Estimation of Non-Response Bias in the Medicare FFS HOS

Nancy McCall, Sc.D., Galina Khatutsky, M.S., Kevin Smith, M.A., and Gregory C. Pope, M.S.

We examined non-response bias in physical component summary scores (PCS) and mental component summary scores (MCS) in the Medicare fee-for-service (FFS) Health Outcomes Survey (HOS) using two alternative methods, response propensity weighting and imputation for non-respondents. The two approaches gave nearly identical estimates of non-response bias. PCS scores were 0.74 points lower and MCS scores 0.51 points lower after adjustment for non-response through imputation and 0.63 and 0.46 lower after adjustment for propensity weighting. These levels are small for component scores suggesting that survey non-response to the FFS HOS does not adversely affect estimates of average health status for this population.

INTRODUCTION

Health surveys are designed to provide information about the health status of some population of interest. In surveys of older adults, a major threat to the validity of survey estimates is non-response. When sampled individuals do not respond to the survey at all or fail to answer key items, non-response occurs. This non-response may lead to bias in survey estimates. Non-response bias is the systematic difference between the outcome scores for survey respondents and the (unknown) scores that would have been obtained if all subjects had completed the survey. The degree of bias is determined by two factors: the difference in scores between respondents and non-respondents and the non-response rate.

In recent years, CMS has begun to collect health status information from Medicare+Choice (M+C) beneficiaries through the conduct of the HOS. Thus, the validity of self-reported health status estimates is important for decisions about using the HOS to report changes in average health status. The presence of non-response bias of any significant degree could be problematic in the use of health status at the M+C plan level as a performance measure. This article reports the results of a non-response bias analysis conducted using the 1998 FFS HOS, which allowed for the inclusion of claims-based measures of health status for both respondents and non-respondents not available at the time of study for M+C enrollees.

Several studies have previously examined various aspects of differences between respondents and non-respondents to mailed health surveys (Fowler et al., 2002; Fowles et al., 1994; Grotzinger, Stuart, and Ahern, 1994; Lasek et al., 1997; Rowland and Forthofer, 1993). In general, previous research has found that non-elderly non-respondents are healthier than respondents are, whereas elderly non-respondents are sicker. For example, Fowles and colleagues (1994), using administrative, claims, and survey data from a mailed survey of enrollees in a large

MCP, found that the pattern of non-response was not consistent across age groups. Non-respondents under age 65 were younger, healthier, and had significantly lower health costs than respondents, in contrast to senior non-respondents, who were sicker and had higher medical costs than respondents of the same age group. In both age groups, those with mental health problems were less likely to respond (Fowles et al., 1994). Hornbrook and Goodman (1995) examined a large survey of non-elderly adult employed members in a large HMO and found that in both pre- and post-survey years, the non-respondents' mean annual medical expenses were significantly lower than that of respondents.

Studies examining claims and administrative data in the Medicare population have generally found that non-respondents are sicker than respondents. For example, Grotzinger, Stuart, and Ahern (1994) concluded that there are substantial differences in self-reported health status and Medicare expenditures between respondents and non-respondents. They also reported that non-respondents to a large Pennsylvania Medicare enrollee survey on drug use were more likely to have had an admission to a hospital or nursing home, longer lengths of stay in both settings, and higher overall hospital charges for a year. As a result, the authors concluded that without adjustment for non-response, there would be significant underreporting of prescription drug use for this population. Non-respondents were also found to have had longer hospital stays and more hospitalizations with medical diagnoses than surgical diagnoses than respondents (Lasek et al., 1997). In addition, Andresen and colleagues (1996) investigated test-retest reliability and response patterns to the SF-36®. Among community residing seniors surveyed twice, non-response to the baseline survey was highest among the oldest old group and for those with higher levels of the Charlson comorbodity index score (Charlson et al., 1987). Non-response to the followup survey was found to be a more significant problem than for the baseline.

In addition, prior research conducted using the M+C HOS showed similar patterns of non-response for some Medicare subpopulations. Khatutsky and Pope (2002) examined the HOS for M+C Medicare beneficiaries and found that the oldest old, Medicaid enrollees, black persons, and institutionalized persons were less likely to respond. In a separate study, the Health Assessment Lab examined longitudinal outcomes in the HOS for M+C and reported that within the elderly cohort non-respondents were older, non-white, male, and low-income (Rogers et al., 2000).

In summary, these studies indicate that among the Medicare population, non-respondents to mailed surveys are more likely to be minorities, age 85 or over, Medicaid enrollees, and young disabled, and they tend to be sicker, have more inpatient stays and be hospitalized for longer periods. These findings underscore the need to evaluate non-response bias in the Medicare FFS HOS and assess the relative influence of these factors on the estimation of average health status at an aggregate level.

The 1998 Medicare FFS HOS was a large, stratified random sample of Medicare beneficiaries. The health status of beneficiaries in this survey was measured by PCS and MCS computed from the SF-36®. The overall response rate was 65.5 percent. Non-respondents included beneficiaries who refused to participate, those who did not respond to repeated survey mailings and telephone contacts, those who had a bad address or telephone number that were not traceable, those who

completed some parts of the survey but not the items needed to compute PCS and MCS scores, those who were too ill or cognitively impaired to participate, including those with a language barrier, as well as those who died before they could complete the survey.

