COMPONENTS OF VARIANCE AND NONRESPONSE ADJUSTMENT FOR THE MEDICARE CURRENT BENEFICIARY SURVEY

David Judkins and Annie Lo, Westat, Inc.
Westat Inc., 1650 Research Boulevard, Rockville, Maryland 20850

Keywords: Response Propensity, Fay's Method

1. Introduction

The Medicare Current Beneficiary Survey (MCBS) is a continuous, multi-purpose panel survey of Medicare beneficiaries sponsored by the Health Care Financing Administration (HCFA). The target population of the study is the aged and the disabled residing in households and nursing homes in the 50 States, District of Columbia, and Puerto Rico. A panel of beneficiaries is interviewed three times a year. MCBS operates in rounds with the first round of data collection conducted in the fall of 1991. A sample of 15,411 individuals was drawn. This sample size was chosen to yield complete annual data on 12,000 beneficiaries. Access to health care, health status and functioning, usual source of care, satisfaction with health care, health insurance, as well as demographic characteristics were collected in round 1. For round 2 through round 10 in calendar years 1992, 1993 and 1994, the emphasis is on information on cost, utilization, and expenditures for health care.

The multistage sample design, the coverage issues, the sampling operations, and round 1 response rates for this national in-person Computer Assisted Personal Interviewing (CAPI) survey are reported in Apodaca, Judkins, Lo and Skellan (1992). This paper presents the components of variance and the nonresponse adjustment of the sample weights over the first three rounds of the survey

2. Nonresponse Adjustment

The major causes of nonresponse for MCBS are refusals and unlocatable sample persons. Like many complex surveys, MCBS uses survey weights to account for differential probabilities of selection and to adjust for nonresponse of beneficiaries. The weights were created in several steps. First, a baseweight was computed by taking the reciprocal of the probability of selection for the beneficiary. The second step involved raking the baseweights to reduce both the undercoverage bias and the variance due to inaccurate measures of size at the PSU and the ZIP code levels. The third step was to create the round 1 final weights by adjusting the raked weights for nonresponse. Round 2 weights were computed by adjusting the round 1 final weights for nonresponse at round 2. Finally, we computed round 3 weights by adjusting round 2 weights for nonresponse at round 3.

Nonresponse adjustment cells were formed on the basis of modeled response propensity to minimize the potential for bias by maximizing the variation in response rates across cells. The response propensity approach assumes that the characteristics of interest are unrelated to response status within an adjustment classs. Logistic regression was used to predict response propensity. The eligible sample was then stratified by the response propensity to form adjustment cells. Within each cell, a weighted nonresponse adjustment factor was computed and the raked weight for each beneficiary was multiplied by the adjustment factor. Raking of the baseweights is presented in Section 2.1. The patterns of response propensity for round 1 through round 3 interviews are presented in Sections 2.2 through 2.4.

2.1 Raking of Baseweights

The baseweights of MCBS beneficiaries were raked to the March 1991 5-percent Health Insurance Skeleton Write-Off (HISKEW) file maintained by HCFA containing persons eligible as of January 1, 1991. The variables used in the row adjustment were age domain (0-44, 45-64, 65-69, 70-74, 75-79, 80-84, and 85+ as of July 1, 1992), gender, and region (Northeast, South, Midwest, West, and Puerto Rico). The column adjustment was by age domain, gender, and race (black, other). This adjustment was equivalent to raking region against race with each age-gender cell. The population total was adjusted from 33,407,262 to 34,205,380 beneficiaries who were eligible for Medicare as of January 1, 1991. The weighted mean of the adjustment factor was 1.02, almost exactly the size of the undercoverage. This result indicates that the variance due to the multistage design is essentially in the age-gender-race-region distribution of the sample, not in the total sample size.

2.2 Round 1 Response Propensity

Demographic, geographic, socioeconomic variables, and medical charges were used as potential predictors of round 1 response rates. Table 1 shows the parameter estimates, standard errors, and the chi-square statistics for the independent variables. The continuous variables PCTPOOR, VISRATIO, REIMBAMT, and MED_INC were shifted to have zero means prior to modeling in order to reduce rounding errors.

