MMRR
Medicare & Medicaid Research Review
2014: Volume 4, Number 2

A publication of the Centers for Medicare & Medicaid Services,
Office of Information Products & Data Analytics

Ever Enrolled Medicare Population Estimates from the MCBS Access to Care Files

Jason Petroski,[1] *David Ferraro,*[2] *and Adam Chu*[2]
[1]Centers for Medicare & Medicaid Services—Office of Information Products and Data Analytics
[2]Westat, Inc.

Objective: The Medicare Current Beneficiary Survey's (MCBS) Access to Care (ATC) file is designed to provide timely access to information on the Medicare population, yet because of the survey's complex sampling design and expedited processing it is difficult to use the file to make both "always-enrolled" and "ever-enrolled" estimates on the Medicare population. In this study, we describe the ATC file and sample design, and we evaluate and review various alternatives for producing "ever-enrolled" estimates.

Methods: We created "ever enrolled" estimates for key variables in the MCBS using three separate approaches. We tested differences between the alternative approaches for statistical significance and show the relative magnitude of difference between approaches.

Results: Even when estimates derived from the different approaches were statistically different, the magnitude of the difference was often sufficiently small so as to result in little practical difference among the alternate approaches. However, when considering more than just the estimation method, there are advantages to using certain approaches over others.

Conclusion: There are several plausible approaches to achieving "ever-enrolled" estimates in the MCBS ATC file; however, the most straightforward approach appears to be implementation and usage of a new set of "ever-enrolled" weights for this file.

Keywords: Medicare, survey research and questionnaire design, access to care, complex sample design, sampling weights, biostatistical methods

ISSN: 2159-0354

doi: http://dx.doi.org/10.5600/mmrr.004.02.a05

Medicare & Medicaid Research Review
2014: Volume 4, Number 2

Mission Statement

Medicare & Medicaid Research Review is a peer-reviewed, online journal reporting data and research that informs current and future directions of the Medicare, Medicaid, and Children's Health Insurance programs. The journal seeks to examine and evaluate health care coverage, quality and access to care for beneficiaries, and payment for health services.

http://www.cms.gov/MMRR/

U.S. Department of Health & Human Services
Kathleen Sebelius
Secretary

Centers for Medicare & Medicaid Services
Marilyn Tavenner
Administrator

Editor-in-Chief
David M. Bott, Ph.D.

The complete list of Editorial Staff and
Editorial Board members
may be found on the MMRR Web site (click link):
MMRR Editorial Staff Page

Contact: mmrr-editors@cms.hhs.gov

Published by the
Centers for Medicare & Medicaid Services.

All material in the *Medicare & Medicaid Research Review* is in the public domain and may be duplicated without permission. Citation to source is requested.

Introduction

The Medicare Current Beneficiary Survey (MCBS) *Access to Care* (ATC) research files are designed to provide quick access to MCBS data, making it an attractive tool for up-to-date research and statistics on Medicare beneficiary access to—and satisfaction with—care, health status, and certain health care cost/use/source-of-payment information. The expedited release of the ATC files is accomplished by focusing on cross-sectional data captured in the fall round MCBS interviews only (as opposed to including all three interviews that are conducted each year) and omitting survey-reported utilization and expenditure data, which are reported in the MCBS *Cost and Use* (CAU) files. These tradeoffs impact the ATC file's use and application, most notably requiring estimates to be made only for the "always-enrolled" Medicare population. Unlike the MCBS CAU files, ATC files must be modified by the user to yield estimates that are roughly generalizable to the "ever-enrolled" Medicare population. Researchers desire the flexibility to produce both types of estimates because there are important differences in people who have continuous enrollment throughout the year (i.e., "always enrolled") versus people who do not (i.e., "ever-enrolled"). In particular, the ever-enrolled population includes beneficiaries who died or are new to Medicare, whereas the always-enrolled population does not. In addition, Medicare administrative reports and statistics are often published on the "ever-enrolled" population; therefore, MCBS estimates would be more consistent if produced using the same criteria for reporting. Finally, the ability to make "ever-enrolled" estimates from the ATC file is important because this file includes certain

variables (e.g., preventative services, satisfaction with—and access to—care) that cannot be estimated using the CAU file.

In this paper, we review features of the MCBS complex sample design and how that design impacts the production of the ATC files, briefly explain how "ever-enrolled" estimates are achieved in the MCBS CAU files, and most importantly, consider different statistical approaches to achieve approximate "ever-enrolled" estimates from the ATC file by comparing survey estimates derived from the different approaches. We also discuss the simplicity, or "ease of use," to potential researchers, and operational considerations (e.g., the impact on release of the file, the need to obtain other data to produce the estimates, etc.) in the overall evaluation of the different options for producing "ever-enrolled" ATC estimates.

Overview of the MCBS and Data Files

The Medicare Current Beneficiary Survey (MCBS) is a continuous, in-person, longitudinal survey of nationally representative samples of the Medicare population conducted by the Office of Information Products and Data Analytics at the Centers for Medicare & Medicaid Services (CMS). The MCBS is a key source of information (both self-reported and administrative data) related to Medicare beneficiary access to—and satisfaction with—care, health status, health care cost and use, and source of payment. Comprehensive in nature, the MCBS provides insight into the Medicare program, its operations, and the people it serves; covering beneficiaries who are aged or disabled, living in the community or a facility, or served by managed care or fee-for-service (FFS).

CMS releases two sets of research files each year. The first is the annual *Access to Care* (ATC) research file. The ATC file includes content on Medicare beneficiaries' health status and functioning, access to care, satisfaction with care, and usual source of care. It is supplemented with both survey and administratively reported demographic and health insurance data. The ATC file represents a somewhat restricted view of the Medicare population because it covers beneficiaries who were enrolled in Medicare as of January of the given calendar year and who were alive during the fall round data collection period. This file is normally released a year after the fall round interviews are completed in the field. The second file is the annual *Cost and Use* (CAU) research file, which represents Medicare beneficiaries enrolled in Medicare anytime during the calendar year regardless of vital status at the end of the year. The CAU file captures survey-reported data for an entire year (up to three interviews per survey participant), includes utilization and cost information (regardless of payer), links Medicare claims data to survey-reported data, and includes an additional processing step to impute relevant data for recently-enrolled beneficiaries. These features of the CAU file mean that its sample and back-end processing are significantly different from those used to construct the ATC file. As such, the CAU data files are released roughly two years after the end of the calendar year in reference.

