Evaluating the Two-Year Follow-up Health Status of Medicare Fee-For-Service Beneficiaries Using the Health Outcomes Survey

Final Report

Prepared by:

Michael G. Trisolini, Ph.D., M.B.A.
Nancy T. McCall, Sc.D.
Gregory C. Pope, M.S.
Michelle Klosterman, B.A.

RTI International

and

Kevin Smith, M.A.

New England Research Institutes

November 25, 2002

Nancy McCall, Sc.D.	**Janet B. Mitchell, Ph.D.**
Project Director	**Scientific Reviewer**

The research presented in this report was performed under Centers for Medicare & Medicaid Services (CMS) Master Order Contract No. 500-95-0058 Task Order No. 2, Zhoowan Jackson, Project Officer. The statements contained in this report are solely those of the authors and no endorsement by CMS should be inferred or implied.

Table of Contents

Page

Table of Contents
(continued)

Page

Table of Tables and Figures

Page

Chapter 2

Chapter 3

Table of Tables and Figures
(continued)

Page

Chapter 4

Chapter 5

Acknowledgements

The following individuals are acknowledged for their contributions to this document: Ann Larsen, Anita Bracero, Norma DiVito, Nora Rudenko, and Michelle Griggs.

1 Introduction and Background

1.1 Introduction

This report presents an analysis of follow-up results for the Health Outcomes Survey (HOS), which was administered to a sample of Medicare Fee-for-Service (FFS) beneficiaries in 1998 and 2000. The primary goal of this analysis is to assess the feasibility of using longitudinal estimates of self-reported health status for cohorts of Medicare beneficiaries to evaluate the care provided to FFS beneficiaries by physician group practices or by FFS providers in small geographic areas. For this study, we evaluated the performance of four large multi-specialty physician group practices (PGPs) and the health care systems serving Medicare FFS beneficiaries in five small geographic areas (SGAs). This analysis builds upon previous work evaluating the feasibility of implementing performance measurement in Medicare FFS that has been conducted by RTI International (McCall *et al*, 1998, 2000; Pope *et al*, 2000; Khatutsky *et al*., 2001) for the Centers for Medicare & Medicaid Services (CMS).

CMS is required by Congress to broadly disseminate information to Medicare beneficiaries to promote informed choice regarding the options they have for receiving treatment under Medicare. This includes information on quality and performance of both managed care plans and FFS providers. The HOS was designed to complement other HEDIS measures for assessment of quality of care for Medicare beneficiaries by enabling

analysis of changes in self-reported health status over time. The HOS contains the SF-36 survey instrument that provides estimates of both physical and mental health through calculation of the physical component summary (PCS) and mental component summary (MCS) scores. Measurement of a series of HEDIS quality indicators, including the HOS, has been applied for several years to Medicare+Choice (M+C) plans.

While other studies have compared FFS and managed care systems and compared different managed care plans serving M+C beneficiaries (Ware *et al*., 1996; HSAG, 2001), feasibility issues related to using self-reported health status, as derived from the HOS, for a nonenrolled, FFS population for performance measurement have not previously been studied. Further, there are two methodological issues related to measuring change in health status that have been contentious topics for many years that required consideration in this project and influenced our methodological approach: reliability of change scores at the individual person level and handling death between the baseline and follow-up time period. Each is discussed briefly, in turn.

Change scores, differences between baseline and follow-up measures, calculated at the individual person level are problematic for two reasons. First, because measures of health status at each time point are subject to random error, the resulting change scores can contain considerable measurement error. As a result, change scores tend to be unreliable. A second problem is that change scores are implicitly based on the assumption that the regression slope for the baseline measure is exactly 1.0. If this assumption is incorrect, then change scores will be negatively correlated with baseline

scores. This problem is one reason that regression adjustments using the baseline as a covariate are frequently advocated for analyzing change.

We believe that the shortcomings of estimating change scores at the individual level is not particularly relevant to our study given that the focus of our evaluation is the feasibility of using follow-up health status as a performance measure. Thus, we focus on deriving the follow-up PCS and MCS means for FFS cohorts as a whole. Random errors for individuals can be expected to cancel out when estimating the group mean. Further, we take a methodological approach that compares expected mean PCS and MCS follow-up scores with actual mean PCS and MCS follow-up scores for our nine FFS cohorts of interest. Individual change scores are not used in our method. However, baseline values of PCS and MCS are used as covariates in our prediction models.

A second issue related to conducting longitudinal studies with traditional health status measures, such as the PCS and MCS, is that they do not provide explicit values for death. Given that approximately 5 percent of Medicare beneficiaries die each year (Gage *et al*., 2000), this issue is especially salient for studies involving Medicare beneficiaries, who have a higher death rate than the general population. Thus, we expected a significant number of the beneficiaries who responded to our baseline survey would die before the follow-up survey was administered two years later.

There is no standard convention for scoring death for either the MCS or PCS. Many longitudinal studies using the SF-36 simply ignore deaths, and analyze changes over two or more years in PCS and MCS scores only for those alive at follow-up. However, Diehr *et al*. (1995) have shown that this approach underestimates changes in

health status, and can significantly bias comparisons of the performance of different health care plans or providers.

Another method is to use a PCS or MCS score of zero for death. This was one of two methods used by Ware *et al*. (1996) for handling death for the PCS in their analysis of data from the Medical Outcomes Study (MOS). An arbitrary score of zero, however, does not represent the "absence" of health and has no explicit meaning on a component score metric, other than being five standard deviations below the general population mean for the PCS. Moreover, the extreme nature of a zero value means that deaths dominate analysis of change scores or follow-up scores.

A second method, employed as the alternate approach by Ware *et al*. (1996) for analyzing change scores, is to collapse individual changes in health status over time in the PCS and MCS into three categories, depending on whether the changes are "better," "worse," or "about the same" as expected. All deaths were assigned to the "worse" category for the PCS in the MOS. However, this approach has two shortcomings. First, it results in a categorization that is considerably less precise than the original continuous PCS score data. Second, it treats all deaths as though they represent the same amount of change in health status. A person with a baseline score of 50 on the PCS who later died is placed in the same category as a much sicker person whose baseline score was 25, even though the declines in health for these two people can be considered quite different. Moreover, Ware *et al*. (1996) treated MCS scores differently, defaulting to the approach we described initially above, where respondents who died between baseline and follow-up were simply excluded from the analysis of MCS results.

More recently, Ware and his colleagues at the Health Assessment Laboratory (HAL) modified their method for handling deaths for PCS scores as part of their analysis of HOS surveys of M+C beneficiaries (Rogers *et al*., 2000; 2001). This new method is also based on categorizing respondents into "better", "same", or "worse" categories at follow-up based upon individual change scores. However, it employs a two-part model for analysis of PCS outcomes, estimating both the probability of death and the probability that PCS scores are the same or better. The outcome assessed is then the expected probability that the beneficiary is alive and has a follow-up PCS score the same or better as the PCS score at baseline. However, the new method continues to treat MCS scores differently, still excluding scores for any respondents who died between baseline and follow-up (Rogers *et al*., 2000; 2001).

A third method, recently proposed by Diehr *et al*. (2001), transforms SF-36 scores into a new metric ranging from 0 (death) to 95 (excellent health). This approach is based on the probability that a person will be alive or healthy at some point in the future. A number of other methods have been proposed as well. (Diehr *et al*., 1995; Rogers *et al*., 2000). These include removing dead participants from both the baseline and follow-up samples, assigning death some other extreme value besides zero, and imputing the lowest observed score. These methods tend to have predictable biases (Rogers *et al*., 2000). The strategies that give less influence to deaths, such as by omitting them, tend to show more favorable average changes in health status over time. Therefore, they tend to show better performance for groups with more deaths. Conversely, the methods most

influenced by deaths, such as assigning a zero value, show negative changes in health status over time. They tend to favor groups with fewer deaths.

We believe that none of the approaches described above for handling deaths between baseline and follow-up are ideal for performance measurement, so we sought to develop a new method of imputing scores for death that would preserve the original continuous metric of the component scales and retain all baseline respondents who died prior to follow-up. To do this, we made use of a concept from economic evaluations of health known as "utilities." Utilities are preferences for health states (Petitti, 2000). Utilities are especially appealing in this context because they are defined on a scale ranging from 0 (death) to 1 (optimal health). Thus, death has a specific value in the utility approach. We relied on questions from the HOS to estimate utilities of various self-reported health states, including death, and then used those values to impute corresponding follow-up PCS and MCS scores for decedents and retained all decedents in the follow-up analyses.

1.2 Methodological Approach

The feasibility of using PCS and MCS follow-up to measure performance in Medicare FFS will be assessed in two ways. First, we conduct a descriptive comparison between baseline and follow-up respondents to determine whether there are systematic differences in respondents to the baseline HOS versus the follow-up HOS, or whether there are systematic differences in the completeness of survey responses as the scoring of the PCS and MCS measures are highly dependent on item response. We also explore

differences in mean PCS and MCS scores between baseline and follow-up and directly examine the effect of setting the PCS score to zero at follow-up for decedents. We include analysis with the follow-up PCS of zero for decedents since, as noted, that was one of the approaches taken for analysis in the MOS (Ware, *et al*., 1996). Our main approach for this study was a different method for imputing values for decedents, but we also included this approach for comparison. Lastly, we explore the degree of retention of baseline respondents at follow-up by the four physician group practices as a measure of face validity of the follow-up scores.

Second, we evaluate the feasibility of using follow-up health status to compare the performance of nine FFS cohorts of interest. We evaluate two alternative methods that compare expected with actual follow-up health status. The first estimation method is developed in this report by RTI. It is based on a regression model of expected mean follow-up health status as a function of baseline health status and a limited set of other independent variables. The second method is the one being used by the Health Assessment Lab (HAL) to evaluate performance of M+C health plans. One of the differences between these two estimation methods is how follow-up physical and mental health status is estimated for baseline respondents who die before the follow-up survey is fielded. A second difference is use of individual change score calculations in the HAL method to estimate better, same or worse health status at follow-up prior to comparing the predicted with actual follow-up health status. Our principal focus, however, is on comparing the two statistical estimation methods in identifying better and worse performing PGPs and SGAs.

1.2.1 Descriptive Comparisons between Baseline and Follow-up Respondents

We begin our descriptive comparison between baseline and follow-up respondents by analyzing response rates and the distribution of respondents across sociodemographic and health status measures at baseline and follow-up to determine if there are systematic changes in the two populations at the two time points. The response rate analysis is conducted for the 10 cohorts and selected sociodemographic and health status characteristics.

Second, we compare and contrast the scoring methods that are used between baseline and follow-up for FFS respondents. In our study, we used two different scoring methods to obtain estimates of PCS and MCS; one based on a 36 question item scoring algorithm (SF-36) and a second based on a 12 question item scoring algorithm (SF-12). Use of the 12 item scoring method allowed us to retain in the study a segment of the FFS population who did not complete the entire HOS survey instrument. Our analysis provides an assessment of the degree of completeness of the survey-based health status responses that are used to produce the baseline and follow-up PCS and MCS scores. In an earlier analysis of differences in mean PCS and MCS scores calculated using the two alternative scoring algorithms, we found that the 12 item scoring algorithm generated less of a range in PCS and MCS scores than did the 36 item scoring algorithm, and slightly higher mean PCS and slightly lower mean MCS scores than the 36 item scoring algorithm for our cohorts of interest (McCall *et al*., 2000). Thus, a significant shift in the

proportions being scored with each method between baseline and follow-up could result in a spurious measured difference in health, *e.g.*, a measured change when no actual change occurred.

Third, we display mean PCS and MCS scores at baseline and follow-up for our FFS beneficiaries. In this analysis, we set the PCS score to 0 for decedents and retain them in the follow-up PCS calculation. However, decedents are removed from the change score calculation for the MCS. This is the method previously used in the MOS to analyze health status as measured by the PCS and MCS. Differences in mean physical and mental health scores are compared, in total, for the national random sample, four physician group practices and five small geographic areas, and across different sociodemographic and health status measures. In this analysis, we directly examine the influence of setting the PCS to zero for decedents on differences in mean scores for our cohorts of interest.

Lastly, we examine the proportion of follow-up respondents that identified their usual source of care as the same physician group practice to which they were assigned at baseline for our four group practices. This answers a face validity question regarding ownership of follow-up health status.