Results from logistic regression indicate that among the elderly (65 and older), males were more likely to be available and cooperative; while among the disabled, females were more likely to respond. The response rates were lower in the Northeast and Midwest. In metropolitan areas with population less than 1 million people, the response rates were higher than in large metropolitan areas with population of more than 1 million people. People in ZIP codes with lower 1990 median income were more likely to

Table 1. Analysis of maximum-likelihood estimates for Round 1 response propensity

Variable	Description	Estimate	SE	Chi-Square	p-value
Intercept		2.1088	0.0575	1346.93	0.0000
Age	< 65	0.1130	0.0390	8.41	0.0037
MSA size	New York, Los Angeles	-0.1591	0.0690	5.32	0.0211
	population 1,000,000+	-0.1292	0.0487	7.05	0.0079
	population 200,000-1,000,000	0.0337	0.0512	0.43	0.5102
	population < 200,000	0.1357	0.0771	3.10	0.0784
Region	Northeast, Midwest	-0.2605	0.0412	39.99	0.0000
	South, Puerto Rico	0.0401	0.0414	0.94	0.3322
Gender	Male	-0.0694	0.0388	3.19	0.0741
Age*Gender	Male, < 65	-0.1399	0.0388	12.99	0.0003
PCTPOOR	Percent of population in a ZIP code whose income was below the poverty level according to the 1980 Census	-0.0423	0.8574	0.00	0.9606
PCP_2	PCTPOOR (squared)	15.5505	5.0512	9.48	0.0021
PCP_3	PCTPOOR (cubed)	-125.6	39.7198	9.99	0.0016
PCP_4	PCTPOOR (quadratic)	191.9	89.1192	4.64	0.0313
VISRATIO	Ratio of 1996/1991 visits fee (determined by Physician Payment Review Commission) for area (usually state)	2.6292	0.3851	46.61	0.0000
VIS_2	VISRATIO (squared)	-1.5338	0.7338	4.37	0.0366
VIS_3	VISRATIO (cubed)	-13.4731	3.5643	14.29	0.0002
REIMBAMT	Reimbursement amount for the individual from HCFA payment records	0.000061	0.000017	12.89	0.0003
MED_INC	1990 median income for ZIP code	-0.00002	0.000006	11.86	0.0006

respond. Higher reimbursed amount was also associated with higher response rates. Although the linear effect of PCTPOOR was not significant, the squared term had a significant negative impact on response propensity. However, the cubic and quadratic terms of PCTPOOR were positively related to response propensity. Finally, response propensity first increased with VISRATIO, then decreased.

The 14,530 eligibles were grouped into 145 nonresponse adjustment cells based on their response propensity. Each cell contained approximately 100 beneficiaries. The response rates ranged from 72% to 99%, with an overall round 1 response rate of 87.3%. Within each cell, the weighted round 1 response rate was calculated. The inverse of the response rate was assigned to each member of the cell that responded at round 1 as their nonresponse adjustment factor. The adjustment factors ranged from 1.02 to 1.45.

2.3 Round 2 Response Propensity

Round 1 respondents eligible for round 2 were included in the round 2 nonresponse adjustment. These round 2 eligibles were first grouped into two groups. The first group contained beneficiaries in long-term care facilities as of round 2. The second group consisted of beneficiaries in the community as of round 2. From the second group, we created separate cells for: (a) the recently deceased; (b) those who were unable to respond for themselves due to illness, frailty or mental incapacity; and (c) beneficiaries who had unusual patterns of item nonresponse at round 1. These items included Medicaid participation, interview conducted by proxy or sample person, income reported, limited social life in past month, lifting difficulty, reaching difficulty, delayed care because of health cost, service in the Armed Forces, race, region, metropolitan statistical area (MSA) size, and length of interview. These items were used as independent variables to model round 2 response propensity for the balance of beneficiaries in the community. The nonresponse rates of these items were quite low. Since no imputation has been performed on these items, a special cell has to be created for the 675 people who did not respond to these items. The response rates for the selected groups are shown in Table 2.