Overview of the MCBS Sample Design

To minimize survey costs, representative samples of beneficiaries are selected for the MCBS through a stratified, three-stage, probability sample design. The first stage of sampling involves the selection of primary sampling units (PSUs) consisting of metropolitan statistical areas (MSAs) or groups of rural counties. The second stage of sampling consists of the selection of ZIP code areas within each sampled PSU. At the third and final stage of sampling, beneficiaries within the selected ZIP codes are stratified by age and subsampled at rates designed to yield self-weighting (equal probability)

samples of beneficiaries within each of seven age groups.[1] Because of the complex features of the sample design, sampling weights and appropriate variance-estimation procedures are generally required to properly interpret estimates derived from MCBS data (Ferraro & Liu, 2005; Briesacher, Tjia, Doubeni, Chen, & Rao, 2012).

The MCBS employs a rotating panel design in which roughly one-quarter of the existing sample is released from the study each year, and a new (nationally representative) sample is selected to replace it. Under the panel rotation scheme, beneficiaries in each newly selected sample (referred to as a "panel") are interviewed three times a year for a maximum of four years. Exhibit 1 summarizes the basic features of the rotating panel design developed for the MCBS, where it can be seen that the fall data collection rounds (e.g., round 55, 58, etc.) involve three continuing panels (i.e., panels that have been in the study for at least one year) and a newly-selected panel (i.e., the supplemental sample) that is introduced in the fall round. The new panel is designed to both (a) compensate for sample losses in the continuing panels, and (b) extend coverage of the MCBS sample to recent Medicare enrollees not represented in the older panels. It is the latter purpose that is critical to ensuring that annual estimates derived from the MCBS provide substantially complete coverage of the current Medicare population. However, the timing of sample selection and the prospective nature of data collection preclude the collection of relevant survey data for the most recent Medicare enrollees, leaving a small gap in coverage of MCBS data. In the following sections, we describe how this gap arises and how it is handled in the estimates derived from the MCBS.

[1] Age groups: (a) 0–44, (b) 45–64, (c) 65–69, (d) 70–74, (e) 75–79, (f) 80–84, (g) 85+.

Timing and Coverage of MCBS Sample Selection

The universe list of beneficiaries from which individuals are selected for the MCBS is referred to as the "sampling frame." The creation of the sampling frame carries certain limitations that are worth noting because they limit generalizing the ATC

Exhibit 1. Rotating Panel Scheme for the MCBS

Year	2009			2010			2011			2012			2013		
Period	W	S	F	W	S	F	W	S	F	W	S	F	W	S	F
Round*	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67
	2005 C	2005 X													
	2006 C	2006 C	2006 C	2006 C	2006 X										
	2007 C	2007 C	2007 C	2007 C	2007 C	2007 C	2007 C	2007 X							
	2008 C	2008 C	2008 C	2008 C	2008 C	2008 C	2008 C	2008 C	2008 C	2008 C	2008 X				
			2009 P	2009 C	2009 C	2009 C	2009 C	2009 C	2009 C	2009 C	2009 C	2009 C	2009 C	2009 X	
						2010 P	2010 C	2010 C	2010 C	2010 C	2010 C	2010 C	2010 C	2010 C	2010 C
									2011 P	2011 C	2011 C	2011 C	2011 C	2011 C	2011 C
												2012 P	2012 C	2012 C	2012 C
															2013 P

NOTE: * Panels are designated by the year in which they are first selected. New panels are introduced in fall data collection rounds and are denoted by a "P" in the above exhibit. A panel remains in the study for 12 rounds of data collection. Continuing panels are denoted by a "C" or "X" in the above exhibit, where an "X" indicates the round in which the panel exits the study. "W", "S", and "F" correspond to MCBS' three annual rounds, or winter, summer and fall.

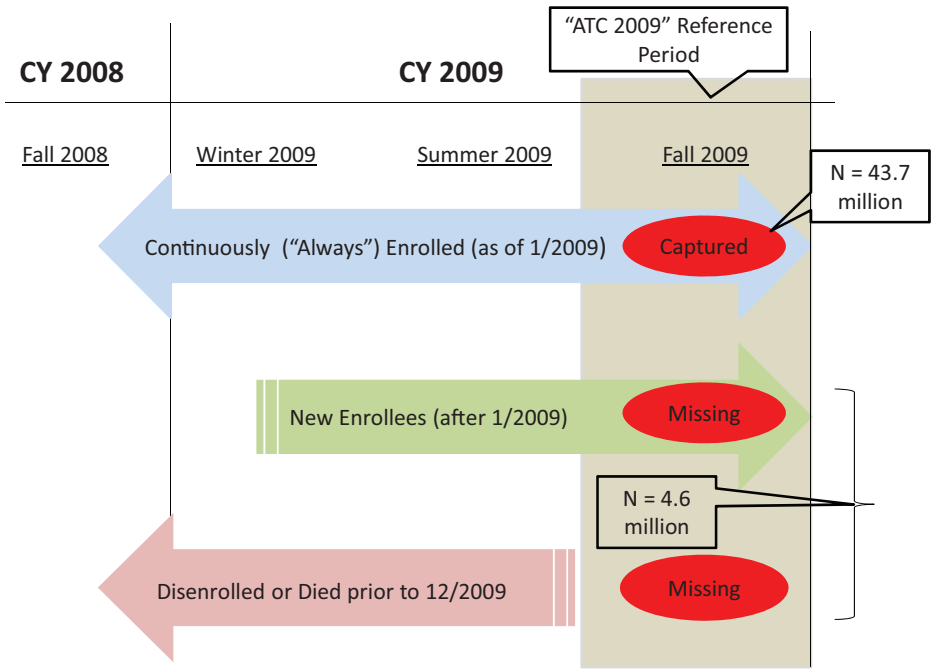
SOURCE: Centers for Medicare & Medicaid Services (CMS), Medicare Current Beneficiary Survey (MCBS), (2014).