1.2.2 Comparison of Expected and Follow-up Health Status Methods

Two alternative methods that compare expected with actual follow-up health status are evaluated. The first method, developed in this report by RTI, compares predicted versus observed changes in mean PCS and MCS scores for each cohort of interest. The method is known as the "regressor variable" approach, and is commonly used in psychometric analysis of longitudinal data for scale scores with two waves of data collection (Menard, 1991; Taris, 2000; De Vaus, 2001). Expected mean PCS and MCS scores for each cohort are estimated as a function of a multivariate regression model using baseline sociodemographic characteristics, disease status, baseline values of PCS and MCS, and other variables. The differences in mean predicted and observed scores are evaluated for statistical significance using a one sample *t* test.

All baseline respondents who die prior to re-survey are retained for analysis of change in health status, and a PCS and MCS value is imputed for each decedent at follow-up. To do this, we use, as noted, the concept in economic evaluations of health known as "utilities". The HOS survey does not directly provide utility assessments. Instead, we relied on questions from the HOS to estimate utilities using the Health and Activity Limitation Index (HALex, Erickson, 1998). The HALex, developed from the 1990 National Health Interview Survey, provides utility scores for combinations of self-reported activity limitations and perceived health. (More detail is presented in Chapter 3 of this report.)

The second method is that being used by the HAL to evaluate performance of M+C health plans (Sinclair & Gandek, 2001; Rogers *et al*., 2001). As noted, it is related

to the approach developed for the MOS to compare outcomes between managed care and fee-for-service (Ware *et al*., 1996). The HAL approach excludes decedents from the MCS analysis but retains them for the PCS analysis. The two primary outcomes for comparison are rates of beneficiaries (1) alive and PCS same or better and (2) MCS same or better, since they were interested in developing a measure that indicates whether a health plan was maintaining or improving the health of its members.

To obtain these two sets of plan-level outcomes in the HAL method, four stages of data analysis are necessary. First, beneficiaries are classified as to whether their actual PCS and MCS scores are better, the same, or worse over the two year period. Beneficiaries who died during the follow-up period are assigned to a dead category for PCS and excluded from the MCS analysis. Beneficiaries alive at follow-up are considered to be in better or worse physical health if their PCS score changed by more than (+/-)5.66 points, and in better or worse mental health if their MCS score changed by more than (+/-)6.72 points.

Second, each beneficiary is assigned an expected change in PCS and MCS status using a multivariate logistic regression model, which adjusts for casemix differences among beneficiaries. The expected outcome for PCS involves estimating a two-stage model of the probability of being alive and the probability of the PCS being the same or better conditional on being alive. For MCS, a single stage probability model of being the same or better is estimated.

For PCS, the third step involves calculating an average expected death rate and the average expected PCS same or better rate for each health plan. These two expected

rates are multiplied to yield a combined estimate of the expected rate of beneficiaries being alive and PCS better or same. For MCS, the third step involves calculating an average expected MCS same or better rate at the health plan level.

Lastly, differences in expected versus actual rates are computed and statistical differences across all plans are assessed using an F test. Individual plans' differences are evaluated for statistical significance using a *t* test.

1.3 Organization of Report

Chapter 2 provides an overview of the Health Outcomes Survey used in this study of the follow-up health status of FFS beneficiaries. We provide a brief description of the HOS survey instrument, assignment of PCS and MCS scores, baseline and follow-up survey samples, and survey operations. We also provide an analysis of response rates, including the extent of missing data, internal consistency in scoring the PCS and MCS measures, the degree of retention of baseline respondents at follow-up by the four physician group practices, and differences in mean PCS and MCS scores between baseline and follow-up.

Chapter 3 describes the novel approach we developed for this study for imputing follow-up health status scores for respondents who died between baseline and follow-up. We provide descriptive analyses of changes in mean PCS and MCS scores at follow-up as a result of the imputation of PCS and MCS scores for decedents. Chapter 4 presents our multivariate statistical analysis, including the development and specification of our model for predicting expected mean follow-up health status scores, and comparisons of

expected versus actual mean follow-up health status scores across the PGP and SGA cohorts in our sample. Chapter 5 presents a comparison of expected versus actual follow-up mean health status scores across the PGP and SGA cohorts in our sample using the HAL method. Chapter 6 presents our conclusions and the policy implications of our findings.

2 Description of the Baseline and Follow-up Medicare Fee-For-Service Health Outcomes Survey

The Medicare Fee-for-Service Health Outcomes Survey (HOS) was used in our study to obtain self-reported estimates of health status from a sample of 10,000 Medicare Fee-for-Service (FFS) beneficiaries in 1998. The HOS was fielded again in 2000 to obtain follow-up estimates of health status for those beneficiaries that responded at baseline. This chapter provides a brief description of the HOS survey instrument, assignment of physical component summary (PCS) and mental component summary (MCS) scores, baseline and follow-up survey sample criteria, and survey operations. We also provide a descriptive comparison between baseline and follow-up respondents to determine whether there are systematic differences in respondents to the baseline HOS versus the follow-up HOS, or whether there are systematic differences in the completeness of survey responses as the scoring of the PCS and MCS measures are highly dependent on item response. We explore differences in mean PCS and MCS scores between baseline and follow-up and directly examine the effect of setting the PCS score to zero at follow-up for decedents. Lastly, we explore the degree of retention of baseline respondents at follow-up by the four physician group practices as a measure of the face validity of follow-up scores as a measure of performance for those providers.

2.1 Background on the Medicare Fee-for-Service Health Outcomes Survey

2.1.1 Questionnaire

The core of the Medicare Fee-for-Service Health Outcomes Survey (HOS) consists of 36 questions (SF-36), which ask the respondent to rate general health, ability to perform certain physical tasks, level of pain, and social and emotional states. The SF-36 Health Survey was developed as part of the Medical Outcomes Study (Ware *et al*., 1994). It was created to fill a need for a self-administered survey that generates an overall assessment of the respondent's mental and physical health (Ware *et al*., 1993). The beneficiary is assigned health 'scores', or levels, on the basis of his/her responses to the SF-36 questions. The HOS includes additional items on the respondent's health, demographic characteristics, and presence of any of 11 chronic conditions. The latter two types of questions may be used for case-mix control. For the FFS HOS, questions were also included regarding beneficiaries' usual source of care.

2.1.2 HOS Component Scores

The SF-36 includes eight scales measuring different aspects of physical and mental health status. They can be summarized into Physical Component Summary (PCS) and Mental Component Summary (MCS) scales (Ware *et al*., 1994). The PCS and MCS are the measures of health status outcomes used in this report. The physical health scales are Physical Functioning, Role-Physical, Bodily Pain, and General Health. The mental

health scales are Vitality, Social Functioning, Role-Emotional, and Mental Health. While these scales are classified into separate physical and mental health categories, all eight scales are used to calculate both the PCS and MCS; the four mental health scales are given less weight in the PCS score, and greater weight in the MCS scores, and vice-versa. As a result, whenever one component summary is calculated the other can be as well; so the number of respondents will always be the same for the PCS and MCS outcome measures.

All 36 SF-36 questions do not need to be answered to calculate the MCS and PCS scores for a respondent. If a respondent answers at least half of the questions in each of the eight component scales, a score can be calculated using the average scores for the completed items to replace the missing items. This means that a respondent may skip individual questions in the survey, but must answer half or more of the questions in each of the scale categories to receive a score. We use this imputation methodology in our scoring of PCS and MCS.

The PCS and MCS may also be computed from a 12-question subset of the SF-36, called the SF-12. The SF-12 includes selected questions representing each of the eight scales included in the SF-36. The SF-12 was designed as a shorter survey, which could produce comparable measures for the PCS and MCS, but with a higher response rate due to the reduced reporting burden it places on respondents. For this reason, the SF-12 was used to define the minimum required survey response for our FFS HOS survey. It allowed us to retain in the study those respondents who did not answer all of the

questions in the SF-36, but were willing to answer the SF-12. In this analysis, we sought to keep the scoring method the same between baseline and follow-up.

When constructing either the SF-12 or SF-36 estimates of PCS and MCS, the two component scores are normalized such that the mean is 50 with a standard deviation of 10 points in the general U.S. population (Ware *et al*., 1993). All published literature to date reflect normalization to the 1990 U.S. general population. We use 1990 population norms in this project to allow for comparison of our results with the published literature. It should be noted, however, that the analysis of follow-up health status in the Cohort 1 M+C HOS normalized their PCS and MCS component scores to a new 1998 standard population. Further, they employed a new imputation methodology for missing data. In the managed care analysis, a missing data estimation (MDE) utility is employed, which allows for calculation of the PCS and MCS if at least one item is answered within each of the eight scales (Rogers *et al*., 2001). Thus, our results are not directly comparable to the managed care results.

2.1.3 Survey Sample

The baseline HOS was administered to 10,000 Medicare Fee-for-Service 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 five SGAs and four PGPs were chosen to provide a variety of contrasts between different geographic locations and types of physician group practices. Thus, the

combined sample is a convenience sample of the Medicare FFS population. The sample of 10,000 Medicare beneficiaries was drawn from the 100 percent Medicare Enrollment Data Base (EDB), which contains Medicare enrollment and entitlement information for all beneficiaries ever enrolled in the Medicare program.

The initial sample was drawn by selecting only beneficiaries with randomly selected numbers using the four terminal digits of their social security number. Medicare beneficiaries were eligible for the initial selection if they had been continuously enrolled in Medicare fee-for-service for all of calendar year 1997 and had complete mailing addresses in the EDB. Beneficiaries were dropped from the initial sample if they were eligible for Medicare through the End-Stage Renal Disease program, were Railroad Board Retirees, or were members of a Medicare+Choice health plan. Further, inclusion in the survey as a part of the Physician Group Practice (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 areas oversampled were located in the states of Arizona, Georgia, Pennsylvania, Wisconsin and Washington. Residency in these states at the time of sampling was a requirement.

A beneficiary was mailed a follow-up survey instrument, if he/she was a respondent at baseline and was alive at the time of re-survey. A respondent is defined as a beneficiary with a calculated PCS or MCS score at baseline. Death prior to re-survey was defined in three ways: a date of death in the CMS EDB prior to fielding the follow-

up survey, notification via telephone or mail in response to the follow-up survey mailing that the beneficiary had died prior to completing the survey instrument, or determination through the CMS EDB that the beneficiary had died during the survey period and did not respond to any survey attempts. All decedents are retained in our analytic file for analysis in this report.

2.1.4 Survey Operations

The baseline HOS was administered from May 1998 through January 1999. The follow-up survey was administered from May 2000 through December 2000. The mode of administration in both waves was mail with telephone follow-up. The New England Research Institutes was the survey vendor for both the baseline and follow-up survey administration. Medicare beneficiaries who did not complete a mail survey after three mailing attempts were referred for telephone follow-up and up to 10 phone calls were placed in an effort to contact the beneficiary. Federal Express mailing was used as the follow-up technique for beneficiaries for whom we did not have a valid telephone number. Prior to re-survey, mailing addresses were evaluated for change using Medicare's EDB. Follow-ups 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 at both time periods.

More detailed information on the sampling of FFS beneficiaries for this project, on the methods for HOS survey administration, and on the characteristics of the baseline (1998) survey data are all available in previous reports for this project (McCall *et al*., 1998; 2000).

2.2 Response Rates to the Baseline and Follow-up Medicare Fee-for-Service Health Outcomes Survey by Beneficiary Enrollment, Sociodemographic, and Health Status Characteristics

The response rate at baseline was 68.5 percent. A total of 320 beneficiaries died after being selected for the survey but prior to completing a baseline instrument and were considered ineligible at baseline. Another 3,046 beneficiaries refused to complete the survey instrument, or were considered nonrespondents for reasons such as unable to be located, institutionalized and severely impaired with either physical or cognitive illness, non-English and non-Spanish speakers, or beneficiaries for whom access was denied by gatekeepers (typically in nursing homes), or did not complete enough questions to calculate a MCS and PCS score.

Response rate at follow-up was 91.7 percent, using the MCS definition of eligible population which excludes the deceased from the eligible population for follow-up, and 92.5 percent when the deceased are considered as both eligible and respondents. A total of 673 beneficiaries died between completing a baseline survey instrument and the completion of the follow-up survey period. Another 496 beneficiaries were considered nonrespondents.

Tables 2-1 through 2-3 display survey response rates for the baseline and follow-up surveys across each sampling unit and by baseline sociodemographic and health status characteristics. Thus, one is able to observe baseline and follow-up rates as a function of baseline characteristics. Differences in rates and proportions across strata are evaluated for statistical significance by using the chi-square test of differences for categorical data at the 0.05 significance level. When statistically significant differences are found, we then conduct a pairwise analysis of differences in rates or proportions using the *z* test statistic with a pooled sample variance. A 0.01 level of statistical significance is used as we are making multiple comparisons.

