CONSIDERATIONS FOR ANALYSIS OF THE MEDICARE CURRENT BENEFICIARY SURVEY (MCBS) ACROSS TIME

Julie O'Connell, Adam Chu, Westat, Inc., R. Clifton Bailey, HCFA
Julie O'Connell, Westat, Inc., 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Cross-sectional analysis; Comparisons of estimates across time; Longitudinal analysis; Generalized variance; Correlation across time; Repeated measurements of subsamples; Health policy statistics.

Disclaimer: The opinions are those of the authors and do not represent policies or opinions of HCFA.

1. Introduction

The MCBS is a continuous, multi-purpose panel survey of Medicare beneficiaries that is intended to provide data on health care utilization and costs. The survey is sponsored by the Health Care Financing Administration (HCFA). Data collected in the survey include Medicare covered services as well as noncovered services. Data on the associated costs are obtained by source of payment.

To date, seven data files have been released that contain data from the MCBS as well as administrative information from HCFA. Guidance for using these files has been provided mainly for cross-sectional analyses in the form of user's guides that accompany each data file. Inquiries from MCBS data users indicate that there has been a growing interest in using the data files to analyze year-to-year change, particularly as more data have become available. As future files are released, the implementation of even relatively simple longitudinal analyses will become increasingly difficult.

There are a number of factors that complicate the analysis of year-to-year changes in the MCBS data files. First, the population of inference is dynamic, incurring losses as a result of deaths and growth as a result of beneficiaries that are newly eligible each year. Thus, standard longitudinal analysis procedures, in which the population of inference is treated as fixed over time, do not reflect changes that result from the natural evolution of the Medicare population.

In addition, there are also changes in the MCBS sample of respondents over time. Sample losses occur as a result of deaths of sample respondents and of nonresponse in later rounds; sample augmentation occurs as the result of sample supplements that are added each year. Weighting adjustments to account for these changes in the sample result in sample weights for analyzing data in later rounds that are different from those that are appropriate for analyzing data at early rounds. Furthermore, samples of respondents at different points in time weight up to different populations.

A standard practice for longitudinal analysis is to treat the population of inference as fixed over time, including persons in the population who may die at a later time. In this framework, an estimate of year-to-year change applies only to the population of beneficiaries who were initially eligible to be sampled. The resulting estimates can be useful in studying gross changes for a fixed population of survivors; however, such approaches do not completely describe changes for the total Medicare population that result from changes in the population composition.

Another type of estimate, that does reflect the changing composition of the Medicare population, is estimate of net change. These estimates can be obtained by simply computing the difference between two "cross-sectional" estimates, using appropriate cross-sectional weights that are provided with each data file. For example, estimates of net change in annual Medicare utilization and reimbursements are among the most important statistics of this type to be derived from the MCBS. To properly interpret net differences, however, estimates of the corresponding standard errors are required. The current guidance on variance estimation that is provided with each data file is not adequate unless one is willing to treat the cross-sectional estimates that make up the difference as statistically independent, in which case the variance of the difference is the sum of the variances of the annual estimates. For MCBS data, this assumption is not correct since a substantial proportion of respondents in any given year are carried over into the sample in the following year. Also, the sample design that is used to select each annual supplement retains the same primary sampling units (PSUs) and most of the secondary sampling units[1] (SSUs). Both of these factors need to be appropriately reflected in calculating the standard errors for net differences.

[1] For MCBS, secondary sampling units are ZIP Code area fragments or clusters of fragments within primary sampling units.

The purpose of this study is to review methods that can be used with MCBS data to compare estimates over time and to examine some approximation methods for estimating variances. Section 2 provides a brief review of the MCBS data files that are currently available and discusses procedures for estimating gross changes using "backward longitudinal" weights that are supplied with each file. Section 3 discusses procedures for estimating standard errors associated with net differences using direct methods and presents some results using variance approximations. Section 4 summarizes the results and suggests some topics for further study.