data to the "ever-enrolled" population (Centers for Medicare & Medicaid Services, 2011a). The MCBS sampling frame is derived from available administrative data, or more specifically, an extract of CMS' enrollment data base (EDB). Although these data can be used to accurately determine Medicare eligibility (and therefore designate a sample for the MCBS), there is a slight lag in the reporting of newly-enrolled and recently dis-enrolled or deceased beneficiaries in the data file. Because we need to select and process the sample before going to the field, this time lag poses operational and analytic challenges for the MCBS, most importantly, the underrepresentation of newly enrolled Medicare beneficiaries. In the following sections, we describe how these aspects of constructing the sampling frame (a) restrict estimates derived from the ATC file to the "always enrolled" population, (b) are handled in the production of the annual MCBS data files, and (c) can be accounted for through the use of various statistical techniques designed to address the underrepresentation of the recent enrollees and those who were enrolled at the beginning of the year, but died or dis-enrolled prior to the fall interview.

"Always" vs. "Ever-enrolled" Estimates: Two Views of the MCBS

CMS releases two complementary restricted use data files each year, ATC and CAU. The two types of data files represent slightly different "views" of the Medicare population. The ATC data file, which is tied to the corresponding fall data collection round (see "ATC 2009 Reference Period" in Exhibit 2), is generalizable to what is referred to as the "always enrolled" population for the particular year. As the name implies, the always-enrolled population includes those beneficiaries who were continuously enrolled in Medicare throughout a given calendar year, and thus is missing beneficiaries who died before the end of the year (i.e., before completion of the fall round interview) and those who were first entitled and enrolled in Medicare after

Exhibit 2. Access to Care File: Illustration of Always vs. Ever Enrolled Populations



SOURCE: Centers for Medicare & Medicaid Services (CMS), Medicare Current Beneficiary Survey (MCBS), (2014).

January of the given year. As illustrated in Exhibit 2, the ATC 2009 file represents the approximately 43.7 million always-enrolled Medicare beneficiaries in 2009. The ATC file does not represent 4.6 million beneficiaries[2] who had partial enrollment/entitlement in Medicare in 2009 because the newly enrolled beneficiaries were not captured in the sampling frame at the time the sample was drawn, or the beneficiary was enrolled at the start of the reporting year, but dis-enrolled or died prior to receiving a fall interview.

CMS addresses this limitation in coverage when it prepares the annual CAU research file using auxiliary data and augmentation procedures to enable generalizing to both the "ever" and "always" enrolled Medicare populations. Since the CAU data releases come from five MCBS panels (see Exhibit 3) comprised of three continuing panels and two partial panels (representing newly enrolled beneficiaries), the CAU data files are released approximately one and a half years after the release of complementary ATC data. This process is described in more detail below.

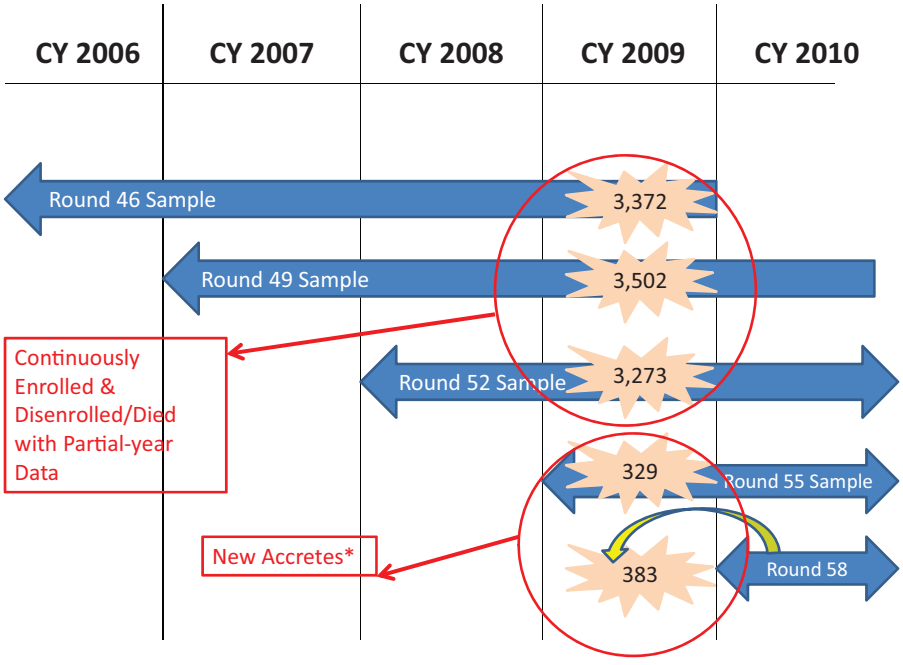
How the Problem is Handled in the Annual CAU Data Files

We use an imputation approach in the CAU files to account for lack of survey-based data on new accretes.[3] Beneficiaries who were enrolled in Medicare, but died or dis-enrolled during the calendar year, are included in the CAU files as

[2] The 4.6 million Medicare beneficiaries not captured in the ATC file represent roughly 10% of the total Medicare population for the year. Of the 4.6 million Medicare beneficiaries noted in 2009, roughly 1.6 million died or disenrolled, and 3 million were newly enrolled.

[3] The CAU data file is not as deficient of Medicare beneficiaries who died or became dis-enrolled during the year as the ATC file is because, unlike the ATC file, the CAU file follows the beneficiary for an entire year and is continuous.