Table 2-1, column 1, displays the number of sampled Medicare FFS beneficiaries, in total, and within the ten strata: one national random sample; five small geographic areas; and four large multi-specialty group practices. The second column contains the number of sampled beneficiaries that were alive at the time of sampling but died prior to completing a baseline survey questionnaire. Subtracting the number of deaths from the sample yields the total number of baseline eligibles within each stratum (column 3). The total number of baseline respondents is displayed in column 4, and the resultant baseline response rate is displayed in column 5.

Table 2-1

Survey Response Rates to Baseline and Follow-up Medicare Fee-For-Service Health Outcomes Survey, by Sample Cohort

	Baseline					Follow-Up				
Cohort	Sampled Beneficiaries (1)	Deaths[1] (2)	Eligibles (3)=(1)-(2)	Baseline Respondents (4)	Response Rate (5)=(4)/(3)	Follow-Up Respondents (6)	Deaths[2] (7)	Death Rate (8)=(7)/(4)	PCS Response Rate[3] (9)=[(6)+(7)]/(4)	MCS Response Rate[3] (10)=(6)/[(4)-(7)]
All	10,000	320	9,680	6,634	68.5%	5,465	673	10.1%	92.5%	91.7%
National	1,000	42	958	617	64.4	484	60	9.7	88.2	86.9
PA SGA[4]	1,000	35	965	601	62.3	492	67	11.1	93.0	92.1
GA SGA	1,000	39	961	615	64.0	482	70	11.4	89.8	88.4
WI SGA	1,000	33	967	767	79.3*#	667	60	7.8	94.8	94.3
AZ SGA	1,000	23	977	604	61.8	481	57	9.4	89.1	87.9
WA SGA	1,000	39	961	597	62.1	497	56	9.4	92.6	91.9
PGP[5] A	1,000	32	968	704	72.7*#	574	88	12.5	94.0	93.2
PGP B	1,000	28	972	768	79.0*#	651	85	11.1	95.8	95.3
PGP C	1,000	19	981	703	71.7*#	588	66	9.4	93.0	92.3
PGP D	1,000	30	970	658	67.8#	549	64	9.7	93.2	92.4

[1]Deaths in the Baseline include all beneficiaries who were alive at sampling, but died before returning a baseline survey (based on NERI Survey Dispositions and EDB Death variable)

[2]Deaths in the Follow-up include all beneficiaries who responded to a baseline survey, but died before returning a follow-up survey (based on NERI Survey Dispositions and EDB Death variable)

[3]The HAL methodology for analyzing two-year change in PCS and MCS scores allows respondents who died between baseline and follow-up to be dropped from the analysis of MCS results. Hence the response rate is calculated two ways for comparison. The MCS Response Rate does not include deaths.

[4]SGA refers to a small geographic area selected for sampling within the given sta

[5]PGP refers to a primary group practice whose members were selected for sampling. One PGP was selected for each state sampled, excluding Geo

*Significantly different from National response rate for the Baseline Sample (pairwise z score, 1% level).

#Significantly different from National response rate for the Baseline Eligibles (pairwise z score, 1% lev

OUTPUT: n03nora

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Table 2-2

Survey Response Rates to Baseline and Follow-up Medicare Fee-For-Service Health Outcomes Survey, by Baseline Demographic and Eligibility Characteristics from CMS's Enrollment Database (EDB)

	Baseline					Follow-up				
	Sampled Beneficiaries (1)	Deaths[1] (2)	Eligibles (3)=(1)-(2)	Baseline Respondents (4)	Response Rate (5)=(4)/(3)	Follow-up Respondents (6)	Deaths[2] (7)	Death Rate (8)=(7)/(4)	PCS Response Rate[3] (9)=[(6)+(7)]/(4)	MCS Response Rate[3] (10)=(6)/[(4)-(7)]
All	10,000	320	9,680	6,634	68.5%	5,465	673	10.1%	92.5%	91.7%
Sex										
Male	4,070	144	3,926	2,738	69.7	2,194	329	12.0	92.1	91.1
Female	5,930	176	5,754	3,896	67.7	3,271	344	8.8	92.8	92.1
Age										
Under 65	948	14	934	540*	57.8	443	27	5.0	87.0	86.4
65-74	3,706	61	3,645	2,665	73.1	2,357	148	5.6	94.0	93.6
75-84	3,865	121	3,744	2615*	69.8	2,136	283	10.8	92.5	91.6
85 and Older	1,481	124	1,357	814*	60.0	529	215	26.4	91.4	88.3
Race										
White	9,264	295	8,969	6,225	69.4	5,133	646	10.4	92.8	92.0
Black	490	21	469	261	55.7	204	18	6.9	85.1	84.0
Asian	56	0	56	33	58.9	29	1	3.0	90.9	90.6
Hispanic	62	1	61	36	59.0	31	2	5.6	91.7	91.2
North American Native	16	0	16	10	62.5	9	0	0.0	90.0	90.0
Other	91	2	89	55	61.8	49	3	5.5	94.5	94.2
Unknown	21	1	20	14	70.0	10	3	21.4	92.9	90.9
Original Reason for Entitlement										
Aged	8,413	280	8,133	5,703	70.1	4,717	596	10.5	93.2	92.4
Disabled	1,572	39	1,533	923*	60.2	742	76	8.2	88.6	87.6
ESRD	4	0	4	3	75.0	3	0	0.0	100.0	100.0
ESRD and Disabled	11	1	10	5	50.0	3	1	20.0	80.0	75.0
Medicaid Status										
No Medicaid	8,828	259	8,569	5,981	69.8	4,980	585	9.8	93.0	92.3
Medicaid	1,172	61	1,111	653*	58.8	485	88	13.5	87.7	85.8

[1]Deaths in the Baseline include all beneficiaries who were alive at sampling, but died before returning a baseline survey (based on NERI Survey Dispositions and EDB Death variable)

[2]Deaths in the Follow-up include all beneficiaries who responded to a baseline survey, but died before returning a follow-up survey (based on NERI Survey Dispositions and EDB Death variable)

[3]The HAL methodology for analyzing two-year change in PCS and MCS scores allows respondents who died between baseline and follow-up to be dropped from the analysis of MCS results. Hence the response rate is calculated two ways for comparison.

*Significantly different from reference category response rate for the Baseline Eligibles (pairwise z score, 1% level). Reference categories are Age 65-74, Aged, and No Medicaid.

OUTPUT: n05

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Table 2-3

Survey Response Rates to Baseline and Follow-up Medicare Fee-For-Service Health Outcomes Survey, by Baseline Sociodemographic Characteristics

	Baseline	Follow-up				
	Respondents (1)	Respondents (2)	Deaths[1] (3)	Death Rate (4) = (3)/(1)	PCS Response Rate[2] (5) =[(2)+(3)]/(1)	MCS Response Rate[2] (6)=(2)/[(1)-(3)]
All	6634	5465	673	10.1	92.5	91.7
Household Income						
Less than $10,000	801	620	106	13.2	90.6	89.2
$10,000-$19,999	1,099	914	122	11.1	94.3	93.6
$20,000-$49,999	1,867	1,629	149	8.0	95.2	94.8
$50,000 or more	619	549	46	7.4	96.1	95.8
Missing/No Response	2,248	1,753	250	11.1	89.1	87.7
Education						
Not a HS Graduate	1,491	1,191	192	12.9	92.8	91.7
High School Graduate or GED	1,858	1,582	165	8.9	94.0	93.4
Some College or 2 year degree	1,100	947	91	8.3	94.4	93.9
4 year college graduate	449	396	35	7.8	96.0	95.7
More than a 4 year college degree	516	447	43	8.3	95.0	94.5
Missing/No Response	1,220	902	147	12.0	86.0	84.1
Marital Status						
Not Married	2,305	1,859	263	11.4	92.1	91.0
Married	3,153	2,734	267	8.5	95.2	94.7
Missing/No Response	1,176	872	143	12.2	86.3	84.4
Home Owner Status						
Owned by Beneficiary or Family Member	4,265	3,678	364	8.5	94.8	94.3
Not Owned by Beneficiary or Family Member	1,016	783	141	13.9	90.9	89.5
Missing/No Response	1,353	1,004	168	12.4	86.6	84.7
Live in a Retirement Community						
Yes	908	728	119	13.1	93.3	92.3
No	4,367	3,727	379	8.7	94.0	93.5
Missing/No Response	1,359	1,010	175	12.9	87.2	85.3

Table 2-3 (continued)

Survey Response Rates to Baseline and Follow-up Medicare Fee-For-Service Health Outcomes Survey, by Baseline Sociodemographic Characteristics

	Baseline	Follow-up				
	Respondents (1)	Respondents (2)	Deaths[1] (3)	Death Rate (4) = (3)/(1)	PCS Response Rate[2] (5) =[(2)+(3)]/(1)	MCS Response Rate[2] (6)=(2)/[(1)-(3)]
Chronic Conditions						
Hypertension or high blood pressure	2,856	2,393	281	9.8%	93.6	92.9%
Angina pectoris or coronary artery disease	992	798	144	14.5	95.0	94.1
Congestive heart failure	458	322	106	23.1	93.4	91.5
Acute myocardial infarction or heart attack	691	530	116	16.8	93.5	92.2
Other heart conditions	1,401	1,117	185	13.2	92.9	91.9
Stroke	555	391	115	20.7	91.2	88.9
Emphysema, asthma, or COPD	744	573	123	16.5	93.5	92.3
Crohn's disease, ulcerative colitis, or inflammatory bowel disease	386	331	35	9.1	94.8	94.3
Arthritis of the hip or knee	2,231	1,879	232	10.4	94.6	94.0
Arthritis of the hand or wrist	1,968	1,649	194	9.9	93.6	93.0
Sciatica	1,419	1,212	122	8.6	94.0	93.4
Diabetes, high blood sugar, or sugar in the urine	948	759	120	12.7	92.7	91.7
Any cancer (other than skin cancer)	971	783	137	14.1	94.7	93.9
Number of chronic conditions reported[3]						
0	1,764	1,388	174	9.9%	88.5%	87.3%
1	958	825	76	7.9%	94.1%	93.5%
2	1,104	958	85	7.7%	94.5%	94.0%
3	992	844	84	8.5%	93.5%	93.0%
4	721	607	77	10.7%	94.9%	94.3%
5	514	414	69	13.4%	94.0%	93.0%
6	277	205	52	18.8%	92.8%	91.1%
7	162	120	26	16.0%	90.1%	88.2%
8+	142	104	30	21.1%	94.4%	92.9%
Number of ADLs the respondent has difficulty with or is unable to do[3]						
0	3,939	3,349	272	6.9%	91.9%	91.3%
1	837	718	81	9.7%	95.5%	95.0%
2	710	586	82	11.5%	94.1%	93.3%
3	349	272	48	13.8%	91.7%	90.4%
4	283	222	45	15.9%	94.3%	93.3%
5	255	169	62	24.3%	90.6%	87.6%
6	261	149	83	31.8%	88.9%	83.7%

Table 2-3 (continued)

Survey Response Rates to Baseline and Follow-up Medicare Fee-For-Service Health Outcomes Survey, by Baseline Sociodemographic Characteristics

	Baseline	Follow-up				
	Respondents (1)	Respondents (2)	Deaths[1] (3)	Death Rate (4) = (3)/(1)	PCS Response Rate[2] (5) =[(2)+(3)]/(1)	MCS Response Rate[2] (6)=(2)/[(1)-(3)]
Health in general						
Excellent	312	277	7	2.2%	91.0	90.8%
Very Good	1,378	1,243	59	4.3	94.5	94.2
Good	2,556	2,191	189	7.4	93.1	92.6
Fair	1,787	1,385	246	13.8	91.3	89.9
Poor	590	360	170	28.8	89.8	85.7
Missing/No Response	11	9	2	18.2	100.0	100.0
Baseline MCS Score Range						
0-30	392	275	76	19.4	89.5	87.0
31-40	895	677	136	15.2	90.8	89.2
41-50	1,280	1,032	155	12.1	92.7	91.7
51-60	2,956	2,551	213	7.2	93.5	93.0
61+	1,111	930	93	8.4	92.1	91.4
Baseline PCS Score Range						
0-20	545	366	137	25.1	92.3	89.7
21-30	1,509	1,136	250	16.6	91.8	90.2
31-40	1,503	1,241	145	9.6	92.2	91.4
41-50	1,615	1,413	83	5.1	92.6	92.2
51+	1,462	1,309	58	4.0	93.5	93.2

[1]Deaths in the Follow-up include all beneficiaries who responded to a baseline survey, but died before returning a follow-up survey (based on NERI Survey Dispositions and EDB Death variable)

[2]The HAL methodology for analyzing change in PCS and MCS scores allows respondents who died between baseline and follow-up to be dropped from analysis of MCS results. Hence the response rate is calculated two ways for comparison. The MCS Response Rate does not include deaths.

