

Analyzing the Need for an Oversample of Low Income Beneficiaries in the Medicare Current Beneficiary Survey

Ying Li, Whitney Murphy, and Kirk Wolter

NORC at the University of Chicago, 55 E. Monroe Street, Chicago, IL 60603

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 through a contract with NORC at the University of Chicago. The MCBS evaluated the potential benefit of oversampling low-income beneficiaries to allow for improved estimation of parameters of health disparities experienced by this population. This paper demonstrates the methods used to conduct the evaluation, which included examining alternative approaches for oversampling low-income beneficiaries, such as using U.S. census data on income at the tract level, data from consumer databases on income at the household level, and geographic variation indicators of low-income status at the beneficiary level, and examining the effects of oversampling low-income beneficiaries on the standard errors of key statistics. Findings from the evaluation revealed that there were already sufficient numbers of low income beneficiaries in the MCBS for analytic purposes.

Key Words: Medicare, survey, sampling frame, low-income, oversample design

1. Background

1.1 Introduction to the MCBS

The Medicare Current Beneficiary Survey (MCBS) is a continuous, in-person, 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. The survey covers a variety of topics, including health care utilization and expenditures, all sources of health insurance coverage, and health status and functioning. Beginning with data collected in 2013, a public use file (PUF) and accompanying documentation is available free for download on the [MCBS PUF page](#). Additionally, two key sets of data from the MCBS are available through Limited Data Set (LDS) files: annual, person-level Survey File and Cost Supplement data.

1.2 The MCBS Sample Design and Frame

The MCBS uses a rotating panel sample design and represents the population of beneficiaries in the 50 states, District of Columbia, and Puerto Rico. Each sampled beneficiary is scientifically selected as part of an annual panel and is interviewed up to three times (Fall, Winter, Summer) per year for four consecutive years to form a continuous profile of their health care experiences. One panel is retired during each summer round, and a new panel is selected to replace it each fall round. Sampled beneficiaries may be

living in the community (e.g., their homes) or a facility (e.g., nursing homes). The MCBS employs a three-stage cluster sample design:

Primary sampling units (PSUs) are made up of major geographic areas consisting of metropolitan areas or groups of rural counties.

Secondary sampling units (SSUs) are made up of census tracts or groups of tracts within the selected PSUs.

Ultimate sampling units (USUs) are Medicare beneficiaries selected from within the selected SSUs.

The sampling frame for the Medicare beneficiaries utilizes Medicare administrative enrollment data. This data is the authoritative source for all Medicare entitlement information, containing information on all individuals entitled to Medicare, including demographic information, enrollment dates, third party buy-in information, and Medicare managed care enrollment. The administrative enrollment data do not include personal income information.

1.3 Analysis Background and Definitions

Analysts have shown increased interest in studying the low-income population in the MCBS across key health outcomes and other variables. Our primary research was aimed at determining whether oversampling of low-income beneficiaries for the MCBS was needed before potential implementation in the 2017 Fall Round. The main goal of the oversampling would be to increase the number of low-income beneficiaries in the MCBS enough to allow for improved precision of estimates of health disparities experienced by this population. For the purposes of our oversampling analyses, we defined various income groups based on Federal Poverty Level (FPL) thresholds according to the U.S. Census Bureau's definition:

Poor beneficiaries are defined as those with incomes at or below 100 percent of the FPL.

Near-Poor beneficiaries are defined as those with incomes greater than 100 percent and less than or equal to 200 percent of the FPL.

Low-Income beneficiaries are defined as those with incomes at or below 200 percent of the FPL (This encompasses both the poor and the near-poor groups above).

In this paper, we first enumerate the low-income population both in the overall U.S. population and in the MCBS sample. We then explore the impact of an oversample of 200 low-income beneficiaries on the resulting MCBS sample and the standard errors of an estimated population proportion. Finally, we describe various low-income classifiers available for use in oversampling and discuss some design options for oversampling low-income MCBS beneficiaries living in the U.S. Throughout the analysis, we excluded Puerto Rico and facility beneficiaries by design.