2. Estimates of Gross Changes for Continuously Eligible Beneficiaries

The Health Care Financing Administration publishes two types of data files that contain information about the Medicare population. Access to Care files provide information on access to and satisfaction with care as well as demographic data and information on health and functioning. Access to Care files have been released for each year from 1991 through 1995. Utilization and Expenditure files provide data on health care cost and utilization. Utilization and Expenditure files have been released for CY 1992 and CY 1993. Data on Utilization and Expenditure for 1994 and 1995 are currently in preparation and are expected to be available by the end of CY 1997. Table 1 identifies the populations that are represented by each sample.

Each data file includes "cross-sectional" sample weights that should be used to produce estimates of population characteristics. These weights will produce estimates that correspond to the population and time period that are represented by the sample. In addition, each file contains "replicate" weights that can be used to produce variances, standard errors and confidence intervals for the estimates. Using the replicate samples to compute estimates of variability results in estimates that correctly reflect the sample design for the MCBS. In general, each set of "cross-sectional" weights should be used only with the data from the same file.

2.1 MCBS "Backward Longitudinal" Weights

In most of the Access to Care data files, additional sets of weights have been provided that are designed for computing estimates of gross changes. Estimates using these weights are for beneficiary populations that were continuously eligible over a period from a previous fall through the current fall.

For example, these weights can be used to estimate transition probabilities from one state to another such as the proportion of beneficiaries in the fall of 1991 who were satisfied with their health care and were still satisfied in the fall of 1994. Such estimates can be computed from the combined observations from 1991 and 1994 for each record with the appropriate weights.

These weights can also be used to compute differences between aggregate estimates at each point in time, for example the proportion that are satisfied with their health care in fall, 1991 and the same proportion in fall, 1994. Separate sets of weights are provided for one-year, two-year and three-year longitudinal comparisons. Each set of weights is greater than zero only for cases in each sample that have been in the MCBS over the entire time period that is associated with the weights.

2.2 Estimates of Gross Change Using Combined Data Files

In order to make estimates of gross differences, a combined data set needs to be created that contains the relevant variables from each annual file along with the appropriate backward longitudinal weights. For longitudinal comparisons involving variables that are provided in the Access to Care data sets, data for the final year will be obtained from the same file as the weights. Data for earlier years will need to be obtained from previous files and merged into a common file.

For merging the data, records should be merged at the beneficiary level[2], and only those beneficiaries who completed all of the relevant interview rounds (that is, sampled records for which the corresponding set of weights is positive) should be included in the final file. Variables should be renamed prior to merging so that variables are properly identified as originating from a specific year or source data file.

3. Estimates of Net Change

Estimates of net change that reflect the changing composition of the Medicare population can be obtained simply by computing the difference between two "cross-sectional" estimates. Each "cross-sectional" estimate is computed in the usual way, using the appropriate "cross-sectional" weights that are included in the data file.

Computation of the associated standard error estimates is complicated by the fact that the two

[2] The variable BASEID is the record identifier for MCBS data files.

samples are not independent. If the two samples were independent, the variance of the difference could be estimated by the sum of the variances for each of the annual estimates. However, the two samples are not independent because many sampled cases are retained in the MCBS sample from year to year and because the sample design for selecting each new supplement uses the same PSUs and many of the same SSUs.

3.1 Estimates of Net Changes Using Direct Methods

Direct estimation of the variance of a differences, can be done using software appropriate for analysis of complex samples, such WesVarPC® (Brick, J. M., et al, 1997) or SUDAAN (Shah, B. V., et al, 1997). Some preliminary data management is required to create a data set for the analysis. Data from each annual file should be combined into a common data set by concatenating the records from each annual file. Weights in the combined file (including the replicate weights if used) should be renamed so that they all have the same name. A variable should also be created that indicates the file in which each record originated.