[3]A zero in either of these categories could also indicate that the recipient did not respond to any of the questions included in the measure.

Output: n08 and a08_resp, joinx02l, joinx03, joinx01a

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Across all strata, the response rate is 68.5 percent; ranging from a low of 61.8 percent in the Arizona small geographic area to a high of 79.3 percent in the Wisconsin small geographic area. There were statistically significant differences in response rates across the 10 strata. In comparison with the national sample, the Wisconsin small geographic area's response rate was significantly higher. Further, all four physician group practices also experienced statistically higher response rates than the national sample.

The response rate for the follow-up survey is a function of the number of respondents to the baseline and follow-up surveys (columns 4 and 6) and the number of beneficiaries who died between completing the baseline and follow-up survey (column 7). Under the HAL method of assessing health status at follow-up, response rate calculations differ between the PCS (column 9) and MCS (column 10).

The calculation of the PCS and MCS response rates are displayed in columns 9 and 10. The overall response rate was in excess of 90 percent for both the PCS and MCS. The PCS response rates ranged from a low of 88.2 percent to a high of 95.8 percent and the MCS scores ranged from a low of 86.9 percent to a high of 95.3 percent. There are no statistically significant differences in the follow-up response rates using either the PCS or MCS definition of eligible.

Table 2-2 presents a similar set of data by eligibility and enrollment information contained in CMS's Enrollment Data Base (EDB) at the time of sampling for the baseline survey. A few statistical differences in baseline response rates are worthy of comment.

Medicare beneficiaries who were under 65 or 85 and over were less likely to respond than beneficiaries age 65 to 84. Blacks were less likely than whites to respond. Beneficiaries originally enrolled in Medicare due to disability and Medicaid enrollees were less likely to respond to those enrolled due to age and not enrolled in Medicaid, respectively. At follow-up, although there continues to be some observed variation in response rates across enrollment and eligibility stratifications, no statistically significant differences were observed. This is likely due to small numbers.

Sociodemographic and health status characteristics available from the Health Outcomes Survey provide additional opportunities to evaluate differential response rates at follow-up as a function of baseline beneficiary characteristics not available from CMS data. Table 2-3 displays response rates at follow-up based on baseline characteristics such as household income, education, marital status, home ownership status, residence in a retirement community, and a variety of comorbidity variables. This table is similar to Table 2-2 with the exception that no baseline response rates are provided because these survey-based measures do not exist for sampled beneficiaries who did not respond to the baseline survey. Thus, only follow-up response rates are reported and evaluated for statistical differences.

Using the PCS definition of response, which includes the deceased between baseline survey completion and fielding of the follow-up survey, there are no statistically significant differences in response rates across any of the sociodemographic or health

status stratifying variables. The lowest response rates at follow-up tended to be for those beneficiaries who had missing data at baseline.

A similar pattern holds using the HAL definition of response with one noted exception. The chi-square statistic for number of chronic conditions reported suggests statistically significant variation in response rates across the nine response categories. Those reporting no chronic conditions had the lowest response rate at follow-up, 87.3 percent; however, this category also includes respondents who may have skipped this series of questions. Other variables, *e.g.*, education, income, etc., that have a discrete category for missing/no response show that the follow-up response rates tend to be lowest for respondents who did not complete the baseline question. In comparison with the zero chronic conditions category, respondents who reported one through four chronic conditions present at baseline had higher response rates. There were no significant differences in response rates at follow-up between those who reported five or more chronic conditions at baseline versus no chronic conditions at baseline (or missing).

2.3 Changes in Sociodemographic and Health Status Characteristics of Respondents at Baseline and Follow-up

An alternative way of evaluating potential differential response rates based upon beneficiary characteristics is displayed in Tables 2-4 and 2-5. In these two tables, the frequency of respondents at baseline and follow-up are displayed across demographic and

Table 2-4

Demographic and Eligibility Characteristics of Baseline and Follow-up Respondents to the Medicare Fee-for-Service Health Outcomes Survey

	Frequencies		Percentage Distribution	
	Baseline Respondents	Follow-Up Respondents[1]	Baseline Respondents	Follow-Up Respondents[1]
All	6,634	5,465	100%	100%
Sex				
Male	2,738	2,194	41.3%	40.1%
Female	3,896	3,271	58.7	59.9
Age*				
Under 65	540	424	8.1	7.8
65-74	2,665	2,064	40.2	37.8
75-84	2,615	2,318	39.4	42.4
85 and Older	814	659	12.3	12.1
Race				
White	6,225	5,133	93.8	93.9
Black	261	204	3.9	3.7
Asian	33	29	0.5	0.5
Hispanic	36	31	0.5	0.6
North American Native	10	9	0.2	0.2
Other	55	49	0.8	0.9
Unknown	14	10	0.2	0.2
Original Reason for Entitlement[2]				
Aged	5,703	4,717	86.0	86.3
Disabled	923	742	13.9	13.6
ESRD	3	3	0.0	0.1
ESRD and Disabled	5	3	0.1	0.1
Medicaid Status				
No Medicaid	5,981	4,980	90.2	91.1
Medicaid	653	485	9.8	8.9

[1]Includes only respondents to the follow-up Survey. Decendents are excluded.
Hence the response rate is calculated two ways for comparison. The Alternate MCS Response Rate does not include deaths.
[2]Categories are defined by baseline status only.

*Significantly different distribution between baseline eligibles and living follow-up respondents (chi-square, 5% level).

OUTPUT: n07, joinx02j

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Table 2-5

Sociodemographic Characteristics of Baseline and Follow-up Respondents to the Medicare Fee-For-Service Health Outcome Survey

	Frequencies		Percentage Distribution	
	Baseline Respondents	Follow-Up Respondents[1]	Baseline Respondents	Follow-Up Respondents[1]
All	6,634	5,465	100%	100%
Household Income*				
Less than $10,000	801	657	12.1	12.0
$10,000-$19,999	1,099	927	16.6	17.0
$20,000-$49,999	1,867	1,622	28.1	29.7
$50,000 or more	619	595	9.3	10.9
Missing/No Response	2,248	1,664	33.9	30.4
Marital Status*				
Not Married	2,305	2,184	34.7	40.0
Married	3,153	2,713	47.5	49.6
Missing/No Response	1,176	568	17.7	10.4
Homeowner Status				
Owned by Beneficiary or Family Member	4,265	3,651	64.3	66.8
Not owned by Beneficiary or Family Member	1,016	657	15.3	12.0
Missing/No Response	1,353	1,157	20.4	21.2
Live in a Retirement Community*				
Yes	908	806	13.7	14.7
No	4,367	3,897	65.8	71.3
Missing/No Response	1,359	762	20.5	13.9
Chronic Conditions				
Hypertension or high blood pressure	2,856	2,839	43.1	51.9
Angina pectoris or coronary artery disease	992	980	15.0	17.9
Congestive heart failure	458	489	6.9	8.9
Acute myocardial infarction or heart attack	691	666	10.4	12.2
Other heart conditions	1,401	1,354	21.1	24.8
Stroke	555	547	8.4	10.0
Emphysema, asthma, or COPD	744	749	11.2	13.7
Crohn's disease, ulcerative colitis, or inflammatory bowel disease	386	371	5.8	6.8
Arthritis of the hip or knee	2,231	2,249	33.6	41.2
Arthritis of the hand or wrist	1,968	1,927	29.7	35.3
Sciatica	1,419	1,372	21.4	25.1
Diabetes, high blood sugar, or sugar in the urine	948	932	14.3	17.1
Any cancer (other than skin cancer)	971	956	14.6	17.5
Number of ADLS with difficulty[2*]				
0	3939	2809	59.4	51.4
1	837	793	12.6	14.5
2	710	711	10.7	13.0
3	349	364	5.3	6.7
4	283	274	4.3	5.0
5	255	289	3.8	5.3
6	261	225	3.9	4.1

Table 2-5 (continued)

Sociodemographic Characteristics of Baseline and Follow-up Respondents to the Medicare Fee-For-Service Health Outcome Survey

	Frequencies		Percentage Distribution	
	Baseline Respondents	Follow-Up Respondents[1]	Baseline Respondents	Follow-Up Respondents[1]
Number of chronic conditions reported[2*]				
0	1,764	812	26.6	14.9
1	958	849	14.4	15.5
2	1,104	1,017	16.6	18.6
3	992	952	15.0	17.4
4	721	733	10.9	13.4
5	514	474	7.7	8.7
6	277	285	4.2	5.2
7	162	171	2.4	3.1
8+	142	172	2.1	3.1
Health in general*				
Excellent	312	197	4.7	3.6
Very Good	1,378	1,080	20.8	19.8
Good	2,556	2,202	38.5	40.3
Fair	1,787	1,542	26.9	28.2
Poor	590	431	8.9	7.9
Missing/No Response	11	13	0.2	0.2
MCS Score Range				
0-30	367	270	5.5	4.9
31-40	867	705	13.1	12.9
41-50	1,235	1,067	18.6	19.5
51-60	2,896	2,319	43.7	42.4
61+	1,269	1,104	19.1	20.2
PCS Score Range*				
0-20	486	417	7.3	7.6
21-30	1,476	1,352	22.2	24.7
31-40	1,520	1,344	22.9	24.6
41-50	1,591	1,254	24.0	22.9
51+	1,561	1,098	23.5	20.1

[1]Includes only follow-up respondents.
[2]A zero in this category could also indicate that the recipient did not respond to any of the questions included in the measure.

*Significantly different distribution between baseline eligibles and living follow-up respondents (chi-square, 5% level).

Output: joinx01a, joinx02l

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

eligibility characteristics assessed at each survey time, allowing one to evaluate whether a shift in the distribution of respondents across characteristics occurred between baseline and follow-up. Only the HAL definition of a respondent at follow-up is used in this series of analyses, thus baseline respondents who died prior to the follow-up survey are included in the baseline frequency counts but are excluded from the follow-up counts. Chi-square tests of differences in proportions were conducted across the strata to identify statistically significant changes in the distributions between baseline and follow-up. Not surprisingly, there is remarkable similarity in the proportion of beneficiaries responding at both baseline and follow-up across the domains of sex, race, original reason for Medicare entitlement and dual Medicare and Medicaid enrollment.

In contrast, there are statistically significant differences in the proportions of respondents based on age when using the HAL change scores definition of response, which reflects the aging of the cohort. The proportion of beneficiaries age 65 to 74 declines by over two percentage points while the proportion of beneficiaries age 75 to 84 increases by three percentage points. There are no significant changes in the proportion under 65 and 85 years of age and older.

Using survey-based measures of sociodemographic characteristics as well as self-reported health status measures, one observes modest, but statistically significant, changes in the distribution of Medicare FFS beneficiaries between baseline and follow-up for all characteristics displayed in Table 2-5, with the exception of MCS scores. Between the two time periods, there is an increase in the proportion of respondents who

report a household income in excess of $50,000, are not married, and who live in a home owned by themselves or a family member. However, it should be noted that these statistical changes appear to be influenced by a changing proportion of cases providing an answer at follow-up relative to the proportion of missing responses at baseline. Z-tests of comparisons of proportions of missing responses at baseline versus follow-up confirm this hypothesis for income and marriage status.

The most noticeable shifts in distribution appear to be related to presence of chronic conditions and activity of daily living limitations. Between baseline and follow-up, the proportion of beneficiaries reporting no chronic conditions fell by almost 12 percentage points, from 26.6 percent to 14.9 percent. The shift in distribution is clearly toward greater number of chronic conditions. For example, the proportion of beneficiaries with four chronic conditions rose from 10.9 to 13.4 percent in the two year interval. The proportion of respondents with specific chronic conditions increased for all listed conditions, except Crohn's disease. The proportion of respondents with hypertension increased by almost 9 percentage points. The proportion of respondents with no limitations in activities of daily living decreased by a significant 8 percentage points. The proportion of beneficiaries that reported their general health as excellent declined from 4.7 to 3.6 percent, a statistically significant change. A similar shift is observed in the distribution of physical component scores; the proportion of beneficiaries with a PCS score of 51 or greater fell from 23.5 percent to 20.1 percent. At the same time, the proportion of beneficiaries whose PCS scores ranged from 21 to 30, indicating

worsening health, rose from 22.2 to 24.7 percent. This is consistent with the literature of declining physical health over time. There were no statistically significant changes in the distribution of respondents across the MCS categories.

