Uses of the Medicare Current Beneficiary Survey for Analysis across Time*

David Ferraro and Hongji Liu
Westat[1]

Keywords: Medicare, MCBS, Panel, Longitudinal, Cross-sectional, Repeated measurements

1. Introduction

The Medicare Current Beneficiary Survey (MCBS) is a continuous, multi-purpose longitudinal survey of Medicare beneficiaries conducted by the Centers for Medicare and Medicaid Services (CMS). The survey provides comprehensive data on access to and satisfaction with health care services, functional status, medical conditions, health care expenditures, health insurance, and other health-related topics. Since its inception in 1991, the MCBS has provided health policy researchers with a rich source of data on the health care utilization and costs for the Medicare population. While most of the analyses conducted with MCBS data have been cross-sectional, researchers have shown a strong interest in examining changes over time in health, insurance coverage, and other factors that affect health care service utilization and expenditures. Iezzoni, Davis, Soukup, and O'Day (2004), for example, used MCBS data to track indicators of functioning over time by comparing reported functional abilities between 1996 and 1997. Mello, Stearns, and Norton (2002) and Mello, Stearns, Norton and Ricketts (2003) conducted an analysis of beneficiaries entering the survey in 1992 and exiting in 1996 in which they modeled the likelihood of Medicare beneficiaries' enrollment in Medicare HMOs. Hoover, Crystal, Kumar, Sambamoorthi and Cantor (2002) also tracked sampled beneficiaries between 1992 and 1996 to estimate elderly beneficiaries' health care expenditures during their last year of life. Yang, Norton, and Stearns (2003) examined the relationship of longevity and health care expenditures using 1992-1998 MCBS data.

These are just a few of the many studies that have used MCBS data to examine changes over time. The majority of these studies have used the cross-sectional weights that are provided in MCBS data files, but these weights are designed specifically for cross-sectional analyses of the Medicare population, and thus may not always be appropriate for longitudinal analyses. An important feature that distinguishes longitudinal analyses from cross-sectional analyses is that the former are based on multiple measurements for the same individuals over time. Starting with the 1992 MCBS data release, separate longitudinal weights have been constructed and are available for analyses of repeated measurements on the same individuals. These weights have not been used to their full extent. The primary purpose of this paper is to describe the types and uses of the longitudinal weights available in MCBS data files and to give some examples of statistical techniques available for use in longitudinal analysis.

Section 2 gives some background about the MCBS and the data files that are available from the study. It also provides guidance on the use of the longitudinal weights currently available in the MCBS data files. Sections 3 and 4 present examples of the types of longitudinal analysis that can be done using these weights. Section 5 provides an example of the use of MCBS data to estimate spell duration. Section 6 discusses pooling data as a means to enhance longitudinal analysis. Finally, section 7 provides some concluding remarks.

2. MCBS Data Files and Longitudinal Weights

The MCBS collects in-person interview data from nationally representative samples of Medicare beneficiaries through the use of a rotating panel design. Under the rotating panel design, each annual MCBS sample (referred to as a "panel") consists of an age-stratified random sample of aged and disabled Medicare beneficiaries who were alive and eligible as of January 1 of the sampling year. Initial interviews for each new panel are conducted in the fall of the year in which the sample was selected. The sampled beneficiaries in each panel are interviewed at four-month intervals for up to four years (and a maximum of 12 data collection rounds). At the end of the four-year cycle, up to three calendar years of complete annual data on cost and utilization of health care will be available for the panel, at which time the panel is retired from the study.

Two sets of public use files** are released annually by CMS: the "Cost and Use" and the "Access to Care." The Cost and Use data files are intended primarily for estimating charges and payments for a complete calendar year. The Access to Care data files, on the other hand, focus on data that describe access to and satisfaction with health care services. Samples for both data files are comprised of beneficiaries from several different MCBS panels. The Cost and Use data files consist of three full panels (plus two partial panels used to represent

* The authors gratefully acknowledge Frank Eppig at CMS for his support, Adam Chu for his comments and technical assistance, Daniel Levine for his review of the paper, and Charles Halla for his programming support. The opinions expressed in this article are those of the authors, and do not represent CMS.

