a facility interview, valuable data collection time could be saved through initial assignment of a sampled beneficiary to the facility component. We learned from our models how to properly match records in the MDS to predict when we would find a sampled beneficiary in an institutionalized setting. Given the results from the third model, our definition of a likely institutional match appears to have merit. Testing our matching process during data collection in a future panel is the next step to assess the value of a priori MDS data to identifying institutionalized sampled beneficiaries prior to starting data collection.

Given the topical areas of the MCBS, maintaining representativeness of beneficiary health status is important to the study. Based on previous literature, we hypothesized that sampled beneficiaries who are less healthy may participate in the study at a lower rate than more healthy beneficiaries. Based on the findings from our analysis of health status on contacting and completing interviews, this may not be the case for the MCBS. Of the three constructs of health status we created from extant administrative data, none had a significant negative impact on our ability to complete an MCBS interview. Conversely, the HCC risk score was a significant predictor of increased likelihood of completing an MCBS interview. It is feasible that this finding is related to the data collection protocol of the MCBS. Proxy respondents are sought for sampled beneficiaries who are unable to complete the interview themselves due to poor health or cognitive decline. Another contributing factor may be that less healthy people are generally home more often than healthy people. While this result offers some assurance that our data is representative of less healthy people, continued monitoring and research are prudent.

Acknowledgements

The research in this article was supported by the Centers for Medicare & Medicaid Services, Office of Enterprise Data and Analytics under Contract No. HHSM-500-2014-00035I, Task Order No. HHSM-500-T0002 with NORC at the University of Chicago. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of NORC at the University of Chicago or the Centers for Medicare and Medicaid Services. Research from the analysis was initially presented at the 2017 Joint Statistical Meetings in Baltimore, MD.

References

Amaya, A., Skalland, B., and Wooten, K. (2013). What's in a Match?. *Survey Practice*, *3*(6). Retrieved from http://www.surveypractice.org/index.php/SurveyPractice/article/view/148/html

Calderwood, L., and Lessof, C. (2009). Enhancing longitudinal surveys by linking to administrative data. *Methodology of longitudinal surveys*, 55-72.

Chatfield, M. D., Brayne, C. E., and Matthews, F. E. (2005). A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *Journal of Clinical Epidemiology*, *58*(1), 13-19.

Cheshire, H., Ofstedal, M. B., Scholes, S., and Schroeder, M. (2011). A comparison of response rates in the English Longitudinal Study of Ageing and the Health and Retirement Study. *Longitudinal and Life Course Studies*, *2*(2), 127.

Couper, M. (2013). Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys. *Survey Research Methods*, *7*(3), 145-156. doi: http://dx.doi.org/10.18148/srm/2013.v7i3.5751

Couper, M. P., & Ofstedal, M. B. (2009). Keeping in contact with mobile sample members. *Methodology of Longitudinal Surveys*, 183-203.

D'Arrigo, J., Durrant, G.B., and Steele, F. (2009). Using field process data to predict best times of contact conditioning on household and interviewer influences. (Working Paper M09/12). University of Southampton: Southampton Statistical Sciences Research Institute.

Harron, K., Goldstein H., and Dibben, C. (2015) *Methodological Developments in Data Linkage*. John Wiley & Sons.

Kennickell, A. (2009). Getting to the Top: Reaching Wealthy Respondents in the SCF. *Federal Reserve Board of Directors*. *http://federalreserve.gov/econresdata/scf/scf_workingpapers.htm*.

Kreuter, F., Müller, G., and Trappmann, M. (2010). "Nonresponse and measurement error in employment research: making use of administrative data." *Public Opinion Quarterly* 74, (5), 880-906.

Lepkowski, J. M., and Couper, M. P. (2002). Nonresponse in the second wave of longitudinal household surveys. *Survey Nonresponse*, 259-272.

Maitland, A., Casas-Cordero, C., and Kreuter, F. (2009). An evaluation of nonresponse bias using paradata from a health survey. In *Proceedings of the Section on Survey Research Methods. Alexandria, VA: American Statistical Association* (pp. 2250-2255).

Sakshaug, J. W., and Kreuter, F. (2012). "Assessing the magnitude of non-consent biases in linked survey and administrative data." In *Survey Research Methods*, vol. 6, no. 2, pp. 113-122.

Schoeni, R. F., Stafford, F., McGonagle, K. A., and Andreski, P. (2013). Response rates in national panel surveys. *The ANNALS of the American Academy of Political and Social Science*, *645*(1), 60-87.

Ward, C., Reimer, R., Corcoran, S., and Schluterman, N. (2017, July). *Predicting Sample Attrition in a National Study of Medicare Beneficiaries.* Presentation at the 2017 Joint Statistical Meetings, Baltimore, MA.

Zabel, J. E. (1998). An analysis of attrition in the Panel Study of Income Dynamics and the Survey of Income and Program Participation with an application to a model of labor market behavior. *Journal of Human Resources*, *33*(2), 479-506.