The Medicare FFS HOS response rate raises concerns that the PCS and MCS scores calculated for respondents may not be representative of the original sample. The purpose of this study was to estimate the magnitude of bias in the component scores using two alternative methods of accounting for non-response—response propensity weighting and regression-based imputations. This study used Medicare enrollment and claims data to develop independent predictors of health status. Further, this study examined the influence of mortality between sample selection and survey administration on the degree of non-response bias in mean PCS and MCS scores.

METHODS

Survey Method

The baseline FFS HOS was administered to 10,000 Medicare FFS beneficiaries evenly divided among 10 samples: a national random sample, five small geographic areas (SGAs), and beneficiaries assigned to four physician group practices (PGPs). The selected SGAs and PGPs were chosen to provide a variety of contrasts between different geographic locations and types of PGPs. A sample was drawn from the 100-percent Medicare enrollment database (EDB), which contains enrollment and entitlement information for all beneficiaries ever enrolled in the Medicare Program. The initial sample was randomly drawn using the four terminal digits of the beneficiary's Social Security number.

Medicare beneficiaries were eligible for the initial selection if they had been continuously enrolled in Medicare FFS for all of CY 1997 and had complete mailing addresses in the EDB. Beneficiaries were omitted from the sampling frame if they were eligible for Medicare through the ESRD program, were Railroad Board Retirees, or were members of an M+C health plan. Further, inclusion in the survey as a part of the PGP sample required that the beneficiary had visited a PGP physician at least once in the prior year and the PGP provided at least as much or more primary care than any other provider. The small geographic area samples were selected from the States of Arizona, Georgia, Pennsylvania, Wisconsin, and Washington; States where the four PGPs were also located. Beneficiary residency in these States at the time of sampling was a requirement.

The FFS HOS was administered May 1998-January 1999. The mode of administration was mail with telephone followup. Medicare beneficiaries who did not complete a mail survey after three mailing attempts were referred for telephone followup and up to 10 telephone calls were placed in an effort to contact the beneficiary. Followups were focused especially on obtaining responses to the 12 items comprising the SF-12® portion of the questionnaire to reduce respondent burden. Proxy respondents were allowed to complete the HOS on behalf of the sampled Medicare beneficiaries.

Calculating PCS and MCS Summary Scores

The core of the HOS consists of the SF-36® questions, which asks the respondent to rate general health, ability to perform certain physical tasks, level of pain, and emotional state (Ware et al., 1994).

Summary scales of the PCS and MCS were calculated using 8 scales based on all 36 questions. Both components were normed such that the mean score was 50 with a SD of 10 points in the general U.S. population.

The component scores could also be computed from a 12-question subset of the SF-36®, the SF-12®. All 12 of these items had to be answered for either the PCS or MCS scores to be computed; no imputations were allowed (Ware et al., 1995). The SF-12® represented the smallest subset of HOS questions that could be used to compute the PCS and MCS. The SF-36® was our preferred scoring method, and was used whenever possible.

Survey Respondent Characteristics

A beneficiary was considered to be a respondent if he or she provided enough survey information to compute PCS and MCS scores using either the SF-36® or SF-12®. The overall study sample included 6,545 respondents and 3,455 non-respondents. Two non-respondents were dropped from all analyses presented in this article because of mismatching with the Medicare EDB. The time period between sampling and surveying was between 3 and 10 months. A total of 379 Medicare beneficiaries died after being selected but prior to completing a survey instrument and were considered non-respondents. Proxy respondents completed 924 surveys. SF-36® scores could be calculated for 82 percent of the respondents.

Health Status Measures Developed from Secondary Data

Three general health status indices were created for respondents and non-respondents using Medicare claims data: the Charlson comorbidity index, the PIP-DCG risk score, and the DCG-HCC risk score.

These indices have been shown to be correlated with health status (Charlson et al., 1987; Pope et al., 1999; 2000). Beneficiaries without a full set of Medicare FFS claims for the period spanning 12 months prior to the date of survey administration or death were subsequently excluded from the claims-based analyses. This included 412 beneficiaries with 1 or more months of the following during the preceding 12-month period: M+C enrollment (n=270); Medicare secondary payer (n=142); and not continuously eligible for both Medicare Parts A and B (n=12). Inclusion of these beneficiaries could bias claims-based estimates of health status since their full medical experience could not be observed if any of the three conditions were present.

The Charlson Clinical Comorbidity index is a weighted sum of selected chronic conditions. The weights take into account the number and the seriousness of the comorbid diseases. We identified these conditions using both principal and secondary diagnoses on inpatient hospital, outpatient hospital, and physician claims, and constructed two alternative Charlson indices: inpatient diagnoses only and one inclusive of inpatient and outpatient diagnoses.

The PIP-DCG model uses both demographic information as well as the principal diagnosis of hospitalizations to predict following year medical expenditures. A risk-adjustment score is calculated by dividing these predicted expenditures by the average cost for the general Medicare FFS population. A risk-adjustment score of 1.0 indicates an average level of predicted future expenditures. Risk-adjustment scores may be used as a measure of relative health status in comparison to the general Medicare population (Pope et al., 1999).