Figures 1 and 2 show the results of applying the direct method to calculate standard errors for changes in mean reimbursements. Mean reimbursements are estimated for various types of reimbursements (total annual Medicare reimbursements, Part A total, Part B total, inpatient services, and outpatient services) and for various levels of data aggregation (overall, by gender, by seven age categories, and by age category within gender). Changes from fall, 1992 to fall, 1993 are shown in Figure 1; and changes from fall, 1994 to fall, 1995 are shown in Figure 2. Values of the abscissa in each figure are standard errors for one-year differences calculated directly using Fay's modification of the balanced repeated replications (BRR) method in WesVarPC®. Values of the ordinate are differences between standard error estimates that would be appropriate for independent samples, based on the standard errors for each annual estimate, and standard errors computed using the direct method.

Differences shown in the figures tend to be positive, indicating that standard errors based on the direct method tend to be slightly lower, in general, than estimates based on independence. This is consistent with a small positive correlation between year-to-year responses for these data. Standard errors for 1994-1995 changes are closer to values based on independent samples than corresponding estimates for 1992-1993 changes, indicating that year-to-year sample correlations are smaller for the later years. This is consistent with a change in the MCBS sample design that was implemented in 1994, in which the design moved from a single longitudinal panel with small annual supplements to a design using a rotating panel system.

Figure 1. Standard Error Estimates for Change in Mean Reimbursements: 1992-1993

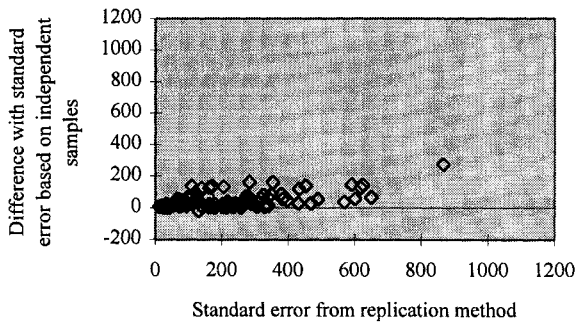
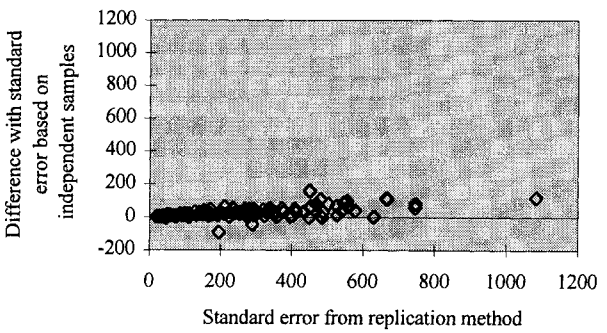


Figure 2. Standard Error Estimates for Change in Mean Reimbursements: 1994-1995

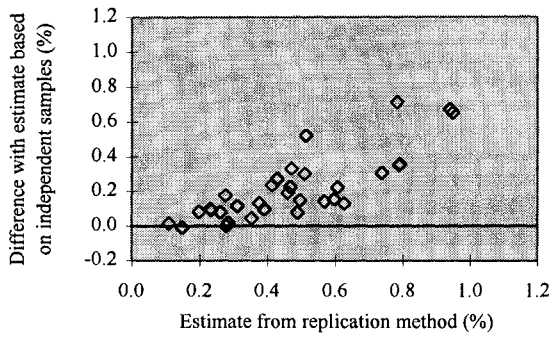


In addition, correlations for comparisons involving longer time periods can be expected to be smaller than corresponding correlations for changes across a single year. Results shown in Figures 1 and 2 are for changes over a one-year period and suggest that standards based on independent samples provides a conservative estimate of the true standard error. For estimates across longer periods of time, we would expect the sample observations to show less dependence and the resulting standard errors to be closer to the estimates based on this assumption.