Exhibit 3. 2009 CAU Panels: Use of "New Accretes" from 2009 & 2010 Samples



NOTE: Counts given in chart are unweighted sample counts in CY 2009 CAU file. * CY 2009 New accretes are not present in Round 46, 49, or 52; therefore, it is necessary to pull just newly enrolled persons from Round 55 and Round 58.
SOURCE: Centers for Medicare & Medicaid Services (CMS), Medicare Current Beneficiary Survey (MCBS), (2014).

"respondents" if sufficiently complete survey data were obtained prior to death/dis-enrollment.[4]

To illustrate how new accretes are accounted for in CAU files, consider the 2009 CAU data release. This data file includes two panels of recent accretions (including new accretes who died in 2009): those who first enrolled in Medicare in 2008 (and were sampled in 2009) and those who first enrolled in 2009 (and were sampled in 2010). Although a rich set of data is available for these beneficiaries from CMS administrative files, e.g., claims data, there are no self-reported cost and utilization survey data for these cases for reasons discussed earlier. Leaving the survey items as missing would understate the corresponding true population value when the sample is inflated (i.e., statistically weighted) to population levels. Thus, we derive, i.e., impute, values of the missing data items for the new enrollees using information from a similar beneficiary in the sample for whom survey data were collected.

The process of identifying a similar beneficiary for this purpose is referred to as "ghost donor imputation." Under this procedure, we find a donor record for each new enrollee (also referred to as a "ghost") by matching on demographic characteristics as well as information collected in the initial fall round interview, such as insurance plan, type of residence, health status, access to care, and utilization (Centers for Medicare & Medicaid Services, 2011b). Because the fall round survey data are critical to the imputation process, it is necessary to delay production of the CAU file until fall round survey data are collected for the recently-sampled new enrollees. The types of data from the matched donor record that are transferred to the ghost record include annual person-level utilization of non-covered services, cost and payment data for non-covered services, and utilization and expenditure data at the person level for covered services for persons in Medicare health plans (i.e., Medicare Advantage Organizations). We believe this procedure, when aggregated with the results for the continuing panels in the annual CAU data releases, provides the best available survey-based estimates of annual cost, utilization, and expenditures for the ever-enrolled Medicare population.

How the Problem is Handled in the Access to Care Data Files

The procedures discussed above to make CAU data files more generalizable to the ever-enrolled population cannot easily be applied to the ATC files, and there has been no formal evaluation of alternative approaches for doing so in the ATC files. A key limiting factor is that the ATC files are designed to provide researchers with a reasonably timely snapshot of the Medicare population and, therefore, auxiliary data cannot be used for new enrollees as it can with the CAU files.

The ATC data files currently include person-level sampling weights that can be used to inflate the respondent data to represent the always-enrolled population. These weights apply to beneficiaries in a given calendar year who were enrolled in Medicare in January of that year, remained in Medicare, and were alive through the end of the fall round data collection. In other words, beneficiaries who died during the year or who accreted into Medicare after January of the given year do not have weights in the ATC files as currently constructed. As long as this is understood, the use of weighted ATC data can provide valuable information about the current state of the always enrolled Medicare population (which is 90% of the Medicare population).

However, there is utility in having the ATC population match what is covered by the CAU data releases and, therefore, also coincide with

[4] Essentially, a valid interview must be completed by proxy for the deceased individual.

administratively reported "ever-enrolled" Medicare statistics.[5] In order to produce estimates from the ATC data files that are approximately comparable with those derived from the corresponding CAU files, CMS analysts have used an internal process that supplements the ATC data and applies statistical adjustments to the traditional ATC weights. These procedures, not currently available to outside researchers, include two main steps. The first is to augment the ATC data files with sample members in the continuing panels who died during the year (and for whom MCBS data are available from the prior fall round). The second step is to post-stratify, i.e., ratio-adjust, the traditional ATC weights of the fall round respondents to account for the new accretions (specifically, those accreting into Medicare after January of the given ATC year and who have survived through the end of the year). The special weighting adjustments are designed to do what the ghost donor matching process does for the CAU data files, but through weighting factors rather than by imputation. In what follows, we evaluate this procedure and compare it with other possible approaches designed to account for new enrollees in the ATC data files.

Methods to Account for Recent Enrollees in ATC Data Files

As mentioned above, CMS currently uses a post-stratification weighting procedure to adjust the traditional ATC sampling weights to represent the ever-enrolled population. The current method uses population counts, derived from CMS administrative files as benchmarks, to adjust the weighted counts from the ATC sample so that they equal the corresponding ever-enrolled population counts by selected age groups. This adjustment, which we refer to as Method 1, does not explicitly account for the accretion year of the beneficiary, i.e., the year in which the beneficiary enrolled in Medicare. Thus, an obvious improvement over the current approach would be to include accretion year as a post-stratification variable along with other demographic characteristics. Failure to account for accretion year in the weight adjustments can distort the distribution of beneficiaries with respect to accretion status. Exhibit 5 demonstrates this where approximately 15% of the (weighted) sample based on the simple post-stratification method (ATC-1) would be estimated to have accreted in 2007 or later, compared to 19% in the CMS administrative files (and other methods where accretion year is explicitly used to develop the weights). Thus, for survey outcomes that tend to be correlated with accretion year, sample-based estimates using the simple post-stratification weights potentially could be biased compared with the other methods. We refer to this modification as Method 2. A third possibility would be to apply an imputation procedure similar to what is currently being used to create the CAU data files. We refer to this as Method 3. While the effectiveness of the imputation procedure for the new enrollees may be limited by the lack of timely and relevant matching variables, an advantage of this approach is that it is analogous to the procedures currently used to construct the CAU files. We provide some additional details about the three methods under consideration below.

Method 1: Weighting Adjustments with Simple Post-stratification

As described in Appendix A of *The Characteristics and Perceptions of the Medicare Population: Data from the 2008 Medicare Current Beneficiary Survey* (Centers for Medicare & Medicaid Services, 2013), CMS accounts for beneficiaries

[5] "Administratively reported" refers to CMS program statistics on the Medicare population reported from administrative sources used for payment, enrollment, etc.