The DCG-HCC model is different from the PIP-DCG model in that it uses diagnoses from inpatient hospital, outpatient

hospital, and physicians as well as certain clinically-trained non-physician diagnoses to predict future Medicare payments (Pope et al., 2000). Diagnoses are grouped into hierarchical diagnostic categories and used together with demographic information to predict future Medicare expenditures. In contrast to the PIP-DCG model, which uses the highest cost diagnosis, the DCG-HCC is a multi-condition additive model, where HCCs are not mutually exclusive. A total predicted expenditure for each beneficiary is the sum of the incremental predicted expenditures associated with each assigned HCC. We used a base prospective payment model that predicts expenditures in the year after survey administration based on the data from the 12 months preceding survey administration or death. A description of the Medicare data used for the health status measures is presented in the Technical Note. (Available on request from authors.)

Response Propensity Weights

Our first non-response adjustment strategy was to estimate the probability of survey response and to weight respondent observations by the reciprocal of this probability (Kessler, Little, and Groves, 1995). The probability of response was estimated as a function of demographic, enrollment, and health characteristics using a logistic regression model. The reciprocal of this probability was used to weight respondents' PCS and MCS scores to represent all survey eligibles. The initial weight was then adjusted so that the sum of the weights equaled the number of respondents, thereby allowing statistical tests of significance to be conducted with proper standard errors.

Imputation of PCS and MCS Scores

Our second strategy used regression-based imputation to impute estimated values for missing component scores for non-respondents (Dillman et al., 2002). Multivariate regression models predicting PCS and MCS scores for respondents were used to derive coefficients for imputing scores. Two models with the same set of independent variables (one predicting PCS and one MCS scores) were estimated. The predictors included basic demographic information, dummy variables for geographic regions, and the alternative claims-based health status measures.

Statistical Significance

Statistical significance of differences in demographic characteristics or claims-based risk scores by response status was assessed through pair-wise one-way ANOVA for continuous variables and with chi-square tests for categorical variables. A 0.05 level of significance (adjusted for multiple comparisons) was used in this article.

RESULTS

Descriptive Results

Three response groups were analyzed to examine non-response due to mortality between sample selection and survey administration and all other causes (respondents, living non-respondents, and decedents). Table 1 compares the demographic and Medicare enrollment characteristics of the three groups. Compared with respondents, living non-respondents were characterized by being older (15.2 versus 10.1 percent, age 85 or over), having a higher proportion of minorities (6.9 versus 3.9 percent,

Demographic	Total Sample[1]			
	Respondents N=6,545	Living Non-Respondents N=3,074	Deceased Before Survey Administration N=379	Statistical Significance
Age	Percent			*, **, ***
Under 65 Years	8.5	13.5	4.8	—
65-74 Years	44.5	36.0	22.7	—
75-84 Years	37.0	35.3	38.5	—
85 Years+	10.1	15.2	34.0	—
Female	58.7	61.0	56.2	—
Race				*
White	93.8	90.1	91.8	—
Black	3.9	6.9	6.1	—
Other/Unknown	2.3	3.1	2.1	—
Dual Medicare-Medicaid Enrollment	10.2	15.5	19.8	*, **, ***
Original Entitlement Due to Disability[2]	5.4	7.1	7.9	*
Census Region				*, ***
Northeast	21.5	23.4	25.1	—
Midwest	28.3	18.0	23.0	—
South	12.9	16.1	15.6	—
West	37.3	42.5	36.4	—
Enrolled in HMO at Least 1 Month [3]	3.0	2.3	1.9	—
Medicare as Secondary Payer[4]	1.5	1.5	0.3	—
Mortality[5]				
Dead 6 Months After Survey Administration	1.4	2.3	NA	*
Dead 12 Months After Survey Administration	3.8	5.6	NA	*

*Statistically significant difference between respondents and living non-respondents at the 0.05 level, adjusted for multiple comparisons.

**Statistically significant difference between respondents and decedents at the 0.05 level, adjusted for multiple comparisons.

***Statistically significant difference between living non-respondents and decedents at the 0.05 level, adjusted for multiple comparisons.

[1] National random sample, five small geographic area samples, and four physician group practice samples. Two respondents were excluded from this analysis because of problems matching their Medicare identification number with the Medicare enrollment database.

[2] Persons age 65 or over on August 1, 1998, originally entitled to Medicare by disability.

[3] Persons with at least 1 month of HMO enrollment in the period between sampling and actual survey administration.

[4] Persons with at least 1 month of Medicare as a secondary payer in the period of 12 months prior to survey administration or death.

[5] Mortality among those alive at time of survey administration.

NOTES: Pair-wise statistical significance of differences by response status are determined with one-way ANOVA with Bonferroni correction for continuous variables and with Chi-square tests for categorical variables. The results of the significance testing is displayed in the last column of the table. Significance levels refer to the entire category, e.g., age, race. NA is not applicable.