** Researchers are required to sign a data use agreement to access these files. Information about the availability of MCBS data files can be found on the CMS website: www. cms.hhs.gov/MCBS.

newly-enrolled beneficiaries), while the Access to Care files consist of four full panels. Additional details about the MCBS sample design and public use data files can be found in O'Connell, Chu, and Bailey (1997) and O'Connell, Lo, Ferraro, and Bailey (1998). Since the Cost and Use data files do not currently include longitudinal weights, the remainder of the discussion in this paper will be confined to the use of the Access to Care data files for longitudinal analyses.

The Access to Care data files include both cross-sectional and "backward" longitudinal weights. The available weights are designed to produce estimates of population characteristics that correspond to the population and time period represented by the sample. The population of inference for the Access to Care data files is often referred to as the "always enrolled" population because it is restricted to beneficiaries who were alive and continuously enrolled in Medicare throughout the entire year. Guidance for using the weights included in Access to Care data files is provided in the user's guides that accompany each data release. Although both cross-sectional and longitudinal weights are described in the user's guides, the use of the longitudinal weights is less straightforward since the weights apply to a data set that must be constructed from two or more annual releases. The longitudinal weights are characterized as "backward" because they apply to surviving respondents in a given Access to Care release and "look back" to data collected in previous releases.

Three sets of longitudinal weights are included in each file for analysis of MCBS Access to Care data: (a) three-year backward longitudinal weights designed for analysis going back three years in time and covering four consecutive years; (b) two-year backward longitudinal weights designed for analyses going back two years in time and covering three consecutive years; and (c) one-year backward longitudinal weights designed for analyses going back one year in time and covering two consecutive years. For example, the three-year backward longitudinal weights in the 2002 Access to Care file apply to fall round interviews conducted from 1999 to 2002. Although the examples in this paper involve the three-year longitudinal weights, they are also applicable to the other types of longitudinal weights.

It is important to note that the longitudinal weights in a particular Access to Care data release apply to specific subsets of cases within the data file and require data from at least one other Access to Care data release. This is illustrated in Table 1 for the two most recent Access to Care data releases. For example, it can be seen that in the 2003 data release, the three-year backward longitudinal weights (3 yr. BL) apply to a single panel, namely the 2000 panel. Moreover, to assemble the longitudinal data (i.e., repeated measurements) needed for analysis, data must be obtained from the previous 2000-2002 data releases in addition to the 2003 release. On the other hand, the one-year backward longitudinal weights in the same file apply to a much larger data set consisting of three panels, namely the 2000 through 2002 panels. In this case, data from the previous 2002 Access to Care data release would be needed to complement the 2003 Access to Care data. Finally, the bottom three rows of the table show the corresponding results for an Access to Care release from an arbitrary year, *T*. The appendix contains additional details about the creation of the analysis files used in the examples described in this paper.

3. Descriptive Statistics Using Longitudinal Weights

The longitudinal weights provided in the Access to Care data files can be used to estimate gross changes for a fixed population of survivors; i.e., beneficiaries who remained alive and eligible for Medicare one or more years after their initial interview. Such estimates apply to a particular cohort of survivors as it ages over the course of the study. Note that beneficiaries who have died after their initial interview do not have longitudinal weights in the Access to Care files.

Table 1. Longitudinal weights in Access to Care files by year of release

Release/weight		Panels to which weight applies			Access to Care data files required			
2002	1 yr. BL	1999	2000	2001	—	—	2001	2002
	2 yr. BL	1999	2000	—	—	2000	2001	2002
	3 yr. BL	1999	—	—	1999	2000	2001	2002
2003	1 yr. BL	2000	2001	2002	—	—	2002	2003
	2 yr. BL	2000	2001	—	—	2001	2002	2003
	3 yr. BL	2000	—	—	2000	2001	2002	2003
Year *T*	1 yr. BL	*T*-3	*T*-2	*T*-1	—	—	*T*-1	*T*
	2 yr. BL	*T*-3	*T*-2	—	—	*T*-2	*T*-1	*T*
	3 yr. BL	*T*-3	—	—	*T*-3	*T*-2	*T*-1	*T*

Table 2 shows estimates of the change in functional limitations from 1999 to 2002 based on weighted data for the 1999 panel in the 2002 Access to Care data file. The 1999 panel consists of 3,434 beneficiaries who completed four fall interviews.[***] The weights used are the three-year backward longitudinal weights provided in the 2002 data release. Data for these cases were obtained from the 1999 through 2002 Access to Care data files. Additional information about the creation of the files used in this example is given in the appendix.