Exhibit 4. Comparison of estimates under alternative procedures and associated average absolute relative differences (AARD)

	Compare weighting for Methods 1, 2P, and 2				Compare Method 2 with Method 3	
	Average relative difference (AARD)				AARD	
	P-value[1]	1 vs. 2P	1 vs. 2	2P vs. 2	P-value[1]	2 vs. 3
Socio-economic characteristics						
Race	NS	0.5	0.2	0.6	NS	2.2
Ethnicity	NS	0.2	0.2	0.0	*	2.1
Number of children	*	0.6	0.7	0.1	NS	0.8
Educational attainment	***	1.1	1.0	0.1	NS	0.8
Marital status	***	1.5	1.3	0.4	***	2.1
Personal income	*	0.6	0.3	0.3	*	0.8
Selected characteristics from administrative files						
Community/Facility Status	***	1.0	0.8	0.2	NS	2.1
Medicare status code	***	0.5	0.5	0.0	***	5.8
Medicaid elibility	***	1.3	1.0	0.3	NS	0.4
Part D plan type - December	**	0.8	0.5	0.2	NS	0.6
GHP plan type - January	***	0.9	0.9	0.2	***	0.8
GHP plan type - December	*	0.8	0.8	0.1	*	0.7
LIS indicator - December	***	1.1	0.7	0.4	NS	0.1
Selected health-related characteristics						
General health status	*	0.6	0.6	0.1	***	2.2
Limitation of activities due to health	NS	0.4	0.4	0.1	**	2.5
Told has hypertension	NS	0.3	0.1	0.2	NS	0.1
Told has Alzheimers	*	0.5	0.3	0.1	NS	0.4
Had mammogram	NS	0.1	0.0	0.1	*	0.8
Had PAP smear	NS	0.1	0.1	0.0	**	1.5
Had digital prostate exam	NS	0.1	0.0	0.1	NS	0.3
Had blood test for prostate cancer	*	0.7	0.9	0.2	NS	0.3
Had flu shot	NS	0.0	0.0	0.0	***	1.3
Had pneumonia shot	***	0.8	0.9	0.1	***	2.8
Selected numeric estimates						
Part A reimbursements	**	1.8	1.9	0.1	NS	1.0
Part B reimbursements	*	1.1	1.0	0.1	NS	1.2
Part B premiums paid	NS	0.2	0.1	0.1	NS	0.4
No. of Inpatient stays	***	1.9	2.1	0.2	NS	0.3
No. of Inpatient days	**	2.1	2.4	0.2	NS	0.6
Inpatient charges	NS	1.4	1.7	0.3	NS	0.7

(Continued)

Exhibit 4 Continued. Comparison of estimates under alternative procedures and associated average absolute relative differences (AARD)

	Compare weighting for Methods 1, 2P, and 2				Compare Method 2 with Method 3	
	Average relative difference (AARD)				AARD	
	P-value[1]	1 vs. 2P	1 vs. 2	2P vs. 2	P-value[1]	2 vs. 3
Number of SNF stays	NS	1.1	1.0	0.2	NS	0.5
Number of SNF days	*	1.2	1.3	0.0	NS	0.5
SNF charges	**	1.5	1.2	0.3	NS	0.1
Number of HHA Visits	NS	0.9	0.6	0.4	NS	3.1
HHA covered charges	NS	0.3	0.0	0.3	NS	1.6
Total outpatient bills	**	1.1	1.0	0.1	NS	1.4
Total physician claims	*	0.7	0.7	0.0	NS	0.4

NOTES: [1] P-value corresponding to test of differences in estimates under alternative methods. NS means not significant at 0.05 level; * means significant at 0.05 level; ** means signficant at 0.01 level; *** means significant at the 0.0001 level.

SOURCE: Data from the 2009 Medicare Current Beneficiary Survey (MCBS, Access to Care files.

not represented in the traditional ATC data files in the following ways. To account for persons who died during the year (and for whom MCBS data are available from the prior fall round), CMS includes the deceased cases in their tabulations, using special sampling weights constructed specifically for this adjustment. To account for the recent accretions for whom the fall round survey data are not available, CMS inflates the traditional ATC weights of the fall round respondents in the relevant MCBS panels using a simple age-group specific ratio adjustment factor defined by the formula: $F_g = (E_g - D_g)/A_g$., where E_g = the number of beneficiaries in age group g who were ever enrolled during the year (derived from CMS administrative files), D_g = the estimated number of beneficiaries in age group g who died during the year (weighted estimate derived from the MCBS sample), and A_g = the number of beneficiaries in age group g who were continuously enrolled during the year (derived from CMS administrative data). The adjustment factor is computed for each of seven age categories (g = 1, 2, ... 7), and applied to the standard ATC sampling weight to obtain the corresponding ever-enrolled weight. It should be noted that the procedures and weights produced under this method are not available to the public and are currently used by CMS internally to create publicly reported data tables from the ATC file. However, these methods (or slight modifications to this approach) hypothetically could be applied by MCBS users and/or statisticians to reproduce the adjusted weights using the ATC files and other information available from CMS.

Method 2: Weighting Adjustments using Ratio Raking

A variant of the simple post-stratification method uses age-group specific weight adjustments (like Method 1), except that the adjustments are developed by relating the standard ATC weights to "known" population counts of the ever-enrolled Medicare population using a technique known as ratio-raking (or simply "raking"). With this technique, the ATC weights are iteratively adjusted to two sets of known population counts referred to as "raking dimensions." The first raking dimension consists of subgroups defined by sex, race (Black versus Others), and age group (under

45 years old, 45 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85 or older). The second raking dimension consists of categories of beneficiaries defined by year-of-accretion (e.g., for the 2009 ATC data file, the categories are: 2004 or earlier, 2005, 2006, and 2007 or later) and age group. Use of accretion year to adjust the ATC weights is important because it allows for the differential adjustment of the weights of the recent enrollees. This procedure is potentially less biased than Method 1 because it inflates the weights of those beneficiaries in the sample that are most like the under-represented recent accretions to a greater degree than those who have been enrolled for many years. Unlike Method 1, if Method 2 were to be implemented, CMS would create a new set of weights and deliver these to MCBS users.