SOURCE: Centers for Medicare & Medicaid Services: Data from the 1998 Medicare FFS Health Outcomes Survey and the Medicare Enrollment Database: 1997 and 1998.

black persons), dual Medicare-Medicaid enrollees (15.5 versus 10.2 percent), beneficiaries originally entitled to Medicare due to disability (7.1 versus 5.4 percent), and higher 6- and 12-month mortality rates than respondents (2.3 versus 1.4 percent and 5.6 versus 3.8 percent, respectively). Geographic differences were also noted between respondents and living non-respondents. Decedents were significantly older than both living non-respondents and respondents (34.0 versus 15.2 versus 10.1

Table 2

Prior Year Mean Medicare Payments, Percent Users of Services, Hospital Use, and Health-Status for Medicare FFS Health Outcomes Survey Eligibles: 1998

Category	Respondents N=6,267	Living Non-Respondents N=2,948	Deceased Before Survey Administration N=371	Statistical Significance
Total Medicare Expenditures	$4,014	$4,125	$23,804	**, ***
Percent Users	95.1	91.0	98.1	*, **, ***
Inpatient Expenditures	$1,871	$1,910	$12,913	**, ***
Percent Users	17.3	18.1	71.7	**, ***
Hospital Outpatient Expenditures	$406	$402	$1,295	**, ***
Percent Users	66.2	62.2	85.7	*, **, ***
Part B (Physician, Professional) Expenditures	$1,173	$1,051	$3,758	*, **, ***
Percent Users	93.6	89.2	97.0	*, **, ***
Home Health Expenditures	$243	$304	$1,722	**, ***
Percent Users	7.9	8.3	40.4	**, ***
Durable Medical Equipment	$111	$134	$559	**, ***
Percent Users	19.5	18.7	48.8	**, ***
Hospice Expenditures	$27	$58	$1,052	**, ***
Percent Users	0.25	0.42	19.1	**, ***
Skilled Nursing Facility Expenditures	$181	$267	$2,506	**, ***
Percent Users	2.8	3.5	28.0	*, **, ***
Hospitalization Utilization				
Mean Number of Hospital Discharges	0.27	0.3	1.62	**, ***
Mean Number of Inpatient Days	1.67	2.46	13.5	*, **, ***
Number of Hospitalizations (Percent)				
0	82.6	81.7	28.0	**, ***
1	11.6	11.2	31.3	**, ***
2	3.6	4.8	16.7	*, ***
3 or More	2.2	2.3	24.0	**, ***
Charlson Comorbidity Index				
Inpatient Diagnoses Only	0.27	0.29	2.81	**, ***
Inpatient and Ambulatory Diagnoses	0.4	0.42	3.12	**, ***
Mean-Risk Adjustment Score				
PIP-DCG	0.98	1.07	2.58	*, **, ***
(Standard Error)	(0.008)	(0.014)	(0.082)	—
DCG-HCC	1.0	1.07	3.35	*, **, ***
(Standard Error)	(0.011)	(0.018)	(0.099)	—

*Statistically significant difference between respondents and living non-respondents at the 0.05 level, adjusted for multiple comparisons.

**Statistically significant difference between respondents and decedents at the 0.05 level, adjusted for multiple comparisons.

***Statistically significant difference between living non-respondents and decedents at the 0.05 level, adjusted for multiple comparisons.

[1] National random sample, five small geographic area samples, and four physician group practice samples.

NOTE: Pair-wise statistical significance of differences by response status are determined with one-way ANOVA with Bonferroni correction for continuous variables and with chi-square tests for categorical variables.

SOURCE: Centers for Medicare & Medicaid Services: Data from the 1998 Medicare FFS Health Outcomes Survey and the Medicare Enrollment Database: 1997 and 1998.

percent, age 85 or over, respectively), and more likely to be a dual enrollee than both living non-respondents and respondents (19.8 versus 15.5 percent versus 10.2 percent, respectively). Geographic variation between the decedents and living non-respondents was also observed.

Table 2 displays health care use rates among the three groups. There were no statistically significant differences in total Medicare expenditures between respondents and living non-respondents. However, respondents had higher physician expenditures than living non-respondents ($1,173

versus $1,051) and used more Medicare services, in total, and outpatient and physician services, in particular (95.1 versus 91.0 percent, 66.2 versus 62.2 percent, and 93.6 versus 89.2 percent, respectively). In contrast, living non-respondents had higher rates of SNF usage than respondents (3.5 versus 2.8 percent). There were no differences in the number of hospitalizations among these two groups; however, living non-respondents had more inpatient days on average (2.5 versus 1.7). Both groups scored similarly on the Charlson comorbidity indices, indicating similar overall disease burden in the year prior to survey. However, living non-respondents' PIP-DCG and DCG-HCC risk scores were 7-9 percent higher, signifying higher predicted Medicare expenditures in the year they were surveyed.

As expected, decedents were vastly different from the other two response groups and represented a substantially sicker group with higher medical expenses. In comparison with respondents and living non-respondents, decedents had substantially higher rates of medical care utilization and incurred substantially higher Medicare expenditures. Total Medicare expenditures for decedents were six times the level of expenditures for respondents and living non-respondents ($23,804 versus $4,014 and $4,125, respectively). This pattern held for each type of service, i.e., inpatient, outpatient, home health, etc. Decedents experienced more episodes of hospitalization and more inpatient days. A total of 24 percent of decedents were hospitalized three or more times during the year prior to their deaths. This was in sharp contrast to only 2 percent of respondents and living non-respondents. The mean number of inpatient days was 13.5 for decedents and only 1.7 and 2.5 days for respondents and living non-respondents, respectively. Decedents also had signifi-

cantly worse health status. The significantly higher Charlson comorbidity scores (3.12 for decedents versus 0.4 for all others) indicated that both a greater percentage of decedents had comorbid conditions and they had a greater number of comorbid conditions. The PIP-DCG and DCG-HCC risk scores were two and one-half and three and one-third times higher than for the other two groups, respectively.