Table 2. Round 2 response rates for selected groups

Group	Number	Number of responses	Response rate
Recently deceased	130	117	90.0
In community at round 2 but too sick to respond for self	1210	1164	96.2
Other eligibles in community at round 2 with unusual round 1 item nonresponse patterns	675	603	89.3
Eligibles in facilities as of round 2	911	895	98.2

To create a reasonable set of nonresponse adjustment cells for round 2, we developed separate models for the facility component and for the balance of the community component. On the facility side, a significant predictor for round 2 response propensity was whether or not the sample person was covered by Medicaid since admission. People with Medicaid coverage were more likely to respond. Eligibles in facilities were grouped into 30 cells by their response propensity. Each cell contained about 30 persons. Across the cells, the response rates ranged from 90% to 100%, with an overall round 2 response rate of 98.2%.

On the balance of the community data set, the following characteristics were associated with persons being **more** likely to respond to round 2: Medicaid participation, reported income at round 1, having delayed health care because of cost, service in the Armed Forces, and people in the 65-69 age group. The following characteristics were associated with persons being **less** likely to respond to round 2: interview conducted by proxy, limits on social activities due to health, much difficulties in lifting 10 pounds, much difficulties in reaching over head, race of white or other nonblack, and residence in the Northeast, in Los Angeles or Chicago, or in other major metropolitan areas other than New York City.

The length of the round 1 interview had complex effects on the round 2 response rate. We found a cubic parabolic effect, where response propensity first increased with the length of the round 1 interview, then decreased, and finally, increased again. Lastly, a large nonresponse adjustment factor in round 1 was associated with a lower response rate in round 2, indicating that whatever underlying factors lead to nonresponse at the initial round, continue to have residual effects at round 2 among the round 1 respondents. Table 3 shows the parameter estimates, standard errors, and the chi-square statistics for the independent variables used to model round 2 response propensity. The continuous variable, length of interview in minutes, was shifted to have zero means prior to modeling.

The 9,630 eligibles in the community were grouped into 96 cells, with each cell containing about 100 persons. The response rates in the cells ranged from 64% to 100%. An overall round 2 response rate of 92.9% was achieved. The round 2 nonresponse adjustment factors ranged from 1 to 1.5.

2.4 Round 3 Response Propensity

Beneficiaries who responded to round 2 and were eligible for round 3 were first categorized into two groups based on their residence. The first group contained people who spent all their time in long-term care facilities at rounds 1 and 2. People who spent at least some time outside long-term care facilities during the reference periods for rounds 1 and 2 fell into the second group. The second group was further divided into: a) people who died between the round 2 and round 3 interviews; b) people who were unable to respond for themselves and required a proxy at round 3; c) people who did not respond to a number of selected questions, such as the total amount of payments for medical care from sources other than Medicare in the round 2 reference period; and d) others. Groups a through c and the facilitiy group were further subdivided by whether or not they were Medicaid recipients at round 2. Each subdivision formed a separate nonresponse adjustment cell. The response rate for selected groups are shown in Table 4.

Response propensity at round 3 was modeled for the "others" group. The following predictors were significant in predicting response propensity at 0.05 level. We found that the higher the total payment by sources other than Medicare, the lower the response propensity. A large nonresponse factor in round 1 was also associated with a lower response rate in round 3. People in Los Angeles were less likely to respond.. Finally, unreported total income at round 1 also resulted in a low response propensity. Higher response rates were related to the following attributes: sample persons who were unable to lift 10 pounds; or month all the time; or the total income of the sample sample persons had limited social life in the past person and spouse was less than $50,000 at round 1. It is interesting that people with difficulty in lifting had higher response rates at round 3, whereas difficulty in lifting was inversely related to response propensity at round 2.

A total of 80 cells were formed for the 8,027 round 3 "others" eligibles. The response rates ranged from 81% to 100%, with an overall round 3 response rate of 94.9%. The round 3 nonresponse adjustment factors ranged from 1 to 1.2.