In addition to Method 2, we also included in the analysis a preliminary version of Method 2 that did not include the more detailed post-stratification by race and sex, but did include post-stratification by accretion year. This version of Method 2, referred to in the subsequent table as "Method 2P," was designed to gauge the extent to which the use of other variables, such as sex and race, in the post-stratification adjustments would affect the weighted estimates.

Method 3: Imputation

This method identifies donor records for the recent enrollees from which selected data items are "borrowed" and transferred to the corresponding new enrollee record. The donors are selected to match the characteristics of the new enrollee (borrower) based on characteristics available in the sampling frame. The imputation approach is similar to the ghost donor matching procedures used to impute survey data for the new enrollees (ghosts) in the annual CAU data sets. Once a donor is identified for a particular new enrollee, all of the relevant data for the donor is transferred over to the new enrollee data record. Since the access-to-care (ATC) data sets are developed a year in advance of the corresponding cost and use (CAU) data sets, most of the relevant characteristics used for CAU ghost donor matching (e.g. recent profiles of Medicare use) are not available for ATC imputation. The only data available for ATC imputation are demographic and administrative data included in the sampling frame. The technique used to identify the set of donors is referred to as "hot deck" imputation.

The first step in the hot-deck imputation process is the separation of the ATC sample into two groups: (1) those with responses (i.e., the always-enrolled sample cases normally included in the ATC data files), and (2) the recent enrollees for whom no fall round survey data are available. Next, a number of matching variables or "keys" are derived based on information available for both groups. Each matching key is assigned a rank from high to low according to desirability, and collectively the keys make up what is referred to as the matching hierarchy. Cells are then formed by cross-classifying a specified number of keys. For example, if key X has 3 distinct values and key Y has 5 distinct values, then the two keys jointly specify 15 cells. Under the given matching hierarchy, the boundaries between the cells are specified to be either "hard boundaries" or "soft boundaries." If a particular key is specified as a hard boundary, then a donor record must be selected from the available records with the same value of the key as the record (new enrollee) with missing data. On the other hand, if the key is specified as a soft boundary, and a donor with the same value of the key as the new enrollee record cannot be found, then the search is expanded to include potential donors in neighboring cells adhering as much as possible to the matching hierarchy established for the imputation.

For the ATC ghost donor imputation used in this evaluation, the hard boundary keys were age group (under 45 years old, 45 to 64, 65 to 69, 70 to 74, 75 to 79, 80 to 84, and 85 or older), race (Black versus Others), sex, and type of plan (Medicare health plans versus others). The soft boundary keys were defined by categories of accretion-year, Medicare enrollment status (enrolled in Part A and B versus others), metropolitan status (MSA versus non MSA), and Census region.

If Method 3 is implemented, CMS would provide an additional set of records for the ghosts (along with imputed data) in its release of the ATC file.

Results

We produced estimates from the 2009 ATC data using each of the methods described above. We first compared these estimates to one another and then contrasted them with 2009 CMS administrative data. The data used in the first part of the analysis included selected variables on demographic characteristics, health status, enrollment characteristics, and Medicare reimbursements and utilization. We limited data used to benchmark the estimates from each of the methods against the corresponding CMS administrative data to variables that were available both through the ATC file and CMS administrative files. All variables used in the analysis are available in the 2009 ATC research file. We excluded beneficiaries who died during the calendar year from the comparisons because this paper focuses on the different ways of handling new accretes and there were no differences in the methods for handling those who died.

We summarize the results of comparisons of the estimates for the 35 variables derived from the different methods in Exhibit 4. The first column in the table to the right of the variable name gives the result of a statistical test comparing the estimates derived under the three weighting schemes.[6] For the categorical variables, the weighted distributions of the characteristic under the different methods were tested for similarity using the Rao-Scott chi-square statistic.[7] For the continuous variables listed in the bottom part of the table, t-tests were used to test the difference between the means under the different methods. The analyses were performed in SAS using the SURVEYFREQ and SURVEYREG procedures[8] to properly account for the complex features of the MCBS sample design such as stratification, clustering, and unequal weighting.[9] The symbol "NS" means that the estimates being compared are not statistically significant, whereas the symbols "*," "**," and "***," mean that the differences are significant at the 0.05, 0.01, and 0.0001 levels, respectively.[10]

Also shown in the table is the average absolute relative difference (AARD) between pairs of estimates. This statistic provides an *ad hoc* overall summary measure of the magnitude of the difference relative to the size of the estimate. Rather than showing the individual differences for each level of a categorical variable, we present the AARD to provide a rough summary view of the difference across all levels of the variable. For example, for

[6] The test is an overall test comparing the three methods jointly. A significant result simply means that not all three methods are the same. In practice, we are mainly interested in comparing 2P or 2 against method 1, but did not conduct formal individual pair-wise tests.

[7] Rao, J. & Scott, A. (1981). The analysis of categorical data from complex sample surveys: chi squared tests for goodness of fit and independence in two way tables. *Journal of the American Statistical Association*, 76, 221–230.

[8] See http://support.sas.com/documentation/cdl/en/statugsurveysamp/61762/PDF/default/statugsurveysamp.pdf

[9] The estimates being compared are not independent; however, we were able to reflect the lack of independence in the statistical tests through the appropriate use of the available MCBS replicate weights.

[10] A conservative adjustment, such as a Bonferroni adjustment, could be applied to account for multiple comparisons. For example, if g comparisons are being made, then to ensure an overall study-wide alpha of 0.05, an individual comparison would have to result in a P-value < .05/g to be deemed statistically significant. Given the ad hoc and exploratory nature of the analysis, we did not apply an adjustment for multiple comparisons.