The table indicates that 14.6 percent of the beneficiaries enrolled in Medicare in 1999 reported no functional limitations in 1999 but some limitations in 2002. Conversely, 9.3 percent of the beneficiaries reported some limitations in 1999 but none in 2002. Overall, the prevalence of functional limitations for this cohort of beneficiaries increased from 43.3 percent in 1999 to 48.6 percent in 2002. In this example, changes in conditions can occur in either direction. However, other conditions, such as strokes, can only have a change in one direction.

Table 2. Percent of beneficiaries in the 1999 panel that reported functional limitations in 1999 and 2002

	Limited in 2002	
Limited in 1999	No	Yes
No	41.7	14.6
Yes	9.3	34.0

The three-year longitudinal weights can also be used to track Medicare reimbursements over time for selected subpopulations. Table 3 presents the mean annual Medicare Part A and Part B reimbursements (in dollars) and associated standard errors (s.e.) by age group. This table summarizes the average total Medicare reimbursements for each of four age groups by year from which differences in mean reimbursements between age groups and differential growth rates across age groups can be calculated. The table indicates that while increases in Medicare reimbursements generally occurred across all age groups, the growth rates differ by age groups.

As another example, Figure 1 shows a graph of the prevalence of hypertension by age group and year. Close to half of the beneficiaries aged 65 to 74 reported that they had the condition in 1999, compared to 44 percent of those aged 85 and over. By 2002, the prevalence of hypertension grew to 64 percent of those aged 65 to 74 compared with 53 percent of those aged 85 years and over. Further analysis of these results is given in the following section.

4. Methods Using Linear Models

Though methods used in the examples in the previous section are useful for describing changes over time, other methods exist that are superior in taking into account the repeated measurements associated with sampled beneficiaries. One such method developed by Koch, Singer, and Stokes (1991) uses weighted least squares to estimate relationships between outcomes observed at different points in time and a set of explanatory variables. A second is based on a traditional regression framework, except that appropriate techniques are used to estimate the variance-covariance structure of the underlying longitudinal data. For a simple one-way analysis of variance models, the two approaches can be shown to yield identical results. For more complex models, the two methods are presumably asymptotically equivalent.

The method by Koch, et al. (1991) assumes the following basic setup. Let y_{ijk} denote the response for k-th subject in the i-th group $(i = 1, 2, \ldots, I)$ observed at time $j = 1, 2, \ldots, J$. Let $\mathbf{g}$ be the column vector of IJ group × time means, $\bar{y}_{ij}$. It is assumed that $\mathrm{E}(\mathbf{g}) = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is an IJ x p matrix that describes the underlying patterns of the $\bar{y}_{ij}$'s, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters to be estimated. Under this general framework, tests of hypotheses involving the model parameters are based on generalized Wald F-statistics that require computation of a variance- covariance matrix reflecting the underlying correlations between (repeated) measurements on the same individuals. Such variance-covariance matrices can be constructed using either replication methods or Taylor series linearization.

When the underlying model holds, there may be interest in testing hypotheses of the form: $\mathrm{H_{oc}}: \mathbf{C}\beta = 0$ where $\mathbf{C}$ is a known $c \times p$ matrix with full rank c. Under $\mathrm{H_{oc}}$, the distribution of the test statistic

$$\mathrm{Q_c} = (\mathbf{bC})'(\mathbf{CV_bC}')^{-1}\mathbf{Cb}$$

is approximately chi-square with c degrees of freedom, where $\mathbf{b} = \left(\mathbf{X}'\mathbf{V}_g^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}_g^{-1}\mathbf{g}$ is the weighted least squares estimate of $\boldsymbol{\beta}$, $\mathbf{V}_\mathrm{g}$ is the variance-covariance matrix of $\mathbf{g}$, and $\mathbf{V}_\mathrm{b}$ is the variance-covariance matrix of $\mathbf{b}$.