Response Propensity Modeling

Respondents' PCS and MCS scores were weighted for non-response by the inverse of the response probabilities derived from a logistic regression model. To choose the best response propensity model, we evaluated five models with various measures of health status. These models are displayed in Table 3. Each model was estimated for living non-respondents and for all non-respondents, including decedents. We started with the basic demographic model that included sex, race, age, program enrollment, and geographic regions. This model is relevant for studies that do not have access to secondary Medicare data. The overall fit for the basic demographic model yielded a chi-square value of 352.11 and a Pseudo R^2 of 0.035.

We then added one health status measure at a time and with each addition assessed model fit, Pseudo R^2, and performed a likelihood ratio test to determine whether the new model was an improvement on the reduced form demographic model. A likelihood ratio test, which employs the G-statistic, was used to evaluate the superiority of the reduced form model relative to the full model. P-values of less than 5 percent led to the conclusion that the tested model was an improvement over the demographic mode. Using the chi-square test, Pseudo R^2 and G-statistic as our measures of overall model fit, Model 1

Table 3

Medicare FFS Health Outcomes Survey Eligibles[1] Alternative Health Status Logistic Regression Models for Estimating Likelihood of Response (Deceased Included): 1998

Variable	Model 1 Odds Ratio	Model 2 Odds Ratio	Model 3 Odds Ratio	Model 4 Odds Ratio	Model 5 Odds Ratio
Intercept	**	**	**	**	**
Female	*0.91	0.92	0.93	0.94	0.95
Black	**0.66	**0.65	**0.67	**0.67	**0.66
Other Non-White	0.96	0.95	0.95	0.97	0.94
Under 65 Years	**0.64	**0.64	**0.64	**0.65	**0.65
75-84 Years	0.92	*0.89	**0.84	**0.82	**0.83
85 Years+	**0.58	**0.55	**0.48	**0.46	**0.48
Originally Disabled	0.84	*0.79	**0.72	**0.69	**0.71
Medicaid	*0.83	**0.77	**0.73	**0.71	**0.72
Midwest Region	**1.68	**1.60	**1.68	**1.69	**1.64
South Region	1.01	0.98	1.00	1.00	1.00
West Region	0.91	0.90	0.92	0.93	0.92
PIP-DCG Score	**0.73	—	—	—	—
DCG-HCC Risk Score	—	**0.81	—	—	—
Charlson Inpatient Index	—	—	***0.84	—	—
Charlson Inpatient+Outpatient Index	—	—	—	**0.87	—
Total Medicare Expenditures	—	—	—	—	**1.00
Overall Chi-Squared	**474.44	**456.18	**455.93	**368.17	**417.03
Pseudo R^2	0.047	0.045	0.045	0.037	0.042
G-Statistic[2]	**449.443	**101.619	**101.184	**16.034	**63.585

*Significant at $p = <0.05$ level.

**Significant at $p = <0.01$ level.

[1] National random sample, five small geographic area samples, and four physician group practice samples.

[2] G-statistic for comparison with reduced form demographic model.

NOTES: $N = 9,568$. Excludes beneficiaries without a complete set of FFS claims over the prior year. Northeast Region is omitted region in the multivariate regression.

SOURCE: Centers for Medicare & Medicaid Services: Data from the 1998 Medicare FFS Health Outcomes Survey and the Medicare Enrollment Database: 1997 and 1998.

(chi-sq-474.44; Pseudo R^2=0.047; G=449.443, p<0.0001), which included the PIP-DCG score as a measure of health status, gave the best claims-based predictor of survey response. One possible reason is that the PIP-DCG score is a better indicator of the disease severity than the DCG-HCC risk score, which contains inpatient and ambulatory care diagnoses, since it focuses only on hospitalizations.

Table 4 displays the preferred model for estimating the likelihood of response (including deceased non-respondents). We observed that all but five of the characteristics in the model were associated with the probability of response at the 0.05 level

of significance or better (Table 4). The following demographic characteristics reduced the probability of responding to the FFS HOS: being female (9 percent less likely than males), black persons (34 percent less likely than white persons), young disabled or those whose age is under 65 (36 percent less likely than Medicare beneficiaries age 65-74 years), and age 85 or over (42 percent less likely to respond than Medicare beneficiaries age 65-74). Enrollment in Medicaid significantly reduced the likelihood of responding to the survey. Beneficiaries residing in the Midwest region of the United States were significantly more likely to respond compared

Table 4

Medicare FFS Health Outcomes Survey Eligibles[1] Logistic Model of Estimating Likelihood of Response: 1998

Variable	Deceased Included[2] Odds Ratio	Deceased Excluded[3] Odds Ratio
Female	*0.91	*0.90
Black Race	**0.66	**0.66
Other Non-White Race	0.96	0.96
Under 65 Years	**0.64	**0.61
75-84 Years	0.92	*0.88
85 Years+	**0.58	**0.58
Originally Disabled	0.84	**0.74
Medicaid	*0.83	**0.78
Midwest Region	**1.68	**1.71
South Region	1.01	0.99
West Region	0.91	0.90
PIP-DCG Risk Score	**0.73	*0.93
Overall Chi-Squared (*p*-value)	474.44 (0.0001)	298.613 (0.0001)
Pseudo R^2	0.047	0.031

*Significant at $p<0.05$ level.