2. Research Outcomes

2.1 Comparison of Current MCBS Low-Income Sample Size to Benchmark

An examination of the 2015 American Community Survey (ACS) data gives a general picture of the size and distribution of the poor, near-poor, and low-income populations in the United States. We used the 2015 ACS as a benchmark for comparison with the low-income sample size in the MCBS. In order to maximize comparability with the MCBS, which collects total income only for beneficiaries and their spouses, we kept only ACS households with a householder age 65 or older, with or without a spouse, and with no other adults in the household. We classified MCBS beneficiaries into the three income-level groups described earlier—poor, near-poor, and low-income—based on self-reported total income in the 2015 Survey File. Throughout all of the analyses, we excluded Puerto Rico and facility beneficiaries by design.

Exhibit 2.1 displays the proportions of the 2015 MCBS sample and U.S. population, age 65 and older, which fall into the income groups described above. For example, 39 percent of the 65 and older MCBS beneficiaries in the sample are defined as low-income for the purposes of our analyses (i.e., they have incomes at or below 200 percent of the FPL), while according to the ACS, 33 percent of the 65 and older U.S. population are defined as low-income using the same criteria. About 11 percent of both the 65 and older MCBS sampled beneficiaries of the U.S. senior population are considered poor for the purposes of our analyses (i.e., they live at or below 100 percent of the FPL).

Exhibit 2.1: Low-income Populations in the MCBS Compared to the U.S Population*

Income Group	Frequency #	Percent of Total Population
2015 MCBS Survey File (SF), Self-Reported Income, Age 65+		
Total 65+ Population	10,556	100.0
Poor (<=100% of the FPL)	1,208	11.4
Near-Poor (>100% and <=200% of the FPL)	2,910	27.6
Low-Income (<=200% of the FPL)	4,118	39.0
2015 American Community Survey (ACS), Age 65+		
Total 65+ Population	23,574,477	100.0
Poor (<=100% of the FPL)	2,529,261	10.7
Near-Poor (>100% and <=200% of the FPL)	5,281,695	22.4
Low-Income (<=200% of the FPL)	7,810,956	33.1

*ACS includes Household size of 1 or 2 members only to be comparable with MCBS population. Puerto Rico and Facility populations excluded in both Files.

2.2 Impact of Oversample

An oversample size of 200 additional low-income beneficiaries per year for four years was proposed in response to the growing interest in this population. Oversampling, if needed, would begin in the Fall 2017 round, and an oversample would be selected every Fall round through Fall of 2020.

Without oversampling, the projected MCBS sample size by 2020 is 14,449 total beneficiaries, of which 6,575 are projected to be low-income. With the proposed oversample, the projected MCBS sample size increases to 14,996 total beneficiaries by 2020, of which 7,122 are expected to be low-income. In other words, taking sample attrition and deaths into account, we expect an increase of low-income beneficiaries in the MCBS sample over the four year period.

An important factor in analyzing the impact of an oversample is determining how the standard errors are affected. Standard error is an indication of the reliability of estimates; thus, a smaller standard error suggests that the corresponding estimate is a more precise reflection of the actual population parameter being estimated.

Exhibit 2.2 displays the standard errors for an estimated population proportion under two scenarios: the current expected low-income beneficiaries, given no oversampling (6,575 low-income completes expected by 2020), and the expected low-income beneficiaries with the proposed oversample (7,122 low-income completes expected by 2020). In these calculations, we assume a true population proportion of 0.5, and a design effect of 2.0.

The overall standard errors decrease, as expected, but only very slightly, with the addition of the oversample. Standard errors for the low-income group (poor and near-poor combined) range from 0.0087 without oversampling to 0.0084 with oversampling. The poor subgroup, which has the smallest sample size among the FPL groups studied here, has the largest standard errors, ranging from 0.0141 without oversampling to 0.0135 with oversampling. Standard errors for the near-poor range from 0.0111 without oversampling to 0.0107 with oversampling. As we can see, there is no significant precision gains of estimated measurement of interest if we oversample low-income beneficiaries as proposed, likely because the additional oversample size is small compared to the already large low-income sample in the MCBS.

Exhibit 2.2: Standard Errors for an Estimated Population Proportion for Oversampling Options in 2020*

Low-Income Group	Oversampling Options	
	No Oversampling	Oversampling: 200 Additional Low-Income/ Year**
Total Low-Income (<=200% FPL)	0.0087	0.0084
Poor (<=100% FPL)	0.0141	0.0135
Near-Poor (>100 to <=200% FPL)	0.0111	0.0107

*Assumes true P = 0.5 and Design Effect = 2.