To illustrate the method, consider the hypertension example from Figure 1. For simplicity, we will analyze the results for the under-65 year-old age group. Let $\bar{y}_j, j = 1999, 2000, 2001, 2002,$ denote the estimated

[***] A small number of cases (23) do not have complete data for all four fall interviews. These "carryover" cases were nonrespondents in an intermediate fall round and are not included in the Access to Care file for that year. However, they subsequently completed the following interview and continued in the survey as respondents. These cases can be imputed by the user or deleted from the analysis.

Table 3. Part A and B reimbursements by age group for the 1999 panel

Age group (age as of 1999)	1999 Mean (s.e.)	2000 Mean (s.e.)	2001 Mean (s.e.)	2002 Mean (s.e.)
Under 65	$3,343 (438)	$4,066 (499)	$4,936 (593)	$5,776 (522)
65-74 Years	2,327 (202)	2,805 (254)	3,118 (295)	4,506 (457)
75-84 Years	3,453 (282)	4,092 (306)	4,798 (304)	5,450 (306)
85 Years and over	4,327 (513)	5,060 (549)	4,723 (512)	7,062 (762)

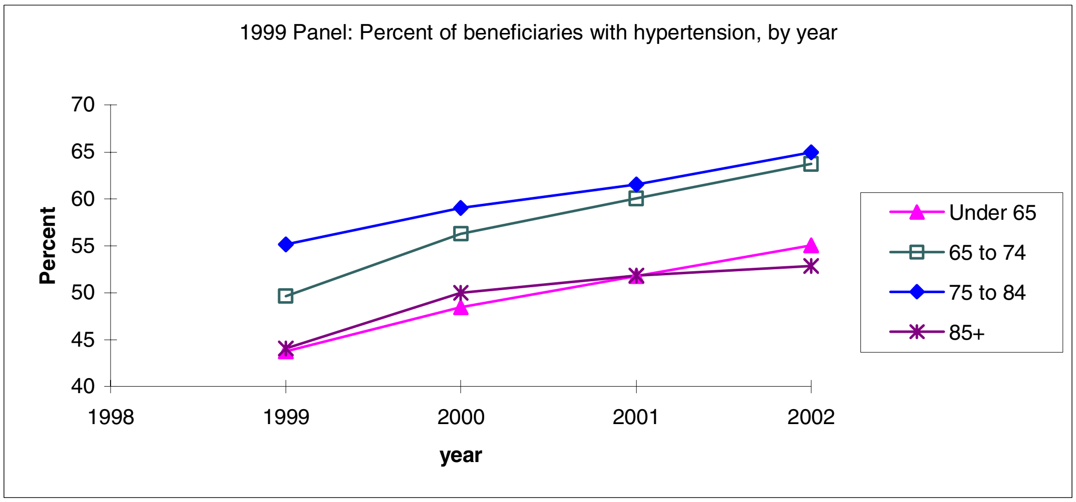

Figure 1. Hypertension by year and age group

hypertension rate in year j for the specified age group, and let $\mathbf{r}$ denote the corresponding vector of these four rates. Note that all of the estimated rates are computed using the three-year backward longitudinal weights. A simple model describing the observed hypertension rates is given by: $\mathrm{E}(r) = \boldsymbol{\beta}$. The hypothesis that the rates of hypertension are equal across the four years, i.e., $E(\bar{y}_1) = E(\bar{y}_2) = E(\bar{y}_3) = E(\bar{y}_4)$ can be specified as $\mathrm{H}_1 : \mathbf{C_1}\beta = 0$, where

$$\mathbf{C}_1 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{pmatrix}.$$

The corresponding adjusted F test is calculated as

$$F = \frac{e-c+1}{ec}(\mathbf{C}_1\mathbf{r})'\left(\mathbf{C}_1\mathbf{V}_b\mathbf{C}_1'\right)^{-1}(\mathbf{C}_1\mathbf{r}),$$

where $\mathbf{V}_\mathrm{b}$, is the estimated variance-covariance matrix of **b**, e denotes the number of degrees of freedom corresponding to $\mathbf{V}_\mathrm{b}(100)$, and c is the rank of $\mathbf{C}_1(3)$. Note that **b** is the weighted least squares estimate of β, which in this example is simply **r.**