**Significant at $p<0.01$ level.

[1] National random sample, five small geographic area samples, and four physician group practice samples.

[2] N=9,568.

[3] N=9,215.

NOTES: Excludes beneficiaries without a complete set of FFS claims over the prior year. Northeast Region is the omitted region in the multivariate regression.

SOURCE: Centers for Medicare & Medicaid Services: Data from the 1998 Medicare FFS Health Outcomes Survey and the Medicare Enrollment Database: 1997 and 1998.

with beneficiaries residing in the Northeast. In addition, increases in PIP-DCG scores suggesting poorer health status led to substantial reductions in the probability of being a respondent. A similar pattern of effects was found when decedents were excluded, although this model had a slightly poorer overall fit. While the model chi-square values were highly significant due to the large sample size, the pseudo R^2 values were small.

Imputation of PCS and MCS Scores for Non-Respondents

Using demographic, program enrollment, and geographic characteristics together with the PIP-DCG score as predictors, we estimated regression models predicting PCS and MCS scores for respondents. In a second step, we used the coefficients from this model to impute the scores for non-respondents. Table 5 pre-sents the regression modeling results. PCS scores were predicted to be about 2 points lower for females, more than 10 points lower for young Medicare beneficiaries with disabilities (under age 65) compared with beneficiaries age 65-74, and almost 7 points lower for elders age 85 or over compared with those age 65-75. Beneficiaries originally entitled to Medicare due to disability were predicted to have scores 9 points lower than Medicare beneficiaries aging into the Medicare Program. Additionally, a 1-unit increase in the PIP-DCG score lowered the PCS score by about 4 points.

Similar characteristics affected MCS scores, but exerted less influence. Beneficiaries entitled to Medicare because of disability (under age 65) and elders age 85 or over were predicted to have MCS scores 11 and 3 points lower than those age 65-74, respectively. Beneficiaries originally disabled were predicted to have scores

Table 5

Medicare FFS Health Outcomes Survey Respondents[1] Regression Models Used for Imputing SF-36® Scores: 1998

Variable	Physical Component Summary Score Parameter Estimate	Mental Component Summary Score Parameter Estimate
Intercept	**47.11	**54.20
Female	**-2.21	*-0.53
Black Race	-0.49	-1.21
Other Non-White Race	0.14	-1.05
Under 65 Years	**-10.15	**-10.55
75-84 Years	**-3.56	**-0.93
85 Years+	**6.91	**-3.27
Originally Disabled	**-9.05	**-2.81
Medicaid	-0.95	**-2.82
Midwest Region	-0.32	**1.36
South Region	-0.28	0.37
West Region	0.30	**1.41
PIP-DCG Risk Score	**-4.00	**-1.91
R^2	0.18	0.13
F-Value	111.39	74.62
P>*F*	0.0001	0.0001

*Significant at *p*<0.05 level.

**Significant at *p*<0.01 level.

[1] National random sample, five small geographic area samples, and four physician group practice samples.

NOTES: N=6,267. Excludes beneficiaries without a complete set of FFS claims over the prior year. Northeast Region is the omitted region in the multivariate regression.

SOURCE: Centers for Medicare & Medicaid Services: Data from the 1998 Medicare FFS Health Outcomes Survey and the Medicare Enrollment Database: 1997 and 1998.

lower by about 3 points compared with scores predicted for those who age into the Medicare Program. Dual Medicare-Medicaid enrollees had a stronger effect on the mental health component score, decreasing the MCS score on average by 2.8 points, whereas respondents from the Midwest and West had higher scores compared with respondents in the Northeast region. The PIP-DCG score also had a smaller effect on MCS scores with a 1-unit increase leading to about a 2-point decrease. The two models explained 18 and 13 percent, respectively, of the variance in the component scores.

Comparison of Weighted and Imputed MCS and PCS Scores

Table 6 compares the unweighted mean PCS and MCS scores with mean scores adjusted for non-response and presented separately for the inclusion and exclusion of decedents. For Medicare FFS HOS respondents, the mean PCS score was 38.38. This was lower than the general population mean because Medicare enrolls an aged and disabled population. The mean MCS score was 50.89, which approximates the norm for the general U.S. population. Adjusted for non-response with propensity weights, the mean PCS score declined by 0.39 points with the deceased excluded as non-respondents. Including decedents, there was a slightly larger downward adjustment (0.63 points). When imputed scores were used for all living beneficiaries, the average PCS score declined by 0.45 points. Including the deceased as non-respondents, there was a 0.74 point decline in the average PCS score.