Table 3. Analysis of maximum-likelihood estimates for round 2 response propensity

Description	Estimate	SE	Chi-Sq	p-value
Intercept	2.1359	0.2575	68.7810	0.0001
Covered by Medicaid	0.5076	0.1558	10.6125	0.0011
Interviewed proxy	-0.9395	0.1581	35.3034	0.0001
Income reported	0.8287	0.1045	62.9432	0.0001
Health limited social life in last month	-0.2229	0.0968	5.2995	0.0213
Difficulty lifting 10lbs	-0.2988	0.1186	6.3462	0.0118
Difficulty reaching over head	-0.3307	0.1318	6.2899	0.0121
Delayed health care because of cost	0.4261	0.1401	9.2489	0.0024
Ever in Armed Forces	0.3148	0.1100	8.1940	0.0042
White	-0.4125	0.1433	8.2881	0.0040
Not white or black	-0.5860	0.2586	5.1328	0.0235
Age 65-69	0.2497	0.1141	4.7847	0.0287
NE Census Region	-0.3030	0.0999	9.1987	0.0024
Los Angeles and Chicago	-0.7567	0.1830	17.0875	0.0001
MSAs 200,000+ less NY, LA and Chicago	-0.2937	0.0919	10.2057	0.0014
Length of interview	0.00962	0.00299	10.3662	0.0013
Lenght of interview (squared)	0.00039	0.000101	15.0147	0.0001
Length of interview (cubed)	0.000004	0.000002	6.1675	0.0130
Logistic transform of the inverse of round 1 nonresponse adjustment factor	0.2822	0.0790	12.7544	0.0004

Table 4. Round 3 response rates for selected groups

Group	Number	Response rate	Response rate (Percent)
Recently deceased	142	135	95.1
Medicaid	25	24	96.0
No Medicaid	117	111	94.9
Ever in community at rounds 1 and 2 but too sick to respond for self	1180	1165	98.7
Medicaid	360	385	99.4
No Medicaid	820	807	98.4
Other eligibles in community at round 2 with rounds 1 and 2 selected item nonresponse	1366	1294	94.7
Medicaid	233	223	95.7
No Medicaid	1133	1071	94.9
Eligibles in facilities in rounds 1 and 2	836	829	99.2
Medicaid	571	569	99.6
No Medicaid	265	260	98.1

Note: Medicaid status refers to round 2.
Vital status and proxy status refer to round 3.

3. Variance Estimation

A form of the balanced repeated replication (BRR) technique, Fay's Method, was used to compute the sampling errors for estimates from the MCBS. Fay's estimate of variance is given by

$$\frac{1}{(1-K)^2}\left[\frac{1}{T}\sum_{r=1}^{T}(\hat{x}_r-\hat{x})^2\right],$$

where T is the total number of replicates employed, "r" of x_r designates that the estimate x_r is based on the r-th replicate, and x is the estimate from the full sample, 100(1-K)% is referred to as the Fay's perturbation factor. Judkins (1990) evaluated several perturbation factors for ratios, regression coefficients, and medians in a Monte Carlo simulation study. His results showed that a perturbation factor in the range of 50-70% performed relatively well in terms of bias and stability of the variance estimates when compared with the standard BRR and the jackknife methods. Smaller values of K were found to be better for medians. Since a substantial number of medians will be estimated for the MCBS, we used a Fay's perturbation factor of 70% (i.e., K=0.3).

A total of 100 strata were formed for variance estimation purposes. Thirty-seven of these variance strata were created from the first-stage noncertainty strata. The noncertainty primary sampling units (PSUs, composed of MSAs and clusters of non-metropolitan counties) were originally selected in pairs for MCBS with two from each stratum. The first PSU in the stratum formed the first variance unit, the second PSU formed the second variance unit. The remaining 63 variance strata were formed by combining secondary sampling units (ZIP codes) in certainty PSUs. Each resulting variance stratum either contained 2 or 3 variance units.