Figure 3 shows similar results for estimates of differences in proportions between 1992 and 1993. Proportions are based on categorical responses to questions on satisfaction with care and health status aggregated over the entire population represented by each sample. Standard errors for these difference estimates suggest that the year-to-year correlations for

proportions are larger than those for mean reimbursements; however, this is at least partially attributable to the fact that the data in Figure 3 are summarized for the total sample rather than by age category. Year-to-year sample overlap is smaller within age categories because of progression of sampled persons to the next age category.

Figure 3. Standard Error Estimates for Differences in Proportions: 1992-1993



3.2 Estimates of Net Changes Using Variance Approximations

For estimates of net change with independent samples, the standard error depends only on the standard errors for each annual estimate. When samples are not independent, the standard error of the differences is also affected by the magnitude of year-to-year correlation and by the amount of sample overlap. Comparisons of standard errors computed using direct methods with estimates based on independent samples (Section 3.1) suggests that independent sample estimates provide a conservative approximation for standard errors of estimate of differences using MCBS data.

For samples selected using simple random sampling. The variance of a difference $d = y_1 - y_2$ is given by

$$V(d) = V^{(SRS)} - \frac{2 \bullet R_{12} S_1 S_2 P_1 P_2}{n_c} \qquad (1)$$

where $V^{(SRS)} = V(y_1) + V(y_2)$ is the corresponding standard error for independent samples, n_c is the sample overlap, R_{12} is the year-to-year correlation, S_1 and S_2 are the standard errors of the variables in each estimate, and P_1 and P_2 are the proportions of overlap in the two samples (Kish, 1965).

Although the MCBS sample is not a simple random sample of beneficiaries, we would like to approximate the variances using formulas similar to (1) with appropriate values of the correlations and proportions of overlap. Figure 4 shows results for standard error estimates using an approximation based on (1). The overall sample overlap between the 1992 and 1993 Access to Care samples is 87.7 percent and 87.8 percent, respectively, based on the unweighted samples. The approximations in Figure 4 are also based on year-to-year reimbursement correlations that range from 0.74 for outpatient reimbursements to 0.17 for inpatient reimbursements in the overlap sample.

There is an analog of (1) for proportion estimates; however, standard errors based on this approximation showed less improvement than those shown in Figure 4. Other methods that could be used to obtain variance approximations include Generalized Variance Functions (Wolter, 1985). Wolter suggests several models including:

$$\log(V^2) = \alpha - \beta \log(X) \qquad (2)$$

where $V^2 = \frac{Var(\hat{X})}{X^2}$ is the relative variance. Plots shown in Figure 5 provide some empirical support for applying this generalized variance model to estimate standard errors for year-to-year differences in proportions with MCBS data.

Figure 4. Standard Error Estimates for Change in Mean Reimbursements: 1992-1993

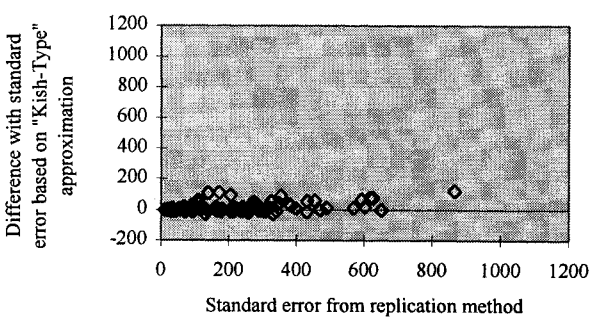
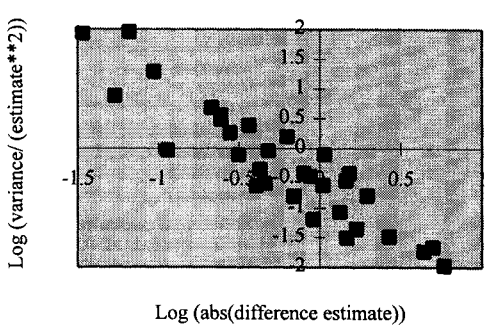


Figure 5. Generalized Variance Model for Standard Errors of Differences in Proportions: 1992-1993



4. Summary and Suggested Topics for Further Study

HCFA has prepared a number of data files containing annual data for the MCBS sample. Each of these files contains weights that can be used to produce "cross-sectional" estimates for the relevant population and time period. In addition, special sets of weights have been supplied with most of the Access to Care data files that will allow the user to make comparisons of gross changes over time for continuously eligible populations.