The MCS scores exhibited even smaller changes. Adjusted for non-response with propensity weights, the mean MCS scores declined by 3/10ths of a point when decedents were excluded and by just under one-half of a point when decedents were included. The mean MCS score was 0.37

Table 6

Medicare FFS Health Outcomes Survey Sample,[1] Differences in Mean PCS Scores and MCS Scores Unweighted, Imputed, and Adjusted for Survey Non-Response with Propensity Weights: 1998

Category	Unweighted Respondents Only	Adjusted for Non-Response with Propensity Weights		Imputed for Non-Response		Difference Between Propensity Weighted and Unweighted Scores		Difference Between Imputed and Unweighted Scores	
		Excluding Deceased	Including Deceased	Excluding Deceased	Including Deceased	Excluding Deceased	Including Deceased	Excluding Deceased	Including Deceased
Mean PCS Score	38.38	37.99	37.75	37.93	37.64	-0.39	-0.63	-0.45	-0.74
Mean MCS Score	50.89	50.56	50.43	50.52	50.38	-0.33	-0.46	-0.37	-0.51

[1] National random sample, five small geographic area samples, and four physician group practice samples.

NOTES: Excludes beneficiaries without a complete set of FFS claims over the prior year. PCS is physical component summary. MCS is mental component summary.

SOURCE: Centers for Medicare & Medicaid Services: Data from the 1998 Medicare FFS Health Outcomes Survey and the Medicare Enrollment Database: 1997 and 1998.

points lower with imputation for living non-respondents, and 0.51 points lower when decedents were included.

CONCLUSIONS

The overall response rate for the Medicare FFS HOS was 65.5 percent when Medicare beneficiaries who died prior to survey administration were included as non-respondents. This rate is not unusual for contemporary mail surveys, but raises questions about how representative the observed health status results are for the FFS sample as a whole. In this article, we used two different approaches to estimate the likely degree of bias in PCS and MCS scores attributable to non-response.

The two approaches—weighting by the inverse of the response propensity and regression-based imputation of missing scores—gave nearly identical estimates of non-response bias. PCS scores were estimated to be 0.74 points lower and MCS scores 0.51 points lower after adjustment for non-response through imputation and 0.63 and 0.46 lower, respectively, after adjustment with propensity weighting. These levels are comparatively small for component scores with SDs of 10 points.

The degree of non-response bias in any survey is a function of the (unknown) magnitude of the difference in scores between respondents and non-respondents, and the proportion of non-respondents. Non-respondents who were alive at the time of survey administration comprised 30 percent of the original sample, but were estimated to have component scores that were only slightly lower than those for respondents. Decedents, on the other hand, were predicted to have considerably poorer health status than respondents, but comprised only 3.8 percent of the sample. As a result, the amount of bias jointly contributed by these two groups of non-respondents was comparatively small.

There is little doubt that FFS survey respondents, living non-respondents, and decedents are different in many respects. We found several differences among these three groups in terms of demographic and enrollment characteristics. One of the strengths of our analysis is that we were able to examine Medicare claims data for all beneficiaries even if they did not complete the HOS. The groups also differed with respect to claims-based alternative indicators of health status, including the prevalence of chronic health conditions, morbidity and mortality indices, and expenditure-based risk indices. These alternative indicators yielded a consistent pattern of results—health status was most favorable for respondents, slightly less favorable for non-respondents who were alive, and considerably worse for decedents.

The usefulness of regression-based imputation depends heavily on the availability of predictors that are strongly associated with outcomes. Our imputation models explained only 13-18 percent of the variance in the component scores. As a result, the range of scores predicted by the model may have been fairly restricted. We found that the risk-adjustment scores for future medical expenditures were more highly correlated with the PCS and MCS than the cruder Charlson indices. When health status measures were individually evaluated, the model with the PIP-DCG risk-adjustment score produced the best fit for purposes of our study and was chosen for both response propensity weighing and imputation. One possible explanation for its superiority is that the PIP-DCG score measures severity of medical conditions as it is based on hospitalizations. It seems that addition of outpatient data is not particularly helpful in achieving better prediction for the outcomes in this study. This finding does not imply that the PIP-DCG model

would outperform other models in terms of predicting future expenditures for Medicare risk-adjusted payment purposes.

The utility of the response propensity approach also depends on how closely our equation approximates the actual non-response mechanism in the FFS HOS. Our model strongly suggests that poor health decreases the likelihood of responding because the PIP-DCG and other surrogates for health status, such as age and disability, were significant explanatory variables. However, the decision to complete the HOS survey is likely to be a complex one that also depends on situational factors, attitudes, and psychological traits. Other reasons for non-response by older adults include such factors as unfamiliarity with surveys, the cognitive tasks imposed by the instrument, privacy concerns, resistance to participation in government-sponsored research, the unavailability of proxies, and the belief that the survey was not salient for those who do not have serious health problems. Adding measures for these factors might have improved the fit of the response propensity model and could have altered our estimates of the amount of bias.

An additional limitation is that most of the analyses were performed on the sample of FFS HOS eligibles with a full set of Medicare claims. Those with at least 1 month of HMO enrollment or Medicare as a secondary payer, as well as those without consistent coverage by Medicare Part A and B were excluded. These beneficiaries could be systematically different from those that were retained.

While the true extent of non-response bias in the HOS FFS cannot be determined, our analyses suggest that the bias is likely to be small. This is particularly true if decedents are classified as ineligible. When non-response adjustments are desired, our models indicate that response propensity weighting and imputation yield adjusted means that are very similar.

REFERENCES

Andresen, E.M., Bowley, N., Rothenberg, B.M., et al.: Test-Retest Performance of a Mailed Version of the Medical Outcomes Study 36-Item Short-Form Health Survey Among Older Adults. *Medical Care* 34(12):1165-1170, 1996.