The variance-covariance matrix $\mathbf{V}_\mathrm{b}$ was estimated using replication (e.g., see Judkins and Lo, 1993, for a description of the balanced repeated replication method used in the MCBS). The resulting adjusted F test is 13.60, corresponding to a p-value of <.0001. Thus, the hypothesis that the hypertension rates are equal is rejected.

This method can also be extended to more complex models such as a two-way model of hypertension rates by age group and year. Details on setting up such a model are given in Koch, et al. (1991), pages 221-222. Though the weighted least squares methodology is straightforward, implementation requires separate calculation of certain variance-covariance matrices and use of software to perform the required matrix operations.

An alternative is to use software packages such as WesVar (Westat, 2002) or SUDAAN (Research Triangle Institute, 2004) to run similar tests. The examples given below use WesVar, but SUDAAN or other packages can be used as well. However, using these software packages involves a change in the structure of the input data set. Rather than using a data file with one record per beneficiary and up to four repeated measurements per record, a concatenated data set must be created from the individual Access to Care data files. As described in the appendix, this process results in a data file in which each respondent appears more than once, and where the variable YEAR is used to identify the time period to which a particular response (e.g., hypertension) applies. The variable YEAR can then be used in a table statement or a regression model to examine differences over time. Depending on the constructed test, the use of these software packages yields results that are consistent with the weighted least squares method.

For instance, in the hypertension analysis described earlier for the under-65 year age group, the same test and results were achieved using the linear regression option in WesVar (e.g., see Westat, 2002, pages 6-12 to 6-14), where HYPERTENSION is specified as the dependent variable and YEAR is defined to be an independent categorical (class) variable. Although it is not obvious from the structure of the input data file, the replication-based variance-covariance matrices generated for the statistical tests properly reflect the correlations of the repeated measurements on the same individuals. In other words, even though the reported data for the same individual at different time points are entered as separate cases in the data file, the replicates created for variance estimation automatically take into account the year-to-year correlations (Westat, 2002, pages E1 to E5).

From Figure 1, it can be seen that the trend in hypertension rates appears to be roughly linear. Under the weighted least squares approach, this relationship can be described by the model: $\mathrm{E}(\bar{y}_j) = \alpha + \beta_j$, where j = year. To test whether the slope of the line is equal to zero, i.e., $\beta = 0$, the general approach in Koch, et al. (1991), can be used here as well. Alternatively, this test can be performed in WesVar by specifying YEAR as a continuous (rather than categorical) variable in the regression model. When this is applied to the data generating the rates in Figure 1, the resulting estimate of α is 0.41 and the estimate of β is 0.035. The adjusted F test associated with the test of the hypothesis that $\beta = 0$ is 37.86 with a p-value of <0.0001, leading to the conclusion that the slope is statistically significant.

More complicated models can also be tested. For example, a two-factor model relating hypertension to year and age group can similarly be specified with YEAR and AGE_GROUP defined as categorical variables, and the interaction term of YEAR*AGE_GROUP included in the model statement. The results of this analysis are shown in Table 4. Both YEAR and AGE_GROUP are highly significant, while the interaction is moderately significant at the 0.05 level.

Table 4. Results of a two-way hypertension model

Effect	Adjusted F test	p-value
Year	56.5111	<0.0001
Age group	9.481	<0.0001
Year by age group	2.414	0.0166

5. Spell Duration

Another useful technique for analyzing data from panel surveys involves the estimation of duration of “spells.” A “spell” in the context of the MCBS would be the length of time a beneficiary is in any type of program or type of health care, for example, home health use.