Exhibit 5. Estimates of Medicare Population Counts from 2009 ATC file versus 2009 CMS Administrative Data

		CMS Administrative Data	ATC–1 (Simple post-strat.)		ATC–2P (Raking adj.–prelim. version)		ATC–2 (Raking adj.–final version)		ATC–3 (Imputed)	
Beneficiary Characteristics		%	%	Diff. from Admin.	%	Diff. from Admin.	%	Diff. from Admin.	%	Diff. from Admin.
N		*46,960,272*	*46,739,586*		*46,739,586*		*46,739,580*		45,867,308	
Age	< 65	16.83	16.83	0.00%	16.83	0.00%	16.82	–0.06%	17.43	3.56%
	65-74	44.45	44.25	–0.46%	44.25	–0.46%	44.24	–0.47%	43.70	–1.69%
	75-84	27.06	27.16	0.36%	27.16	0.36%	27.15	0.33%	26.97	–0.32%
	85+	11.65	11.77	0.99%	11.77	0.99%	11.76	0.94%	11.90	2.15%
Sex	Male	44.71	45.01	0.67%	45.20	1.10%	44.52	–0.42%	45.05	0.77%
	Female	55.29	54.99	–0.54%	54.80	–0.89%	55.48	0.34%	54.95	–0.62%
Race	White	77.82	77.18	–0.83%	77.26	–0.72%	77.11	–0.91%	76.84	–1.26%
	Non-White	22.02	22.82	3.65%	22.74	3.27%	22.89	3.95%	23.16	5.18%
Accretion Year	2006	5.69	6.70	17.79%	5.66	–0.54%	5.66	–0.54%	5.69	0.02%
	2007 or later	19.03	15.37	–19.22%	18.87	–0.82%	18.87	–0.82%	19.18	0.80%
Medicare Plan	Yes	23.25	25.59	10.08%	25.42	9.35%	25.44	9.42%	26.00	11.81%
	No	76.75	74.41	–3.05%	74.58	–2.83%	74.56	–2.85%	74.00	–3.58%
Part D	Yes	58.07	59.83	3.03%	59.44	2.36%	59.57	2.58%	60.02	3.35%
	No	41.93	40.17	–4.20%	40.56	–3.28%	40.43	–3.58%	39.98	–4.64%
Medicare Entitlement Status	Aged	82.14	81.70	–0.53%	81.86	–0.34%	81.86	–0.34%	80.09	–2.49%
	ESRD	0.53	0.82	55.30%	0.79	49.00%	0.79	49.06%	0.80	50.86%
	Disabled	17.56	18.14	3.30%	17.98	2.37%	17.98	2.39%	19.76	12.52%

NOTE: CMS administrative statistics and ATC estimates exclude beneficiaries that died during the year.

SOURCE: 2009 CMS Administrative data from CMS' Chronic Condition Warehouse (CCW), 2009 Beneficiary Annual Summary File.

each level of a particular categorical variable, we compute the absolute value of the difference between estimates. The level-specific absolute difference is then divided by the average of the estimates derived from the two methods to obtain the corresponding absolute relative difference. A simple average of the absolute relative differences over all levels of the variable is then computed to obtain the AARD. We use the AARD to help discern whether certain differences are meaningful, even though they may be statistically significant. For numeric variables, the AARD is simply the absolute difference between estimated means divided by the average of the two estimates.

We first compared the estimates derived under the three weight-adjustment methods 1, 2P, and 2. , The majority of the differences (22 of the 35) are moderately to highly significant (Exhibit 4), which suggest that the estimates produced by the three weighting methods generally do not give similar results. However, looking at the average absolute relative differences, we see that the estimates from Methods 2P and 2 are generally similar, whereas the estimates for Method 1 differ appreciably from estimates derived from both Methods 2P and 2, though the differences are moderately small and are largest for the numeric variables. The latter result seems reasonable as both Method 2P and Method 2 explicitly account for accretion year in the weight adjustments, while Method 1 does not. The fact that Method 1 produces estimates (see Exhibit 4) that are generally different from those produced under either Method 2P or Method 2 suggests that it may be less effective in accounting for the most recent enrollees than the other methods. Although Methods 2P and 2 produce similar results, we prefer Method 2 because it is better at reproducing Medicare population counts by sex and race (Black versus Others). The second set of analyses we conducted involved comparisons between Method 2 (the "preferred" weighting approach) and Method 3 (the imputation approach). Of the 35 variables included in the analysis, the weighted estimates for 24 (including all of the numeric variables shown in the bottom part of Exhibit 4) do not differ significantly by method. This indicates that Method 3 often, but not always, produces results that are similar to Method 2 based on statistical testing. Apart from some categorical variables where the base of the relative difference is very small (resulting in a large relative difference), the magnitude of the differences between methods is generally small. Because the imputation method used in this evaluation is not the same as the procedures normally used to create the CAU files, it is likely to be less effective at producing consistent estimates for the ever-enrolled population for some types of categorical variables. On the other hand, differences for the numeric variables considered in this analysis are not statistically significant, and all are generally small. Based on these results, we tentatively conclude that both Method 2 and Method 3 are somewhat better at reflecting the ever-enrolled population than Method 1, although neither is likely to produce estimates that are exactly equivalent to those derived from the standard CAU methodology.

Discussion

As demonstrated in this paper, there are a number of feasible alternatives towards achieving "ever-enrolled" estimates from the MCBS ATC file, including some that are relatively straightforward and convenient to implement. Evaluation of each of the alternatives primarily involved a comparison of the estimates produced by each method for a limited set of variables, but there are also other considerations worth discussion when determining the best approach moving forward.