2.4 Scoring of PCS and MCS at Baseline and Follow-up

In our study, we used two different scoring methods to obtain estimates of PCS and MCS; one based on a 36 question item scoring algorithm and a second based on a 12 question item scoring algorithm. Use of the 12 item scoring method allowed us to retain in the study a segment of the FFS population who did not complete the entire HOS survey instrument. Our analysis provides an assessment of the degree of completeness of the survey-based health status responses that are used to produce the baseline and follow-up PCS and MCS scores. A significant shift in the proportions being scored with each method between baseline and follow-up could result in a spurious measured difference in health, *e.g.*, a measured change when no actual change occurred.

Table 2-6 displays the proportion of beneficiaries for whom we were able to calculate PCS and MCS scores using their baseline scoring algorithm method. In our initial scoring approach, we focused on using the same scoring method at both baseline and follow-up. We were able to achieve considerable consistency in response patterns between baseline and follow-up. Roughly 20 percent of our sample was scored using the SF-12 at baseline and follow-up, as shown in Table 2-6. Only six baseline respondents

Table 2-6

Scoring Method for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey

	Scoring Method[1]							
	SF-12 Only		SF-12/SF-36		SF-36 Only		SF-36/SF-12	
	#	%	#	%	#	%	#	%
Site								
All	1,097	20.7	6	0.1	4,361	79.8	1	0.0
National	129	26.7	1	0.2	353	72.9	1	0.2
PA SGA	117	23.8	1	0.2	374	76.0	0	0.0
GA SGA	119	24.7	0	0.0	363	75.3	0	0.0
WI SGA	111	16.6	1	0.1	555	83.2	0	0.0
AZ SGA	88	18.3	2	0.4	391	81.3	0	0.0
WA SGA	112	22.5	0	0.0	385	77.5	0	0.0
PGP A	118	20.6	1	0.2	455	79.3	0	0.0
PGP B	109	16.7	0	0.0	542	83.3	0	0.0
PGP C	90	15.3	0	0.0	498	84.7	0	0.0
PGP D	104	18.9	0	0.0	445	81.1	0	0.0

[1]In the baseline and follow-up surveys, we counted a survey as complete if there were enough answers given to calculate a PCS and MCS score using either the SF-36 and/or the SF-12. Our initial approach was to use the same scoring method (SF-36 or SF-12) for both the baseline and the follow-up whenever possible. We started with beneficiaries with baseline SF-36 scores. If they also had a follow-up SF-36 score, then the SF-36 scoring method was used in calculating both their baseline and follow-up survey scores ("SF-36 Only"). If there were not complete SF-36 responses for both the baseline and the follow-up surveys, we tried to use only completed SF-12 responses for both ("SF-12 Only"). When only the SF-12 was completed in the first round, and only the SF-36 was completed in the second round (i.e., the SF-36 could be completed but the SF-12 could not, due to the pattern of missing data), or vice-versa, the beneficiaries had scores using one scoring method in the baseline survey, and the other scoring method in the follow-up.
("SF-12/SF-36" or "SF-36/SF-12.")

SGA= Small Geographic Area
PGP= Primary Group Practice

Output: n07_5

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

who were scored using the SF-12 at baseline were unable to be scored at follow-up using the same algorithm, but were able to be scored using the SF-36 algorithm, due to the pattern of missing data. It is important to note that some beneficiaries in the category of "SF-12 only" could have also been scored using the SF-36 algorithm at either baseline or follow-up.

For the remaining 80 percent of our sample that was scored at baseline using the SF-36, we were also able to score their follow-up surveys using the SF-36 scoring algorithm. There was only one respondent for whom the SF-12 was used at follow-up and the SF-36 had to be used to calculate a baseline score due to the pattern of missing data.

2.5 Differences in Mean Physical and Mental Health Between Baseline and Follow-up

This section of the report focuses upon differences in mean PCS and MCS scores during the two year time period between baseline and follow-up surveying. We compared differences between the two time periods in average PCS and MCS scores across our FFS beneficiaries, in total, by sample cohort, and by sociodemographic characteristics, Medicare enrollment and eligibility characteristics, and self-reported health status characteristics. In this analysis, we follow the lead of the Medical Outcomes Study, which assigned the value of zero to follow-up PCS scores for baseline respondents who died between the time of completing the baseline survey and completing a follow-up survey (Ware *et al*., 1996). Decedents are removed totally from the MCS change score analysis.

Statistical comparisons of differences in mean PCS and MCS scores are made between the two time periods using a two-tailed *t*-test for differences in means. No adjustment for multiple comparisons has been made. We highlight those differences that are statistically significant at the 0.01 level as an alternative approach for correcting for multiple comparisons. However, we are also concerned with identifying "clinically significant" differences in average scores. We consider two approaches. One approach to defining a "minimally clinically important difference" is to apply conventional statistical standards for "effect" sizes. Cohen's (1988) conventions are the most widely known and used. He defines small effects as 0.2 standard deviations, medium effects as 0.5 standard deviations, and large effects as 0.8 standard deviations. Since the MCS and PCS are normalized to have standard deviations of 10 points, these conventions translate into differences of 2, 5, and 8 points on the component scales. The SF-36 developers have themselves endorsed this approach (Ware and Kosinski, 2001).

Another approach is to relate score differences to external factors that are considered to be important or interpretable. In other contexts, changes in component scores due to job loss or divorce could provide an interpretable metric. In the analysis of health status, the impact of chronic diseases on health scores provides a natural benchmark. Ware *et al.*, (1994) show that the effect of comorbidities (asthma, COPD, angina, etc.) on the PCS range from 2 to 6 points. Comorbidities other than clinical depression tend to have much smaller impacts on the MCS (Ware *et al.*, 1994). Ware and Kosinsky (2001) provide additional examples of the clinical correlates of the different

effect sizes. For example, an improvement of 2 points on the PCS or MCS has been correlated to pre/post drug treatment for migraine headaches. Since the threshold of two points is justified by both approaches, we consider PCS or MCS differences of 2 points or more between any two groups or between two time periods for the same group to be minimal "clinically important" differences.

Tables 2-7 and 2-8 display mean PCS and MCS scores by sampling unit and by demographic and eligibility characteristics and self-reported health status for the baseline and follow-up surveys. In Table 2-7, the mean PCS scores at follow-up are displayed for all respondents, including decedents whose PCS scores are set to zero, and for survivors only. This allows us to directly examine the impact of setting PCS scores to zero in the follow-up period when a respondent to the baseline survey dies prior to completing a follow-up survey. In Table 2-8, MCS scores at follow-up are displayed for respondents only as decedents are removed from the eligible population.

Across all strata of respondents, the mean PCS score baseline is 38.51, and declines by over 5 points during the two year follow-up period. This is both a statistically significant and clinically significant decline. Restricting our analysis to survivors only, we observe a modestly higher average baseline PCS score of 39.48 but a considerably smaller, although statistically and clinically significant, decline in the average PCS score of roughly 2 points. Although beneficiaries who were alive at the time of follow-up are

Table 2-7

Difference in Mean PCS Score Between Baseline and Follow-up by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (followup PCS of decedents = 0)

	Decedents & Survivors							Survivors					
	PCS Follow-Up Respondents	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Change Due to PCS=0 for Decedents[1]	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
ALL	6,138	38.51	33.23	-5.27	0.16	***	59.2	89%	39.48	37.33	-2.15	0.14	***
Site													
National	544	38.34	32.65	-5.68	0.58	***	59.1	89.0%	39.02	36.70	-2.32	0.42	***
PA SGA[3]	559	39.03	33.20	-5.84	0.54	***	57.3	88.0	40.21	37.72	-2.49	0.40	***
GA SGA	552	37.51	31.87	-5.64	0.57	***	60.3	87.3	38.74	36.50	-2.24	0.44	***
WI SGA	727	37.78	32.98	-4.80	0.45	***	49.3	91.7	38.38	35.94	-2.43	0.35	***
AZ SGA	538	39.76	34.59	-5.17	0.56	***	59.2	89.4	40.80	38.69	-2.11	0.40	***
WA SGA	553	39.56	34.54	-5.03	0.54	***	56.3	89.9	40.62	38.43	-2.20	0.40	***
PGP[4] A	662	37.55	32.22	-5.33	0.52	***	71.5	86.7	38.68	37.16	-1.52	0.36	***
PGP B	736	37.85	32.26	-5.59	0.46	***	57.7	88.5	38.84	36.47	-2.37	0.34	***
PGP C	654	39.06	34.02	-5.04	0.49	***	59.7	89.9	39.87	37.84	-2.03	0.35	***
PGP D	613	39.11	34.38	-4.73	0.51	***	62.1	89.6	40.18	38.39	-1.79	0.38	***
Sex													
Male	2,523	39.44	33.54	-5.91	0.27	***	64.5	87.0	40.66	38.57	-2.10	0.19	***
Female	3,615	37.85	33.02	-4.83	0.20	***	54.7	90.5	38.68	36.50	-2.19	0.16	***
Race													
White	5,779	38.64	33.28	-5.35	0.17	***	59.6	88.8	39.64	37.47	-2.17	0.12	***
Black	222	35.57	31.16	-4.41	0.87	***	56.6	91.9	35.82	33.91	-1.91	0.68	**
Asian	30	40.47	38.30	-2.17	2.05	ns	34.9	96.7	41.04	39.62	-1.41	1.97	ns
Hispanic	33	36.09	32.77	-3.32	1.74	ns	34.4	93.9	37.07	34.88	-2.18	1.63	ns
North American Native	9	41.52	39.47	-2.05	4.25	ns	0.0	100.0	41.52	39.47	-2.05	4.25	ns
Other	52	37.96	35.71	-2.26	1.23	ns	48.2	94.2	39.06	37.89	-1.17	1.13	ns
Unknown	13	32.56	21.71	-10.84	3.33	**	40.7	76.9	34.66	28.23	-6.43	3.13	ns
Original Reason for Entitlement													
Aged	5,313	39.71	34.12	-5.58	0.18	***	57.7	88.8	40.79	38.43	-2.36	0.13	***
Disabled	818	30.73	27.47	-3.25	0.39	***	74.7	90.7	31.11	30.29	-0.82	0.29	**
ESRD	3	47.13	47.01	-0.12	5.77	ns	0.0	100.0	47.13	47.01	-0.12	5.77	ns
ESRD and Disabled	4	28.57	19.73	-8.84	6.20	ns	58.8	75.0	29.95	26.31	-3.64	4.79	ns
Medicaid Status													
No Medicaid	5,565	39.15	33.92	-5.23	0.17	***	58.1	89.5	40.09	37.90	-2.19	0.13	***
Medicaid	573	32.29	26.60	-5.69	0.55	***	69.8	84.6	33.15	31.42	-1.72	0.41	***

Table 2-7 (Continued)

Difference in Mean PCS Score Between Baseline and Follow-up by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (followup PCS of decedents = 0)