The baseweight was adjusted by a perturbation factor to form the replicate weight. For MCBS, 100 replicate weights were formed. The values of the perturbation factor depended on the composition of the variance strata, that is, whether the first and second half-samples within the variance stratum consisted of one or two variance units. Raking was repeated for each of the 100 replicates. Nonresponse adjustments were recomputed for each of the 100 replicates using the perturbed baseweights and the original nonresponse adjustment cells.

The variance estimates calculated using Fay's method account for clustering, stratification, unequal probabilities of selection, and ratio adjustments. Estimates and estimated variances have been computed for seven selected items: poor health status, hypertension, difficulty with bathing, Medicaid participation, high school graduate, Hispanic origin, and income below $25,000 per annum. These items were cross-tabulated by region, MSA size, and by age domain and gender.

Variances from the MCBS design can be decomposed into two major components: between-PSU and within-PSU. Between-PSU variance is the extra component of variance that results from restricting the sample into 107 PSUs. The PSUs were formed by expanding the 1981 Westat general purpose sample of 100 PSUs. We estimated within-PSU variance by re-assigning variance strata and units and then repeating the weighting procedures. For each subdomain estimate of an item, a direct variance estimate was computed for the total and within PSU variances. To estimate between-PSU variances, we subtracted estimated within-PSU variances from total variances. The existence of some large design effects mainly arises from between-PSU variance. The additional clustering by ZIP code within PSUs does not appear to have had a major effect on variances. The importance of between-PSU variance varies widely across the statistics we examined. Relative variance estimates for the prevalence of selected variables are shown in Table 5.

Table 5. Relative variance estimates for selected variables

Prevalence	%	Total	Within-PSU	Between-PSU	Design effects
Fair or poor health status	30.9	.00040	.00022	.00018	2.26
Hypertension	45.0	.00012	.00012	.00000	1.24
Difficulty w/bathing	18.5	.00047	.00031	.00015	1.35
Medicaid participation	12.3	.00099	.00051	.00048	1.75
High school graduate	51.4	.00022	.00009	.00013	2.84
Hispanic origin	4.6	.01588	.00238	.01349	9.62
Income <$25K per annum	63.7	.00014	.00006	.00008	2.74

For some statistics, such as the prevalence of the need for assistance in bathing for specific age-by-gender subdomains, between-PSU variances were trivial and the corresponding design effects were small. Between-PSU variances and total design effects were larger for regional estimates than for national estimates, but nowhere near as large as those for estimates by metropolitan status. For example, for the prevalence of non-metropolitan beneficiaries with income below $25,000 per annum, the between-PSU variance accounts for 94% of the design effect of 20+. Between-PSU variance and total design effects were also quite large for Hispanic estimates.

Although Fay's method makes the estimation of the sampling variance of any statistic straightforward, the estimation process is computationally intensive and costly for multivariate surveys like the MCBS, in which a number of comparisons among the resulting parameter estimates are of interest. To reduce the work required to calculate sampling errors for each estimate, generalized variances are used as an alternative approach.

4. Generalized Variance Functions

Direct variance estimates are themselves subject to sampling errors. As reported in Apodaca et. al. (1992), the design effect for 70-74 year olds with income below the median was 1.03, while the design effects for neighboring age brackets (65-69 and 75-79) were 1.72 and 1.75. Some kind of smoothing is thus required before analyzing the design effects. One of the smoothing techniques, generalized variance functions (GVFs) are used for MCBS. With GVFs, variances are simultaneously estimated for groups of statistics, resulting in a possibly more stable set of estimates and still accounting for the effects of a complex sample design.

GVFs relate the relative variance of a survey estimator to the expectation of the estimate. We adopt the following model for MCBS:

$$V^2 = \frac{\sigma_{\hat{X}}^2}{X^2} = \alpha + \frac{\beta}{X}.$$

where V^2 represent the relative variance of an estimator, $\hat{X}$ of some population total X.

The model was fitted to three subgroups: (1) age group by gender; (2) region; and (3) region by metropolitan size. To compute the coefficients α and β for each fitted model, an iterative procedure using weighted least square was used. GVFs results are presented in Table 6.