The Kaplan-Meier conditional probability technique (Miller, Lepkowski, and Kalton, 1992) is applicable to spells observed to start during the life of the panel. Generally speaking, spell duration estimation requires specific start and stop dates of the spells. Spells that start and stop in the panel period are known as uncensored spells. The Kaplan-Meier method only requires that spells be observed to start during the life of the panel and includes those that stop (uncensored) and those that do not stop (right censored) during that period. The method excludes spells that are in existence at the beginning of the panel and stop during the period (initially censored) and those that last the entire period (doubly censored).

Within these guidelines, a hazard function is estimated by calculating the proportion of spells that end at month t among all spells known to have lasted t or more months. The discrete time Kaplan-Meier estimate for the hazard function is given by

$$\hat{h}(t) = \frac{d_t}{\sum_{x=t}^{\infty} d_x + \sum_{x=t+1}^{\infty} c_x}$$

where d_t denotes the number of spells ending at time t and c_t denotes the number of spells censored at time t. The survival function S(t) is then the probability that a spell will last for more than t months and is estimated from h(t) as

$$\hat{S}(t) = \prod_{x=1}^{t} \left[1 - \hat{h}(x)\right]$$

Kaplan-Meier curves of the type described above can be estimated using the KAPMEIER procedure in SUDAAN.

6. Pooling Data

For longitudinal analyses of MCBS data it is tempting to pool data from several years to increase sample size and statistical power. An example of this possibility involving the three-year backward longitudinal weights can be seen in Table 1. As indicated in this table, independent three-year longitudinal samples are available from the 2002 and 2003 Access to Care data releases. Combining these samples would double the sample size. However, the three-year backward longitudinal weights in the 2002 data release apply to the 1999 panel while the three-year backward longitudinal weights in the 2003 data release apply to the 2000 panel. Thus, the populations represented by the two samples are different. The time periods covered by the two samples are also different. The sample from the 2002 release covers the period from 1999 to 2002, whereas the sample from the 2003 release covers the period from 2000 to 2003. In view of these considerations, there may be little to be gained by pooling the results if it is necessary to analyze the two samples separately. On the other hand, where there is no reason to believe that such differences will have an appreciable effect on inferences, pooling can obviously enhance analyses. An example where this might be the case are analyses involving rates of change over time. It is important to note that such assumptions can be tested using the methods described earlier.

Unlike the three-year longitudinal samples, the two- and one-year backward longitudinal samples from different years are not mutually exclusive since there is an overlap of panels between adjacent years. The same considerations mentioned previously also apply to the case of combining several two- or one-year longitudinal samples. That is, the samples being combined represent different populations and cover different time periods. However, there is an additional consideration involving these samples. For example, if the one-year longitudinal sample from 2002 (consisting of panels 1999, 2000, and 2001) is pooled with the one-year longitudinal sample from 2003 (consisting of panels 2000, 2001, and 2002), the 2000 and 2001 panels will be common to both samples. Except for losses due to attrition, the panels in common will contain essentially the same set beneficiaries. Thus, the apparent gain in sample size is not completely realized and the calculation of variances of estimates derived from the pooled sample must take this overlap into account. Because of the way the variance strata and replicates are constructed for the MCBS, the overlap will be taken into account in the variance estimation when the samples are concatenated for pooled analyses.

7. Conclusion

The MCBS is a survey that produces data for both cross-sectional and longitudinal analyses. For example, the MCBS Access to Care data releases are produced annually with both cross-sectional and longitudinal weights. Although the cross-sectional weights have been extensively used in analysis, the longitudinal weights, which are designed for estimation of gross changes over time, have not. This paper describes differences in the use of the longitudinal weights to analyze data over time. More specifically, the paper suggests that more attention be directed to longitudinal analysis of MCBS data and describes a number of statistical techniques which enhance the analysis across time using longitudinal weights.

References

Hoover, D.R., Crystal, S., Kumar, R., Sambamoorthi, U., and Cantor, J.C. (2002). Medical expenditures during the last year of life: Findings from the 1992-1996 Medicare Current Beneficiary Survey. *Health Services Research*, 37(6), 1625-42.