	Decedents & Survivors							Survivors					
	PCS Follow-Up Respondents	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Change Due to PCS=0 for Decedents[1]	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Age													
Under 65	470	32.04	29.67	-2.37	0.47	***	65.5	94.3	32.30	31.48	-0.82	0.37	*
65-74	2,505	41.99	38.25	-3.75	0.23	***	49.3	94.1	42.55	40.65	-1.90	0.17	***
75-84	2,419	37.94	32.04	-5.91	0.27	***	56.5	88.3	38.85	36.28	-2.57	0.20	***
85 and Older	744	32.68	22.50	-10.18	0.56	***	73.6	71.1	34.33	31.64	-2.69	0.41	***
Household Income													
Less than $10,000	726	32.83	26.99	-5.84	0.48	***	66.1	85.4	33.58	31.60	-1.98	0.34	***
$10,000-$19,999	1,036	36.53	30.94	-5.58	0.39	***	58.3	88.2	37.40	35.07	-2.33	0.28	***
$20,000-$49,999	1,778	39.87	35.36	-4.51	0.29	***	53.7	91.6	40.68	38.60	-2.09	0.21	***
$50,000 or more	595	43.85	39.92	-3.93	0.50	***	63.4	92.3	44.70	43.26	-1.44	0.37	***
Not given	483	36.92	31.06	-5.85	0.62	***	68.3	85.1	38.36	36.51	-1.86	0.46	***
Missing/No Response	2,003	38.79	32.81	-5.99	0.31	***	59.9	87.5	39.88	37.48	-2.40	0.23	***
Education													
Not a HS Graduate	1,383	35.06	29.49	-5.57	0.35	***	64.8	86.1	36.21	34.24	-1.96	0.37	***
High School Graduate or GED	1,747	38.69	33.80	-4.89	0.29	***	54.5	90.6	39.55	37.33	-2.22	0.21	***
Some College or 2 year degree	1,038	40.06	35.53	-4.53	0.38	***	57.1	91.2	40.89	38.95	-1.94	0.28	***
4 year college graduate	431	40.85	36.12	-4.73	0.60	***	55.6	91.9	41.42	39.32	-2.10	0.43	***
More than a 4 year college degree	490	41.94	37.50	-4.44	0.56	***	61.6	91.2	42.81	41.10	-1.71	0.39	***
Missing/No Response	1,049	38.63	31.76	-6.87	0.45	***	60.2	86.0	39.67	36.94	-2.74	0.34	***
Marital Status													
Not Married	2,122	39.65	35.19	-4.46	0.22	***	58.2	91.1	40.49	38.63	-1.86	0.16	***
Married	3,001	36.80	31.17	-5.63	0.28	***	59.3	87.6	37.87	35.58	-2.29	0.20	***
Missing/No Response	1,015	38.69	31.76	-6.93	0.46	***	60.2	85.9	39.73	36.97	-2.76	0.35	***
Home Ownership													
Owned by Beneficiary or Family Member	4,042	39.46	34.72	-4.74	0.19	***	55.8	91.0	40.26	38.16	-2.09	0.14	***
Not Owned by Beneficiary or Family Member	924	34.38	28.58	-5.81	0.42	***	65.8	84.7	35.71	33.72	-1.99	0.32	***
Missing/No Response	1,172	38.45	31.76	-6.69	0.43	***	62.9	85.7	39.56	37.08	-2.48	0.31	***
Retirement Community													
Yes	847	37.34	31.23	-6.11	0.46	***	62.5	86.0	38.62	36.34	-2.29	0.34	***
No	4,106	38.75	34.14	-4.61	0.19	***	57.6	90.8	39.56	37.61	-1.96	0.14	***
Missing/No Response	1,185	38.50	31.54	-6.96	0.42	***	60.2	85.2	39.77	37.01	-2.77	0.32	***

Table 2-7 (Continued)

Difference in Mean PCS Score Between Baseline and Follow-up by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (followup PCS of decedents = 0)

	Decedents & Survivors							Survivors					
	PCS Follow-Up Respondents	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Change Due to PCS=0 for Decedents[1]	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Chronic Conditions													
Hypertenstion or high blood pressure	2,674	36.92	31.99	-4.93	0.24	***	59.1	89.5	37.76	35.75	-2.02	0.18	***
Angina pectoris or coroary artery disease	942	33.31	27.59	-5.72	0.41	***	67.4	84.7	34.43	32.57	-1.87	0.31	***
Congestive heart failure	428	28.98	21.91	-7.07	0.66	***	82.1	75.2	30.39	29.13	-1.26	0.49	**
Acute myocardial infarction or heart attack	646	33.06	26.83	-6.22	0.51	***	71.6	82.0	34.47	32.71	-1.77	0.38	***
Other heart conditions	1,302	34.09	28.71	-5.38	0.35	***	68.5	85.8	35.15	33.46	-1.69	0.26	***
Stroke	506	30.63	23.42	-7.22	0.61	***	75.5	77.3	32.07	30.30	-1.77	0.48	***
Emphysema, asthma, or COPD	696	32.03	26.59	-5.44	0.49	***	79.0	82.3	33.44	32.30	-1.14	0.37	**
Crohn's disease, ulcerative colitis, or inflamatory bowel disease	366	32.68	29.49	-3.19	0.58	***	73.1	90.4	33.47	32.61	-0.86	0.46	ns
Arthritis of the hip or knee	2,111	33.22	28.95	-4.28	0.26	***	66.2	89.0	33.97	32.52	-1.44	0.20	***
Arthritis of the hand or wrist	1,843	34.00	29.88	-4.12	0.26	***	62.9	89.5	34.92	33.39	-1.53	0.20	***
Sciatica	1,334	33.41	29.89	-3.52	0.31	***	65.9	90.9	34.10	32.89	-1.20	0.25	***
Diabetes, high blood sugar, or sugar in the urine	879	34.10	28.36	-5.75	0.41	***	59.8	86.3	35.15	32.84	-2.31	0.30	***
Any cancer (other than skin cancer)	920	35.78	30.11	-5.67	0.42	***	68.3	85.1	37.18	35.38	-1.80	0.14	***
Number of Chronic Conditions Reported[5]													
0	1,562	42.48	35.97	-6.51	0.35	***	53.7	88.9	43.49	40.48	-3.01	0.25	***
1	901	44.12	39.27	-4.85	0.44	***	58.4	91.6	44.91	42.89	-2.02	0.32	***
2	1,043	41.04	36.14	-4.90	0.38	***	50.2	91.9	41.79	39.35	-2.44	0.29	***
3	928	37.02	32.56	-4.46	0.40	***	57.5	90.9	37.70	35.80	-1.90	0.29	***
4	684	33.93	29.52	-4.41	0.45	***	66.4	88.7	34.75	33.26	-1.48	0.35	***
5	483	30.70	25.72	-4.99	0.57	***	73.8	85.7	31.31	30.00	-1.31	0.43	**
6	257	30.19	23.68	-6.51	0.81	***	75.8	79.8	31.26	29.69	-1.57	0.59	**
7	146	27.34	22.41	-4.93	0.99	***	84.2	82.1	28.05	27.27	-0.78	0.76	ns
8+	134	24.57	18.97	-5.60	1.00	***	81.5	77.6	25.48	24.45	-1.04	0.78	ns
Number of ADLs Difficult or Unable to Dó[5]													
0	3,621	44.49	38.81	-5.69	0.22	***	45.4	92.5	45.08	41.96	-3.11	0.16	***
1	799	35.60	31.04	-4.56	0.46	***	67.3	89.9	36.03	34.54	-1.49	0.34	***
2	668	30.13	26.06	-4.07	0.48	***	82.8	87.7	30.41	29.71	-0.70	0.35	*
3	320	27.73	23.64	-4.09	0.67	***	91.7	85.0	28.15	27.81	-0.34	0.50	ns
4	267	24.97	20.63	-4.34	0.73	***	98.1	83.1	24.90	24.82	-0.08	0.48	ns
5	231	23.57	17.72	-5.85	0.81	***	100.0	73.2	24.08	24.22	0.13	0.58	ns
6	232	24.52	17.64	-6.88	0.92	***	100.0	64.2	26.01	27.46	1.45	0.75	ns

Table 2-7 (Continued)

Difference in Mean PCS Score Between Baseline and Follow-up by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (followup PCS of decedents = 0)

	Decedents & Survivors							Survivors					
	PCS Follow-Up Respondents	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Change Due to PCS=0 for Decedents[1]	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Health in General													
Excellent	284	54.13	48.16	-5.97	0.69	***	20.2	97.5	54.14	49.38	-4.76	0.53	***
Very Good	1,302	48.61	43.40	-5.22	0.35	***	38.6	95.5	48.66	45.46	-3.20	0.24	***
Good	2,380	40.78	35.09	-5.69	0.27	***	46.8	92.1	41.14	38.11	-3.03	0.20	***
Fair	1,631	29.56	25.25	-4.32	0.32	***	94.3	84.9	29.98	29.73	-0.25	0.24	ns
Poor	530	22.78	16.73	-6.05	0.57	***	100.0	67.9	23.07	24.63	1.55	0.39	***
Missing/No Response	11	31.95	24.33	-7.61	4.43	ns	59.2	81.8	32.85	29.74	-3.11	3.66	ns
Baseline PCS Score													
0-20	503	17.20	16.09	-1.11	0.52	*	100.0	72.8	17.26	22.12	4.86	0.39	***
21-30	1,386	25.71	22.46	-3.25	0.34	***	100.0	82.0	25.80	27.41	1.60	0.23	***
31-40	1,386	35.36	30.37	-4.99	0.35	***	69.5	89.5	35.44	33.92	-1.52	0.25	***
41-50	1,496	45.90	39.35	-6.55	0.33	***	34.9	94.5	45.93	41.66	-4.27	0.23	***
51+	1,367	54.41	46.67	-7.74	0.34	***	26.7	95.8	54.42	48.74	-5.68	0.22	***

NOTES:

[1]This value is calculated by the formula 1-(Mean Follow-Up PCS for Survivors/Mean Baseline PCS for all)

[2]Percentage of baseline respondents alive as of the followup survey.

[3]SGA refers to a small geographic area selected for sampling within the given state.

[4]PGP refers to a primary group practice whose members were selected for sampling. One PGP was selected for each state sampled, excluding Georgia.

[5]A zero in either of these categories could also indicate that the recipient did not respond to any of the questions included in the measure.

*p<0.05, **p<0.01, ***p<0.001; ns=not statistically significant at 0.05 level

Output: n09a, a09resp, n10, a10_resp, joinx03

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Table 2-8

Difference in Mean MCS Score Between Baseline and Follow-up by Subsample and Baseline Demographic Characteristics

	MCS Follow-Up Respondents	Mean Baseline MCS Score	Standard Error of the Mean	Mean Follow-Up MCS Score	Standard Error of the Mean	Mean Difference in MCS	Standard Error of the Mean Difference	P
ALL	5,465	51.46	0.14	51.01	0.15	-0.46	0.13	***
Site								
National	484	49.41	0.53	49.84	0.53	0.44	0.45	ns
PA SGA[1]	492	50.19	0.50	49.55	0.52	-0.63	0.45	ns
GA SGA	482	51.08	0.49	49.95	0.53	-1.13	0.45	*
WI SGA	667	51.43	0.41	51.33	0.42	-0.10	0.38	ns
AZ SGA	481	52.16	0.47	51.52	0.49	-0.64	0.45	ns
WA SGA	497	51.70	0.46	52.07	0.47	0.37	0.43	ns
PGP[2] A	574	50.65	0.45	49.43	0.47	-1.22	0.43	**
PGP B	651	51.93	0.39	51.80	0.39	-0.13	0.38	ns
PGP C	588	53.37	0.37	52.23	0.39	-1.14	0.39	**
PGP D	549	52.24	0.44	51.85	0.45	-0.40	0.42	ns
Sex								
Male	2,194	51.78	0.22	51.39	0.23	-0.39	0.21	ns
Female	3,271	51.25	0.19	50.75	0.19	-0.50	0.17	**
Race								
White	5,133	48.10	0.80	48.43	0.85	0.34	0.75	ns
Black	204	51.86	1.73	46.99	1.97	-4.86	2.09	*
Asian	29	50.75	2.94	44.80	4.35	-5.95	5.20	ns
Hispanic	31	50.41	1.81	49.64	1.64	-0.77	1.70	ns
North American Native	9	56.97	2.19	53.71	3.13	-3.25	3.31	ns
Other	49	39.85	2.18	42.77	2.24	2.92	2.26	ns
Unknown	10	51.67	0.15	51.20	0.15	-0.47	0.14	***
Original Reason for Entitlement								
Aged	4,717	52.63	0.14	52.02	0.15	-0.61	0.14	***
Disabled	742	44.07	0.48	44.50	0.47	0.43	0.43	ns
ESRD	3	54.42	3.71	60.74	0.71	6.32	3.05	ns
ESRD and Disabled	3	41.67	7.84	50.01	4.05	8.35	4.89	ns
Medicaid Status								
No Medicaid	4,980	52.11	0.14	51.53	0.15	-0.58	0.14	***
Medicaid	485	44.84	0.55	45.67	0.58	0.83	0.56	ns

Table 2-8 (Continued)

Difference in Mean MCS Score Between Baseline and Follow-up by Subsample and Baseline Demographic Characteristics