**Assumes the oversampled low-income completes are distributed across the poor and near-poor groups in the same proportions as in the current sample.

2.3 Methods for the Selection of an Oversample of Low-Income Beneficiaries

Since the administrative enrollment data do not include income information, we conducted exploratory analyses on income resources available for MCBS if low-income oversampling were to be implemented. We identified three low-income classifiers for potential use in oversampling and compared self-reported income with each of the three low-income classifiers to determine their accuracy and coverage.

2.3.1 Census Tract-Level Income Data

Summary income data are available at the tract level through the 2015 ACS data file. These data allow us to identify the proportion of each tract in the U.S. that falls below various FPL thresholds. Using this data, we classified the MCBS tracts as “low-income” if at least 50 percent of the 65 and older population within the tract has income at or below 200 percent of the FPL, and “not low-income” otherwise. We applied these tract-level classifications to beneficiaries based on the Federal Information Processing Standardization (FIPS) code (i.e., tract) associated with the beneficiary’s geocoded address, and we then compared that classification to the beneficiary’s self-reported income from the 2015 MCBS Survey File to validate the accuracy of the census classification in predicting actual income level.

Exhibit 2.3.1 displays the results of this comparison. The columns reflect the tract-level classification described above; “Low-Income” columns enumerate beneficiaries living in tracts classified as low-income, and “Not Low-Income” columns enumerate beneficiaries living in tracts classified as not low-income. The rows reflect income classification based on the 2015 incoming panel beneficiaries’ self-reported income in the 2015 MCBS Survey File: “Low-Income” rows enumerate beneficiaries with self-reported income at or below 200 percent of the FPL, and “Not Low-Income” rows enumerate those who self-report income above 200 percent of the FPL.

Exhibit 2.3.1: Conditional Probabilities of Self-Reported Income vs. Census Tract-Based Income Classification in 2015* SF

Self-Reported Income Classification of 2015 Incoming Panel	Census Tract-Level Income Classification					
	Low- Income # (%)	Not Low- Income # (%)	Total # (%)	Low- Income row %	Not Low- Income row %	Total row %
Low-Income (≤200% FPL)	448 (69.1)	1,390 (41.6)	1,838 (46.1)	24.4	75.6	100.0
Not Low-Income (>200% FPL)	199 (30.8)	1,950 (58.4)	2,149 (53.9)	9.3	90.7	100.0
Total	647 (100.0)	3,340 (100.0)	3,987 (100.0)	16.2	83.8	100.0

*Includes 2015 incoming panel only (these beneficiaries first started participating in the MCBS in the Fall 2015 Round). Excludes beneficiaries residing in Puerto Rico and Facility.

The census tract-based income classification proves to be somewhat accurate, with 448 (69.1 percent) of the 647 census-coded low-income beneficiaries actually self-reporting as

low-income. However, among the 1,838 beneficiaries self-reporting as low-income, only 448 (24.4 percent) were accurately coded as low-income via the census classification. This is partly due to the restriction that at least 50 percent of the tract be at or below 200 percent of the FPL in order to be identified as low-income; a lower threshold would increase the number identified, but would decrease the accuracy overall. Thus, while this census flag may be helpful in identifying low-income beneficiaries, there are many beneficiaries that it misses.

Another way of describing the accuracy of these classifiers is through the concepts of sensitivity (i.e., the true positive rate) and specificity (i.e., the true negative rate). In this case, the census tract classifier has low sensitivity (448/1838), but high specificity (1,950/2,149).

2.3.2 CCW MBSF Income Indicators

Several low-income related indicators are available in the Chronic Condition Warehouse Master Beneficiary Summary File (CCW MBSF), including monthly state buy-in indicators, low-income subsidy variables, and Medicaid eligibility information. The state buy-in indicators identify whether the state Medicaid program paid for the beneficiaries' Medicare premiums; the low-income subsidy variables identify whether a beneficiary is receiving a subsidy to help for prescription drugs (limited to part D enrollees); and the Medicaid eligibility code identifies whether a beneficiary is eligible for both Medicare and Medicaid. Using these indicators, a beneficiary is classified as low-income if at least one of the indicators identifies that the beneficiary receives these benefits for at least half of the calendar year.