Charlson, M.E., Pompei, P., Ales, K.L., et al.: A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *Journal of Chronic Diseases* 40(5):373-383, 1987.

Dillman, D.A., Eltinge, J.L., Groves, R.M., et al.: Survey Non-response in Design, Data Collection, and Analysis. In Groves, R.M., Dillman, D.A. Eltinge, J.L. et al. (eds). *Survey Non-response*. John Wiley & Sons, Inc. New York., 2001.

Fowler, F.J., Gallagher, P.M., Stringfellow, V.L. et al.: Using Telephone Interviews to Reduce Non-response Bias to Mail Surveys of Health Plan Members. *Medical Care* 40(3):190-200, 2002.

Fowles, J.B., Weiner, J.P., Knutson, D., et al.: *A Comparison of Alternative Approaches to Risk Measurement*. Park Nicollet Medical Foundation. Final Report. Physician Payment Review Commission Grant Number 93-G07. Minneapolis, MN. November 1994.

Grotzinger, K.M., Stuart, B.C., and Ahern, F.: Assessment and Control of Non-response Bias in a Survey of Medicine Use by the Elderly. *Medical Care* 32(10):989-1003, 1994.

Hornbrook, M.C. and Goodman, M.J.: Assessing Relative Health Plan Risk with the RAND-36 Health Survey. *Inquiry* 32(1):56-74, 1995.

Kessler, R.C., Little, R.J., and Groves, R.M.: Advances in Strategies for Minimizing and Adjusting for Survey Non-response. *Epidemiologic Reviews* 17(1):192-204, 1995.

Khatutsky, G. and Pope, G.C.: Analysis of Non-Response Bias in the Medicare Health Outcomes Survey. Health Economics Research, Inc., Waltham, MA. February 2002.

Lasek, R.J., Barkley, W., Harper, D.L., et al.: An Evaluation of the Impact of Non-response Bias on Patient Satisfaction Surveys. *Medical Care* 35(6):646-652, 1997.

Pope, G.C., Liu, C.F., Ellis, R.P., et al.: Principal Inpatient Diagnostic Cost Group Models for Medicare Risk Adjustment. Health Economics Research, Inc., Waltham, MA. February 24, 1999.

Pope, G.C., Ellis, R.P., Ash, A.S., et al.: Diagnostic Cost Group Hierarchical Condition Category Models for Medicare Risk Adjustment. Health Economics Research, Inc. Waltham, MA. December 21, 2000.

Rogers, W.H., Ware, J.E., Gandek, B., et al.: Examining Longitudinal Outcomes in the Medicare Health Outcomes Survey: Preliminary Recommendations. Boston, MA. September 2000.

Rowland, M.L. and Forthofer, R.N.: Adjusting for Non-response Bias in a Health Examination Survey. *Public Health Reports* 108(3):380-386, 1993.

Ware, J.E., Snow, K.K., Kosinski, M., et al.: *SF-36® Physical and Mental Health Summary Scales: A User's Manual*. Boston. The Health Institute, New England Medical Center, 1994.

Ware, J.E. Kosinski, M., and Keller, S.D.: *How to Score the SF-12® Physical & Mental Health Summary Scales*. Second Edition. The Health Institute, New England Medical Center. Boston, MA. 1995.

Reprint Requests: Nancy T. McCall, Sc.D., RTI International, 1615 M Street NW, Suite 740, Washington, DC 20036. E-mail: NMcCall@rti.org

TECHNICAL NOTE

Construction of the Charlson Comorbidity Index and the DCG-HCC and PIP-DCG Scores for Medicare FFS Beneficiaries Selected for the 1998 FFS HOS

This technical note provides a description of three claims-based measures of health status used to study non-response bias: the Charlson comorbidity index, the PIP-DCG risk score, and the DCG-HCC risk score. It should be noted that the DCG-HCC model is not identical to the CMS-HCC model used for M+C reimbursement, but they are closely related. Both are prospective, using the prior year's diagnoses to predict expenditures, and they both utilize the same underlying diagnostic classification and demographic categories. But the CMS-HCC model includes a smaller number of diagnostic categories that were selected for use in M+C payments.

Demographic, program participation, and HMO enrollment information for 1997 and 1998 was obtained from the Medicare EDB. Principal and secondary diagnoses, service utilization, and Medicare expenditure data were extracted from the 1997 and 1998 100 percent Medicare provider analysis and review files that contain inpatient hospital services, and the 100 percent national claims history standard analytic files that contain hospital outpatient, SNF, home health, hospice, and DME services.

Since the risk-adjustment indices require 12 months of data to predict future Medicare expenditures, we used 1 year as the timeframe for creating a comprehensive health profile for all three health status measures. The FFS HOS sampling frame was constructed based on a March 5, 1998, writeoff of the EDB. The FFS HOS was administered on a staggered schedule. Mailing dates varied for the 10 different subsamples ranging from May 1998-January 1999. We used the date of the first mailing for each sample to calculate a backward-looking 12-month period for selecting claims. For people who died in the months between the survey frame construction and the actual survey mailing date (N=379), we examined claims with the discharge or service date falling into the period of 12 months preceding the date of death.