Iezzoni, L.I., Davis, R.B., Soukup, J., and O'Day, B. (2004). Physical and sensory functioning over time and satisfaction with care: The implications of getting better or getting worse. *Health Services Research*, 39(6) Pt. 1, 1635-52.

Judkins, D., and Lo, A. (1993). Components of variance and nonresponse adjustment for the Medicare Current Beneficiary Survey. Proceedings of the Survey Research Methods Section of the American Statistical Association, 820-825.

Koch, G., Singer, J., and Stokes, M. (1991). Some aspects of weighted least-squares analysis for longitudinal categorical data. *Statistical Models for Longitudinal Studies of Health*, 215-258.

Mello, M.M., Stearns, S.C., and Norton, E.C. (2002). Do Medicare HMOs still reduce health services use after controlling for selection bias? *Health Economics*, 11, 323-340.

Mello, M.M., Stearns, S.C., Norton, E.C., and Ricketts III, T. (2003). Understanding biased selection in Medicare HMOs. *Health Services Research*, 38(3), 961-992.

Miller, D.P., Lepkowski, J., and Kalton, G. (1992). Estimating duration of food stamp spells from the SIPP. *Proceedings of the Survey Research Methods Section of the American Statistical Association,* 560-565.

O'Connell, J., Chu, A., and Bailey, R. (1997). Considerations for analysis of the Medicare Current Beneficiary Survey (MCBS) across time. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 645-650.

O'Connell, J., Lo, A., Ferraro, D., and Bailey, R. (1998). Sampling and estimation in the Medicare Current Beneficiary Survey. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 156-161.

Shea, D., Stuart, B., Vasey, J, and Soma, N. (1999). Medicare physician referral patterns. *Health Services Research*, 34(1) Pt. 2, 331-48.

Sudaan (2004). *Sudaan Language Manual*, *Release 9.0*. Research Triangle Park, NC: Research Triangle Institute.

Westat (2002). *WesVar 4.2 User's Guide*. Rockville, MD: Westat.

Yang, Z., Norton, E.C., and Stearns, S.C. (2003) Longevity and health care expenditures: The real reasons older people spend more. *Journal of Gerontology*, 58B(1), S2-S10.

Appendix

As shown in Table A1, three sets of longitudinal weight files are provided in each Access to Care data release. The longitudinal weights are contained in separate files referred to as "RICs." For example, the file named RIC X4 contains the three-year backward longitudinal weights, L4YRSWGT and corresponding replicate weights, L4YRS0001-L4YRS100. For longitudinal analysis, data from two or more Access to Care data files are required. In general, variables of interest that change over time should be obtained from the relevant Access to Care files and saved under different names. For example these variables would include: age; income; marital status; health status; functioning; and chronic conditions. On the other hand, for certain "constant" demographic and other classification variables, such as race/ethnicity, gender, and education level, only the variables from the oldest data file should be retained.

Two analysis files were created for the examples given in this paper. The first was a person-level multiple year data file with data from the 1999-2002 Access to Care data releases recorded on a single record under different names. For example, the year-specific variables on hypertension were renamed in the longitudinal data file to identify the data year (i.e., HYPERTENSION99, HYPERTENSION00, etc.). The required three-year backward longitudinal weights were then obtained from the RIC X4 file in the 2002 Access to Care data release. This file was used in the examples of descriptive statistics and models employing the weighted least squares methodology.

The second file used in the examples was a concatenated file with one record per beneficiary per year. The required weights were merged onto the concatenated data file from the RIC X4 in the 2002 Access to Care data release. In this case, the variables are not renamed. To distinguish the values from different years, a year variable was added to the concatenated file to identify the data year (e.g., YEAR = 1999 for the 1999 Access to Care file data, 2000 for the 2000 Access to Care data file, and so on). This file was used in the examples involving traditional regression models.

Table A1. Longitudinal weights

Longitudinal sample	Longitudinal and replicate weights	Data file
3-year	L4YRSWGT, L4YRS001-L4YRS100	RIC X4
2-year	L3YRSWGT, L3YRS001-L3YRS100	RIC X3
1-year	L2YRSWGT, L2YRS001-L2YRS100	RIC X2