Again, as part of our analyses, we undertook a comparison of this CCW MBSF low-income classification code to the 2015 incoming panel beneficiaries' self-reported income from the 2015 MCBS survey file to validate the efficacy of this low-income code in predicting actual income level. Exhibit 2.3.2 displays these results.

The CCW MBSF low-income classification code proves to be quite accurate, with 820 (95.8 percent) of the 856 MBSF-coded low-income beneficiaries actually self-reporting as low-income. However, among the 1,838 beneficiaries self-reporting as low-income, only 820 (44.6 percent) were accurately coded as low-income via the MBSF data (i.e., were fully eligible for at least one of the programs identified by the MBSF data for at least half of the year). Similar to the results for the census tract classifier, this reflects a level of "undercoverage" in the MBSF data classifier. In other words, while the MBSF flags are highly accurate for beneficiaries' self-reporting income, they do not identify the entire low-income population in the MCBS.

In terms of sensitivity and specificity, the CCW MBSF classifier displays a higher (but still modest) sensitivity (820/1,838) compared to the census tract level classifier, and very high specificity (2,113/2,149).

Exhibit 2.3.2: Conditional Probabilities of Self-Reported Income vs. CCW MBSF Low-Income Classification in 2015* MCBS Survey File

Self-Reported Income Classification of 2015 Incoming Panel	CCW MBSF Income Classification					
	Low-Income # (%)	Not Low- Income # (%)	Total # (%)	Low- Income row %	Not Low- Income row %	Total row %
Low-Income (≤200% FPL)	820 (95.8)	1,018 (32.5)	1,838 (46.1)	44.6	55.4	100.0
Not Low-Income (>200% FPL)	36 (4.2)	2,113 (67.5)	2,149 (53.9)	1.7	98.3	100.0
Total	856 (100.0)	3,131 (100.0)	3,987 (100.0)	21.5	78.5	100.0

*Includes 2015 incoming panel only. Excludes beneficiaries residing in Puerto Rico and Facility.

2.3.3 Commercial Vendor-Provided Income Data

Vendors use consumer and other data sources to predict a variety of demographic characteristics at the household or person level, including age, race/ethnicity, income, and marital status. Using one such vendor, Marketing Systems Group (MSG), we matched our MCBS sample to the vendor's income data via address to create an income flag. For each beneficiary, we compared the vendor-provided income with self-reported income. This analysis was conducted on an internal, unreleased data file. The conditional probability of the beneficiaries with vendor-coded income under \$25,000 annually who also self-reported annual income under \$25,000 is lower than for other income classifiers. Thus, the vendor-provided income data did not add any information beyond that provided by other indicators.

3. Conclusions and Outcomes

Because it was determined that there are already sufficient numbers of low-income beneficiaries in the MCBS for analytic purposes and that the proposed oversample would not noticeably increase the precision of estimates, CMS determined that a low-income oversample was not needed. However, our research shows that a Census tract-level income classifier or a CCW MBSF income classifier would be viable options for use in the selection of an oversample of low-income beneficiaries or some subgroup of low-income beneficiaries, if desired.

Finally, an income-to-poverty ratio indicator (IPR) for the MCBS was recently developed and will be released starting with 2015 MCBS data. Income-to-poverty ratios, e.g. income divided by the appropriate poverty threshold, are used to normalize incomes across family types and provide context for a better understanding of the depth of poverty (or lack thereof) of a family. The IPR is a useful analytic tool that can help CMS and other researchers to easily identify the percentage of Medicare beneficiaries living in different poverty levels; or how health care access and use may differ across different thresholds of interest. We provide this derived IPR variable to users in the MCBS data releases (both the Public Use File and Limited Data Set) to help them receive the proper construction and comparability to other federal health survey estimates and conduct analysis of low-income populations.

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-500-2014-000351, 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

- Wolter, K., Murphy, W., Davis N. (2015). "Improvements in the MCBS Sample Design". *Joint Statistical Meeting*, 3652-3659.
- Allen, T., Petroni, R., Singh, R. (1993). "The Effectiveness of Sampling Low Income Households in the Survey of Income and Program Participation". *US Census Bureau*. https://www2.census.gov/prod2/sipp/wp/SIPP_WP_182.pdf.
- K. Davis and C. Schoen (1978). "Health and the War on Poverty: A Ten Year Appraisal." *Washington, D.C.: Brookings Institution*.