The higher values of R^2 for the models of within-PSU variances indicate that the direct estimates of within-PSU variances were much more stable than those of total variances. The design effects resulting from within-PSU clustering were smaller than or equal to those that would have otherwise obtained from simple random sampling. This suggests that the stratification and post-stratification are highly effective in reducing within-PSU variances. In fact, they seem to have largely counteracted the effects of differential sampling and ZIP-level clustering.

Table 6. Generalized variance modeling results

	Age by gender	Region	Region by metro. size
Total variance:			
R^2	0.44	0.24	0.33
Design effect	1.0	2.5	10.6
α parameter	-0.000029	-0.000194	-0.00115
β parameter	2491	6359	27194
Within PSU variance:			
R^2	0.77	0.81	0.75
Design effect	0.7	0.9	1.0
α parameter	-0.00005	-0.000047	-0.000061
β parameter	1794	2306	2606
Between-PSU Variance:			
Percent	28	64	90

However, the picture is not so rosy for total variance involving a domain such as metropolitan areas that is not perfectly reflected in the stratification and not involved at all in the post-stratification. Between-PSU appears to be the major problem. The between-PSU variance by region-by-metropolitan size accounts for 90% of the design effect of 10.6. Since metropolitan status was not used as one of the raking factors, it is not surprising that its between-PSU variance was larger than regional and demographic estimates. The between-PSU variance accounts for 28% of the total design effect in the age-by-gender subgroup, which is within our expectation given the number of PSUs.

5. Improving Metropolitan and Nonmetropolitan Estimates

The high between-PSU variances for metropolitan and nonmetropolitan estimates are likely caused by changes in the definitions of MSAs (between 1980 and 1990). Subsequently, HCFA has revised the metro status variable by geocoding each beneficiary in the HISKEW and the round 1 MCBS sample using information from administrative records. The newly geocoded HISKEW provides a very powerful tool to create a better set of round 1 weights.

As reported in Section 2.1, we used two dimensions in the raking. One dimension was age by gender by region. The other dimension was age by gender by race. To improve metro/nonmetro estimates, we have defined four dimensions for the raking. Dimension 1 was by age by gender by race. Dimension 2 was by region by metro. Dimension 3 was by region by age. Dimension 4 was by metro by age. Raking will ensure good comparability with the HISKEW for each of the named dimension. Comparability will not be as good for unnamed dimensions such as race by metro.

The round 1 final weights and replicate weights have subsequently been revised as a result of using the new dimensions in raking. Table 7 presents the total relative variance estimates for percentages of various characteristics in the non-metropolitan area, computed using the original and the revised round 1 final weights. The revised weights are labeled as New Weights.

Table 7. Total relative varance estimates for selected variables in non-metropolitan area

Prevalence	Percentage		Total relative variance		
	Original weights	New weights	Original weights	New weights	Percent change
Fair or poor health status	36.4%	36.1	.00180	.00148	-17.78
Hypertension	48.7%	48.4	.00063	.00067	6.35
Difficulty with bathing	21.7%	21.6	.00130	.000106	-18.46
Medicaid participation	14.3%	14.4	.00518	.00471	-9.07
High school graduate	47.7%	48.4	.00200	.00174	-13.00
Hispanic origin	3.1%	3.1	.39927	.38352	-4.19
Income <$25K per annum	81.6%	81.8	.00053	.00048	-9.43

With the exception of hypertension, the total relative variances for the prevalence of all other characteristics shown in Table 7 have been reduced by 4% to 18%. Incorporating the revised metro status in raking markedly improves the precision of metropolitan and non-metropolitan estimates.

REFERENCES

Apodaca, R., Judkins, D., Lo, A., Skellan, K. (1992). Sampling From HCFA Lists. ASA Proceedings of the Section on Survey Research Methods.

Hosmer, D.W. and Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley & Sons.

Judkins, D. (1990). Fay's Method for Variance Estimation. Journal of Official Statistics, 3, 223-240.

Valliant, R. (1987). Generalized Variance Functions in Stratified Two-Stage Sampling. Journal of American Statistical Association, 82, 499-508.

Wolter, K.M. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.