Results from the comparison of estimates derived from weighting Methods 1, 2P, and 2

showed that while there were statistically significant differences in the results produced under the different methods, the average absolute relative differences are generally small (usually 1.0 percent or less). Comparing Method 2 (the preferred weighting method) with Method 3 (the imputation method), we found no statistically significant differences between the estimated means of the numeric variables considered. However, we did find a number of significant differences for the categorical variables, indicating that the ever-enrolled estimates are dependent to some extent on the methodology and/or approach used. When the three methods were benchmarked against CMS administrative data, Methods 2 and 2P produced estimates closest to those reported in CMS administrative records as expected, but there were some differences for Method 3, which were likely due to the limited data available for imputation (see Exhibit 4). In terms of methodological soundness, all three methods and variants (1, 2 & 2P, & 3) employ well-known and well-used statistical techniques.[11] Methods 2 and 3 are expected to be somewhat more robust than Method 1 because they explicitly account for year-of-accretion in the adjustments and, as seen in the comparison to CMS Administrative data in Exhibit 5, do a better job compensating for the absence of new accretes in the ATC file.

In addition to a review of the estimates provided by each of these alternatives, we also consider the relative "ease of use" of each of the alternatives in applied research using the MCBS. For example, Method 1 is an example of a technique that could be applied currently by external users without support from CMS. MCBS users looking to adjust "always enrolled" weights using Method 1 would need to carry out several steps and adjustments. From a practical sense, therefore, Method 1 is much more burdensome for the user than simply using a new set of weights (Method 2 & 2P) or imputed data (Method 3). If Method 1 was applied by an MCBS user, there is also a greater potential that the user may arrive at different results than CMS or other users making adjustments. Methods 2, 2P, and 3 should yield consistent results across all users since CMS would potentially provide these adjustments. For these reasons, we identified Method 1 as a more complicated option for users and, therefore, we consider it to be less favorable from this perspective. The ATC file is meant to provide quick access to MCBS data and, for this reason, we reviewed each of the alternatives based on their operational simplicity to CMS. We considered Methods 1, 2, and 2P to be the simplest operationally since they require the least amount of auxiliary data, are relatively quick to implement once programmed, and would not delay the release of the ATC files under the current schedule. Method 3 was regarded the most complex alternative from an operational standpoint because of its use of complex imputation methods, more extensive use of auxiliary data, and potential to delay ATC research file release.

Exhibit 6 summarizes the authors' evaluation of estimates from each of the alternatives, as well as the findings pertaining to ease of use and operational simplicity. This study suggests that Method 2 or 2P is the most attractive overall approach (of the methods considered) for making "ever-enrolled" estimates using the ATC file, from a methodological, ease of use, and operational standpoint.

This study has several limitations that should be noted. First, the statistical analysis was limited to a relatively small subset of the variables that are included in ATC data files; hence, there is no assurance that the patterns noted in Exhibit 4 or Exhibit 5 will also hold for other variables not

[11] Examples of ongoing studies that employ weighting techniques similar to those used in this paper include the Medical Expenditure Panel Survey—MEPS (http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr24/mr24.pdf) and the Current Population Survey—CPS (http://www.census.gov/prod/2006pubs/tp-66.pdf).

Exhibit 6. Results of Comprehensive Evaluation of Estimation Alternatives

Methods: Criteria	Method 1: Simple Post-stratification[1]	Method 2 & 2P: Ratio Raking Adjustments	Method 3: Imputation
Methodological/Statistical Soundness	**	***	***
Ease of Use to Researcher	*	***	***
Operational Simplicity	***	***	**

NOTES: Rankings: * = poor; ** = fair; *** = excellent.

[1]This method would be created by the user, where as other methods would be provided by CMS.

SOURCE: Centers for Medicare & Medicaid Services (CMS), Medicare Current Beneficiary Survey (MCBS), (2014).

included in the study. Second, the review of ease of use and operational simplicity represent a somewhat subjective view of the authors, yet we feel qualified to make this analysis given our involvement in the MCBS project. Third, there currently is no true test of what the best approach would be to estimate the characteristics of "ever-enrolled" beneficiaries based on a sample survey, such as the MCBS, given the limitations of sample frame construction and data collection. More research is needed to address the limitations discussed above; however, this evaluation does provide an important starting point for development of robust and appropriate weights for analysis of the ever-enrolled population using MCBS ATC data.

Disclaimer

The authors have been requested to report any funding sources and other affiliations that may represent a conflict of interest. The authors reported that there are no conflict of interest sources.

Correspondence

Jason Petroski, Ph.D., Centers for Medicare & Medicaid Services—Office of Information Products and Data Analytics, 7500 Security Blvd., B2-29-04, Baltimore, MD 21244, jason.petroski@cms.hhs.gov, Tel. 410-786-4681, Fax. 410-786-5515

Acknowledgment

The authors would like to acknowledge Richard Apodaca for his support of this study and over 20 years of contributions to the MCBS project.

References

Briesacher, B. A., Tjia, J., Doubeni, C. A., Chen, Y., & Rao, S. R. (2012). Methodological Issues in Using Multiple Years of the Medicare Current Beneficiary Survey. *Medicare & Medicaid Research Review*, *2*(1). http://dx.doi.org/10.5600/mmrr.002.01.A04

Centers for Medicare & Medicaid Services.(2013). *2011 Characteristics and Perceptions of the Medicare Population*. Baltimore, MD. Retrieved from the CMS Web site at: http://www.cms.gov/Research-Statistics-Data-and-ystems/Research/MCBS/Data ables.html

Centers for Medicare & Medicaid Services (CMS). (2011a). *Medicare Current Beneficiary Survey – CY 2009Access to Care File Documentation.* Section 5: Sample Design and Guidelines for Preparing Statistics. Baltimore, MD.

Centers for Medicare & Medicaid Services. (2011b). *Medicare Current Beneficiary Survey—CY 2010 Cost and Use File Documentation.* Introduction. Baltimore, MD.

Ferraro, D., & Liu, H. (2005). Uses of the Medicare Current Beneficiary Survey for Analysis Across Time. *Proceedings of the Survey Research Methods Section of the American Statistical Association* [CD-ROM], Alexandria, VA: American Statistical Association.