	MCS Follow-Up Respondents	Mean Baseline MCS Score	Standard Error of the Mean	Mean Follow-Up MCS Score	Standard Error of the Mean	Mean Difference in MCS	Standard Error of the Mean Difference	P
Age								
Under 65	443	41.56	0.63	42.50	0.62	0.94	0.55	ns
65-74	2,357	53.09	0.19	52.74	0.20	-0.35	0.19	ns
75-84	2,136	52.17	0.22	51.41	0.22	-0.76	0.21	***
85 and Older	529	49.66	0.45	48.77	0.48	-0.89	0.48	ns
Household Income								
Less than $10,000	620	46.51	0.49	46.69	0.49	0.18	0.45	ns
$10,000-$19,999	914	50.73	0.37	50.54	0.37	-0.19	0.33	ns
$20,000-$49,999	1,629	52.80	0.23	52.46	0.25	-0.33	0.22	ns
$50,000 or more	549	54.26	0.37	53.94	0.38	-0.31	0.36	ns
Not given	411	50.07	0.54	49.73	0.56	-0.34	0.49	ns
Missing/No Response	1,753	51.49	0.25	50.50	0.26	-0.99	0.25	***
Education								
Not a HS Graduate	1,191	49.27	0.32	48.55	0.34	-0.72	0.31	*
High School Graduate or GED	1,582	51.64	0.26	51.64	0.27	0.01	0.24	ns
Some College or 2 year degree	947	52.33	0.33	52.06	0.34	-0.27	0.29	ns
4 year college graduate	396	53.72	0.49	53.33	0.48	-0.39	0.41	ns
More than a 4 year college degree	447	54.06	0.41	54.10	0.41	0.04	0.39	ns
Missing/No Response	902	50.88	0.36	49.47	0.37	-1.41	0.38	***
Marital Status								
Not Married	1,859	52.71	0.18	52.15	0.19	-0.55	0.17	***
Married	2,734	49.86	0.26	50.01	0.27	0.15	0.24	ns
Missing/No Response	872	50.99	0.37	49.53	0.38	-1.46	0.39	***
Home Ownership								
Owned by Beneficiary or Family Member	3,678	52.30	0.16	52.02	0.17	-0.27	0.15	ns
Not Owned by Beneficiary or Family Member	783	48.17	0.43	48.15	0.43	-0.02	0.39	ns
Missing/No Response	1,004	50.99	0.34	49.52	0.35	-1.48	0.36	***
Retirement Community								
Yes	728	51.38	0.39	51.09	0.39	-0.29	0.37	ns
No	3,727	51.66	0.17	51.38	0.18	-0.29	0.15	ns
Missing/No Response	1,010	50.79	0.34	49.58	0.35	-1.20	0.36	***

Table 2-8 (Continued)

Difference in Mean MCS Score Between Baseline and Follow-up by Subsample and Baseline Demographic Characteristics

	MCS Follow-Up Respondents	Mean Baseline MCS Score	Standard Error of the Mean	Mean Follow-Up MCS Score	Standard Error of the Mean	Mean Difference in MCS	Standard Error of the Mean Difference	P
Chronic Conditions								
Hypertenstion or high blood pressure	2,393	51.26	0.21	51.03	0.22	-0.23	0.19	ns
Angina pectoris or coroary artery disease	798	50.21	0.37	49.99	0.38	-0.22	0.35	ns
Congestive heart failure	322	48.74	0.59	49.05	0.62	0.31	0.60	ns
Acute myocardial infarction or heart attack	530	50.32	0.45	49.93	0.46	-0.39	0.42	ns
Other heart conditions	1,117	50.16	0.32	49.94	0.34	-0.21	0.30	ns
Stroke	391	48.55	0.57	48.11	0.60	-0.45	0.59	ns
Emphysema, asthma, or COPD	573	49.57	0.47	48.93	0.49	-0.64	0.44	ns
Crohn's disease, ulcerative colitis, or inflamatory bowel disease	331	47.98	0.66	47.76	0.64	-0.21	0.54	ns
Arthritis of the hip or knee	1,879	50.61	0.25	50.39	0.26	-0.22	0.22	ns
Arthritis of the hand or wrist	1,649	50.27	0.27	50.19	0.28	-0.08	0.24	ns
Sciatica	1,212	49.74	0.32	49.43	0.33	-0.31	0.30	ns
Diabetes, high blood sugar, or sugar in the urine	759	50.06	0.39	49.91	0.41	-0.16	0.38	ns
Any cancer (other than skin cancer)	783	51.79	0.37	51.79	0.37	-0.01	0.34	ns
Number of Chronic Conditions Reported[3]								
0	1,388	51.66	0.29	50.70	0.29	-0.95	0.28	***
1	825	53.37	0.33	52.36	0.35	-0.64	0.30	*
2	958	53.03	0.30	52.66	0.32	-0.37	0.29	ns
3	844	51.16	0.36	51.29	0.36	0.13	0.32	ns
4	607	50.92	0.44	50.48	0.45	-0.44	0.40	ns
5	414	48.20	0.55	48.49	0.58	0.29	0.55	ns
6	205	48.88	0.79	48.52	0.82	-0.36	0.73	ns
7	120	48.30	1.05	47.42	1.09	-0.88	0.97	ns
8+	104	46.75	1.14	46.71	1.03	-0.04	1.03	ns
Number of ADLs Difficult or Unable to Do[3]								
0	3,349	53.37	0.16	52.59	0.17	-0.79	0.16	***
1	718	51.51	0.40	51.24	0.41	-0.26	0.36	ns
2	586	49.93	0.47	49.54	0.46	-0.39	0.43	ns
3	272	45.98	0.65	47.07	0.74	1.08	0.67	ns
4	222	45.65	0.79	45.78	0.78	0.14	0.80	ns
5	169	44.27	0.92	44.62	0.97	0.35	0.80	ns
6	149	41.30	0.96	42.34	1.03	1.04	1.01	ns

Table 2-8 (Continued)

Difference in Mean MCS Score Between Baseline and Follow-up by Subsample and Baseline Demographic Characteristics

	MCS Follow-Up Respondents	Mean Baseline MCS Score	Standard Error of the Mean	Mean Follow-Up MCS Score	Standard Error of the Mean	Mean Difference in MCS	Standard Error of the Mean Difference	P
Health in General								
Excellent	277	57.19	0.38	56.32	0.42	-0.88	0.43	*
Very Good	1,243	56.11	0.20	55.17	0.23	-0.94	0.22	***
Good	2,191	52.91	0.20	52.26	0.21	-0.64	0.21	**
Fair	1,385	47.19	0.29	47.10	0.30	-0.09	0.29	ns
Poor	360	38.86	0.63	40.08	0.65	1.22	0.64	ns
Missing/No Response	9	45.02	5.31	44.42	4.95	-0.60	2.29	ns
Baseline MCS Score								
0-30	275	25.17	0.27	34.87	0.76	9.70	0.75	***
31-40	677	36.14	0.11	42.16	0.40	6.02	0.40	***
41-50	1,032	46.06	0.09	47.74	0.30	1.68	0.30	***
51-60	2,551	56.36	0.05	54.03	0.16	-2.33	0.16	***
61+	930	62.97	0.07	57.55	0.27	-5.43	0.27	***

NOTES:

[1]SGA refers to a small geographic area selected for sampling within the given state.

[2]PGP refers to a primary group practice whose members were selected for sampling. One PGP was selected for each state sampled, excluding Georgia.

[3]A zero in either of these categories could also indicate that the recipient did not respond to any of the questions included in the measure.

*p<0.05, **p<0.01, ***p<0.001; ns=not statistically significant at 0.05 level

Output: n11, a11_resp, joinx03

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

modestly healthier than the cohort that includes those that died between baseline and follow-up, the principal factor that is underlying the observed differences in mean PCS decline (5.27 versus 2.15 points) is the setting of PCS scores to zero for decedents. In fact, almost 60 percent of the difference in the mean PCS scores is due to setting PCS to zero[1].

A similar pattern of statistically and clinically significant declines in mean PCS scores are observed across all ten sampling strata. We also observe the same pattern that survivors have a modestly higher average PCS score at baseline and a more modest decline in mean PCS scores as compared to the cohort that contains both survivors and decedents. The percent of difference in mean PCS scores directly attributable to setting the PCS score to zero for decedents ranges from 49 percent to 72 percent.

Men exhibit a higher average PCS score at baseline compared to women, 39.4 versus 37.9). However, men also experience a larger decline at follow-up, a 5.9 point decline versus a 4.8 point decline. And, the percent of difference in mean PCS scores directly attributable to setting the PCS score to zero for decedents is greater for males than for females, 65 percent versus 55 percent.

Although mean PCS scores at baseline varied across the six racial groups, all racial groups experienced clinically meaningful declines in physical health during the

1 The percent of mean PCS score difference due to setting PCS = 0 for decedents is calculated as follows: [1 – (ratio of mean difference in scores for survivors to mean difference in scores for survivors and decedents)]*100.

two year period. However, whites and blacks experienced the largest declines, 5.4 and 4.4 points respectively. These were also statistically significant changes. Medicare beneficiaries originally entitled to Medicare due to disability had a baseline PCS score that was nine points lower than the average PCS score for Medicare beneficiaries entitled due to age. However, the latter group of beneficiaries experienced a considerably larger decline over the two year period, 5.6 points. The disabled experienced a 3.3 point decline; 75 percent of the decline was due to decedents have a follow-up score of zero. Thus, the disabled who survived to follow-up declined less in physical health than the aged who survived. This is likely due to the disabled being worse off than the aged at baseline. We observe no appreciable differential rate of decline in physical health based upon enrollment in Medicaid.

Not surprisingly, we observe Medicare beneficiaries age 85 and over experiencing the largest decline in physical health, 10 points during the two year period. Almost 75 percent of the decline is attributable to decedents with PCS scores of zero. Differences in mean scores for physical health for survivors age 85 and over is similar to the decline observed for beneficiaries age 75 to 84 who survive to the time of follow-up. Although there is some variation in level of decline, we observe similar patterns of decline across all strata of other demographic variables contained in Table 2-7.

The presence of congestive heart failure (CHF), acute myocardial infarction (AMI), and stroke at baseline appear to have a more significant effect on physical health than other chronic conditions contained in the HOS instrument. Mean PCS scores

decline by over 7 points, on average, for beneficiaries with CHF or stroke, and by 6.2 points for those with a history of an AMI. Mortality has a significant influence on the change in mean scores. Over 80 percent of the change in the mean PCS score for beneficiaries with CHF is due to mortality and setting PCS equal to zero. CHF survivors, on average, do not experience a clinically meaningful reduction in physical health during the two year period. In fact, only diabetics and hypertensives experience a decline in physical health that approximates a clinically meaningful change.

Although we observe a linear decrease in average PCS scores as the number of chronic conditions or limitations in activities of daily living increase, there is no consistent pattern of increasing decline in mean physical health scores as the number of chronic conditions or limitations in activities of daily living increase. Nor is there any particular pattern of decline based upon self-reported general health at baseline, although there are significant differences in average scores at baseline. We observe the average PCS score for those in excellent health decline by roughly the same amount as the average PCS score for those in poor health at baseline, or 6 points. For those who survive to follow-up, we observe regression to the mean, which has been observed in previous studies of change in health status over time. The mean PCS score for those in poor health at baseline rises at follow-up, while the mean PCS score for those in excellent health at baseline declines.

Lastly, we examined changes in average PCS scores as a function of baseline PCS scores. Once again, we observe regression to the mean. Those in poorest health at

baseline experienced on average the least amount of degradation in physical health over the two year period as measured by the PCS; while those with the highest average PCS scores at baseline experienced the greatest degradation of health, 1.1 point decline versus 7.7 point decline. In fact, those with PCS scores at 30 or below who survived the two-year period experienced, on average, an increase in their physical component score.

In contrast, there is relatively little change in average mental component scores between the two time periods (Table 2-8). The mean MCS score at baseline is 51.46 and 51.01 at follow-up. This is an insignificant clinical change. There are relatively few statistically or clinically meaningful differences in average MCS scores across the sampling cohorts or across the various stratifications of the sociodemographic and Medicare enrollment and eligibility characteristics.

The noted exception is change in average MCS scores evaluated as a function of baseline MCS scores. We observe a marked improvement in average follow-up MCS scores for those beneficiaries who had a baseline score of less than 30; almost a 10 point improvement in mean MCS. As the baseline score increases to the normative mean of 50, the magnitude of positive change declines. As expected, average MCS scores for beneficiaries who scored 51 points or higher at baseline actually experienced a statistically and clinically meaningful decline in average mental health during the two year period. The cohort that scored 61 or higher at baseline experienced the greatest decline in mental health, the average MCS score declined by over 5 points.