Because of the dynamic nature of the Medicare population, these estimates are appropriate only for subsets of the population that have continuous eligibility over the time period of interest. However, these approaches do not completely describe changes for the total Medicare population over time that are caused by changes in the population composition.

Estimates of net change, that reflect the changing composition of the Medicare population can be obtained by computing the difference between two "cross-sectional" estimates, using appropriate cross-sectional weights that are provided with each file. To properly interpret net differences, estimates of the corresponding standard errors are required that reflect not only the sample design, but also the correlation and overlap between samples. This study has reviewed some methods that can be used with MCBS data to compare estimates over time. In addition, we examined some results using approximations that reflect the observal year-to-year correlations and algebraic equivalencies that are based on simple random sampling methods. These results showed good agreement with approximations derived from direct methods.

5. References

Brick, J. M., Broene, P., James, P., and Severynse, J. (1997), *A User's Guide to WesVarPC*, Rockville, MD: Westat, Inc.

Bureau of the Census (1978). *The Current Population Survey, Design and Methodology*, Technical Paper 40: U.S. Department of Commerce.

Curtin, Lester and Feinlieb, Manning (1995). "Considerations in the Design of Longitudinal Surveys of Health".

Duncan, Greg J. and Kalton, Graham (1987). "Issues of Design and Analysis of Surveys Across Time", *International Statistical Review*, 56,1, 97-117.

Kalton, Graham and Citro, Constance F. (1995). "Panel Surveys: Adding the Fourth Dimension", *Innovation*, 8, 1.

Kish, Leslie (1965). *Survey Sampling:* Wiley.

Shah, B. V., Barnwell, B. G., Bieler, G. S., Boyle, K. E., Folsom, R. E., LaVange, L., Wheeless, S. C., and Williams, R. (1997), *Technical Manual: Statistical Methods and Algorithms Used in SUDAAN Release 7.0*, Research Triangle Park, NC: Research Triangle Institute.

Wolter, Kirk M. (1985). *Introduction to Variance Estimation:* Springer-Verlag.

Table 1. MCBS data files

Type of data	Year	Population represented	Estimates
Utilization and Expenditure[a]	1992	Beneficiaries with eligibility during CY 1992	CY 92
	1993	Beneficiaries with eligibility during CY 1993	CY 93
	1994	Beneficiaries with eligibility during CY 1994	CY 94
	1995	Beneficiaries with eligibility during CY 1995	CY 95
Access to Care[b]	1991	Beneficiaries eligible on or before 1/1/91 who remained alive and eligible through fall, 1991	Fall, 1991
	1992	Beneficiaries eligible prior to 9/1/92 who remained alive and eligible through fall, 1992	Fall, 1992
	1993	Beneficiaries eligible on or before 1/1/93 who remained alive and eligible through fall, 1993	Fall, 1993
	1994	Beneficiaries eligible on or before 1/1/94 who remained alive and eligible through fall, 1994	Fall, 1994
	1995	Beneficiaries eligible on or before 1/1/95 who remained alive and eligible through fall, 1995	Fall, 1995
	1996	Beneficiaries eligible on or before 1/1/96 who remained alive and eligible through fall, 1996	Fall, 1996

(a) Each Utilization and Expenditure data file contains a sample of beneficiaries that represents the "ever eligible" population for a calendar year. This population includes beneficiaries that die during the year, as well as beneficiaries that acquire eligibility during the year and "always enrolled" beneficiaries that are eligible throughout the year.

(b) Each Access to Care data file focuses on the population of Medicare beneficiaries who were enrolled in one or both parts of the program throughout the calendar year but also includes beneficiaries who died in the fall.