Appendix Tables 1 and 2 contain data similar to Tables 2-7 and 2-8, but only for those respondents where the SF-12 was used for measuring PCS and MCS. They provide comparisons to the data for all respondents in Tables 2-7 and 2-8. Appendix Table 3 contains supplemental data on the consistency of chronic disease condition reports.

2.6 Usual source of care at Baseline and Follow-up

The last issue examined in this chapter is the proportion of follow-up respondents that identified their usual source of care as the same physician group practice to which they were assigned at baseline for our four group practices. This answers a face validity question regarding ownership of follow-up health status. Because there is no concept of enrollment in FFS, we are interested in examining at what rate beneficiaries who said that their usual source of care was one of the four physician group practices at baseline continued to use the same group practice for their usual source of care at follow-up.

At baseline, beneficiaries were assigned to a group practice if that group provided at least one primary care visit in 1997 and provided equal or more primary care than any other group practice. In both the baseline and follow-up survey, we asked the respondent if their usual source of care was one of the four group practices participating in our study. Table 2-9 displays the number of baseline and follow-up respondents who answered the usual source of care question at both baseline and follow-up, and who reported that their

Table 2-9

Agreement Between Assignment of Beneficiaries to Physician Group Practices (PGPs) and Beneficiary Perceptions of Doctor Visits

	BASELINE					FOLLOW-UP				
		'Most Doctor Visits Last Year at This Clinic?'[3]					'Most Doctor Visits Last Year at This Clinic?'[3]			
Assigned to PGP[1]	Respondents[2]	Yes	No	Missing	Percent Agreement[4]	Respondents[5]	Yes	No	Missing	Percent Agreement[4]
A	704	437	83	184	84.0	435	275	117	43	70.2%
B	768	540	46	182	92.2	391	277	64	50	97.9%
C	658	443	42	173	91.3	417	314	57	46	84.6%
D	703	491	77	135	86.4	482	379	50	53	88.3%

NOTES:

[1] Based on frequency of physician visits recorded on Medicare physician claims.

[2] Surveyed beneficiaries for whom an SF-36 or SF-12 score can be computed.

[3] Question posed on the fee-for-service Health Outcomes Survey.

[4] Number answering "Yes" divided by sum of those answering "Yes" or "No".

[5] Number of follow-up respondents who had identified the group practice as their usual source of care at baseline.

SOURCE: RTI International analysis of the 1998 (baseline) and 2000 (follow-up) Medicare Fee-for-Service Health Outcomes Survey.

usual source of care at both baseline and follow-up was the group practice to which they had been assigned at baseline.

At baseline, 88 percent of respondents who answered this question said that the group practice to which they had been assigned was their usual source of care. The percentage of agreement is defined as number answering yes divided by total number of respondents answering this item. PGP A had the lowest rate of agreement, 84 percent, while PGP B had the highest rate of agreement, 92.2 percent. For the follow-up, we consider only respondents who answered yes at baseline and provided an answer at follow-up. Of the 1,725 affirmative respondents at baseline, 1,533 also responded at follow-up, or 89 percent. For these respondents, 81 percent affirmed that the PGP to which they had been assigned at baseline remained their usual source of care. The percentage of beneficiaries who remained most often affiliated with their assigned PGP ranged from 70.2 percent to 97.9 percent. We consider this to be a relatively high rate of retention, given the lack of an enrollment concept in FFS.

2.7 Summary

In this chapter, we explored four selected issues related to the feasibility of using health status follow-up scores in Medicare FFS for performance measurement. First, we were interested in determining whether there were systematic differences in respondents to the baseline HOS versus the follow-up HOS. We analyzed this issue through an analysis of response rates and distribution of respondents across strata of interest.

Second, we were interested in determining whether there were any systematic differences in the completeness of survey responses as the scoring of the PCS and MCS measures are highly dependent on item response. Third, we were interested in exploring differences in mean PCS and MCS scores between baseline and follow-up and directly examining the effect of setting the PCS score to zero at follow-up for decedents. Fourth, we were interested in exploring the degree of retention of baseline respondents at follow-up by the four physician group practices.

The analyses presented in this chapter revealed no systematic differences in response rates or material differences in characteristics of follow-up respondents that were not expected. The overall follow-up response rate was an impressive 92 percent. We found no differences in response rates across the national sample, four group practices, and five geographic areas. Further, we observed a high rate of retention of beneficiaries by the four physician group practices. Over 80 percent of beneficiaries remained with their usual source of care providers between baseline and follow-up.

We did not observe any systematic differences in the item completion rate of the survey instrument between baseline and follow-up. The same scoring method could be used for all but a handful of respondents; thus we believe any observed differences in mean PCS and MCS scores between the two time periods would be a reflection of true changes in average physical and mental health and not an artifact of changes in scoring methods.

As expected, we observed aging of the cohort and some degradation of physical health as assessed through an increase in number of chronic conditions and activities of daily living limitations as well as a reduction in the proportion of respondents that rated their health as excellent or who received a PCS score above 50. We did not observe much change in average mental health status over the two-year period.

We did observe a statistically and clinically significant decline in average physical health scores of roughly 5 points at follow-up. This pattern held for all 10 sampling strata. There was some variation across the level of physical health decline between men and women, across racial cohorts, and across the age groups. The very old experienced the largest decline, 10 points on average between the two years. Those originally entitled to Medicare due to disability saw a significantly larger decline than those entitled due to age. The disabled also experienced a disproportionately larger decline in average physical health than the non-disabled. Beneficiaries with selected chronic conditions, such as CHF, also experienced larger average declines in physical health than those without the disease. And, not surprisingly, we observed those with the highest baseline PCS scores showing the largest average decline in health over the two-year period relative to those with the lowest PCS scores. This is the regression to the mean phenomenon that has been observed in other studies of changes in health status.

The observed differences in mean PCS scores is heavily influenced by setting the PCS to zero for decedents. Restricting our analysis to only survivors, we observed a decline in average PCS of roughly two points, barely clinically significant. Many of the

observed statistical and clinical differences that we observed across the strata are a function of a disproportionate rate of death across strata. Removing the decedents from the follow-up analysis revealed minimal differences in mean follow-up PCS scores across the strata. Removing all decedents from the follow-up health status score estimation appeared to underestimate decline, on average, while setting death to zero appeared to overestimate physical health decline, on average. This raises an issue of fairness of comparisons using differences in mean scores across different health care systems or providers in Medicare FFS, which may penalize providers or health systems if, at baseline, they have a pool of patients with higher than average risk of mortality. This suggests that an alternative approach to estimating physical health at follow-up for decedents should be considered. Further, a methodological approach for analyzing follow-up health status scores that controls baseline selection with respect to probability of mortality should be considered. Two alternative approaches to handling these two issues are considered in the following chapters.

3 Imputing PCS and MCS Scores for Respondents Who Died Between Baseline and Follow-up

3.1 Introduction

As noted, approximately 5 percent of Medicare beneficiaries die each year (Gage *et al*., 2000). Thus we expected a significant number of the beneficiaries who responded to our baseline survey would die before the follow-up survey was administered two years later. This raises a well-known problem, introduced in Chapter 1, in conducting longitudinal studies with traditional health status measures, since they do not provide explicit values for death. Many longitudinal studies using the SF-36 simply ignore deaths, and analyze changes only for those alive at follow-up. However, Diehr *et al*. (1995) have shown that this approach severely underestimates changes in health status, and can significantly bias comparisons of performance.

Several methods for scoring death have appeared in the literature. As noted, one method is to use a PCS or MCS score of zero for death. This was one of two methods used by Ware et al. (1996) for handling death for the PCS in their analysis of data from the Medical Outcomes Study (MOS). An arbitrary score of zero, however, does not represent the "absence" of health and has no explicit meaning on a component score metric. Moreover, the extreme nature of a zero value means that deaths dominate analysis of change scores or follow-up scores.

As noted, a second method, employed as the alternate approach by Ware *et al*. (1996) for analyzing PCS scores, is to collapse changes over time in the PCS and MCS

into three categories, depending on whether the changes are "better," "worse," or "about the same" as expected. However, this approach results in a categorization that is less precise than the original continuous data and treats all deaths as though they represent the same amount of change in health status

More recently, Ware and his colleagues at the Health Assessment Laboratory (HAL) modified their method for handling deaths for PCS scores as part of their analysis of HOS surveys of M+C beneficiaries (Rogers *et al*., 2000; 2001). This new method is also based on categorizing respondents at follow-up. However, it employs a two-part model for analysis of expected PCS outcomes, estimating both the probability of death and the probability that PCS scores are the same or better. The new method continues to treat MCS scores differently, still excluding scores for any respondents who died (Rogers *et al*., 2000; 2001). As noted, a number of other methods have been proposed for handling death of respondents between baseline and follow-up. (Diehr *et al*., 1995; Rogers *et al*., 2000).

3.2 Associating the Utility Score for Death with PCS and MCS Scores

We believe that each of the approaches used up to now is limited and imprecise, so we sought to develop a new method for imputing scores for death that would preserve the original continuous metric of the component scales. To do this, we made use of a concept used in economic evaluations of health known as "utilities." Utilities are preferences for health states and are appealing in this context because they are defined on

a scale ranging from 0 (death) to 1 (optimal health). Thus death has a specific value in the utility approach.

Utility values are elicited by presenting respondents with scenarios involving various health problems and determining, through how they respond to potential time trade-off choices and standard gamble assessments, how those health conditions compare with optimal health. The results are used to assign utility values to specific health states on a scale starting from death at one extreme.

The HOS survey does not directly provide utility assessments. Instead, we relied on questions from the HOS to estimate utilities using the Health and Activity Limitation Index (HALex; Erickson, 1998). The HALex, developed from the 1990 National Health Interview Survey (NHIS), provides utility scores for combinations of self-reported activity limitations and perceived health. It was designed to provide monitoring data for federal health promotion and disease prevention efforts. Utility values were assigned to health states based on a multiplicative, multiattribute model. Values for selected health states were derived from the Health Utilities Index Mark 1 (Erickson, 1998).

Despite its recent development, the HALex method has begun being used by the research community for studying health service issues. In one study its utility scores were found to be negatively associated with the number of comorbidities in a sample of post-acute myocardial infarction patients (Bradley *et al*., 2000).

Using HOS items, we constructed a table cross-classifying three levels of activity limitations by five levels of perceived health. Table 3-1 shows the corresponding HALex

Table 3-1

HALex Utility Scores and Sample Size by Activity Limitation and Perceived Health

	Perceived Health									
	Excellent		Very Good		Good		Fair		Poor	
	Utility Score	Sample Size	Utility Score	Sample Size	Utility Score	Sample Size	Utility Score	Sample Size	Utility Score	Sample Size
No Activity Limitation	1.00	234	0.92	946	0.84	1263	0.63	360	0.47	21
Some Activity Limitation	0.87	12	0.79	209	0.72	856	0.52	994	0.38	297
Unable To Do At Least One ADL	0.47	5	0.41	12	0.36	63	0.21	165	0.10	126

SOURCES: RTI International analysis of the baseline Medicare fee-for-service Health Outcomes Survey; Erickson, 1998.

utility and count of the number of FFS baseline survey respondents in our study falling in each cell. The activity limitation categories were defined using ADL questions in the HOS. No activity limitation means the respondent indicated no difficulty in performing any of the six ADLs. Some activity limitation means the respondent reported difficulty with at least one ADL. The last category indicates the respondent reported being unable to do at least one ADL. Perceived health was defined using a question which asked respondents to rate their own health as "Excellent", "Very Good", "Good", "Fair", or "Poor". The utilities range from 0.1 for respondents in poor health who were unable to do one or more basic activities of daily living, to 1.0 for those reporting excellent health with no activity limitations. The mean HALex utility score for our sample was 0.70 (SD = 0.21). The overall distribution of respondents in the table was similar to that found in the NHIS, with proportionally more respondents in the less healthy cells. This was expected since our Medicare FFS sample is older than the general population sample interviewed for the NHIS.

To define scores equivalent to death on the PCS and MCS scales, we analyzed the relationship between utilities and HOS component scores. Pearson correlations with the utility values were 0.75 for the PCS and 0.45 for the MCS. We regressed the PCS and MCS scores on the HALex utilities, and found that nonlinear, logarithmic models fit the data somewhat better than linear equations. The model was estimated using baseline PCS and MCS scores for the entire baseline sample. Measurement error in a predictor will attenuate slopes in a regression model such as this. However, since the utility values are

deterministic (that is, calculated from the HALex algorithm rather than reported directly by respondents), this type of random reporting error is not likely to be present in these data. As a result, reliability adjustments to correct the slopes for measurement error were not made.

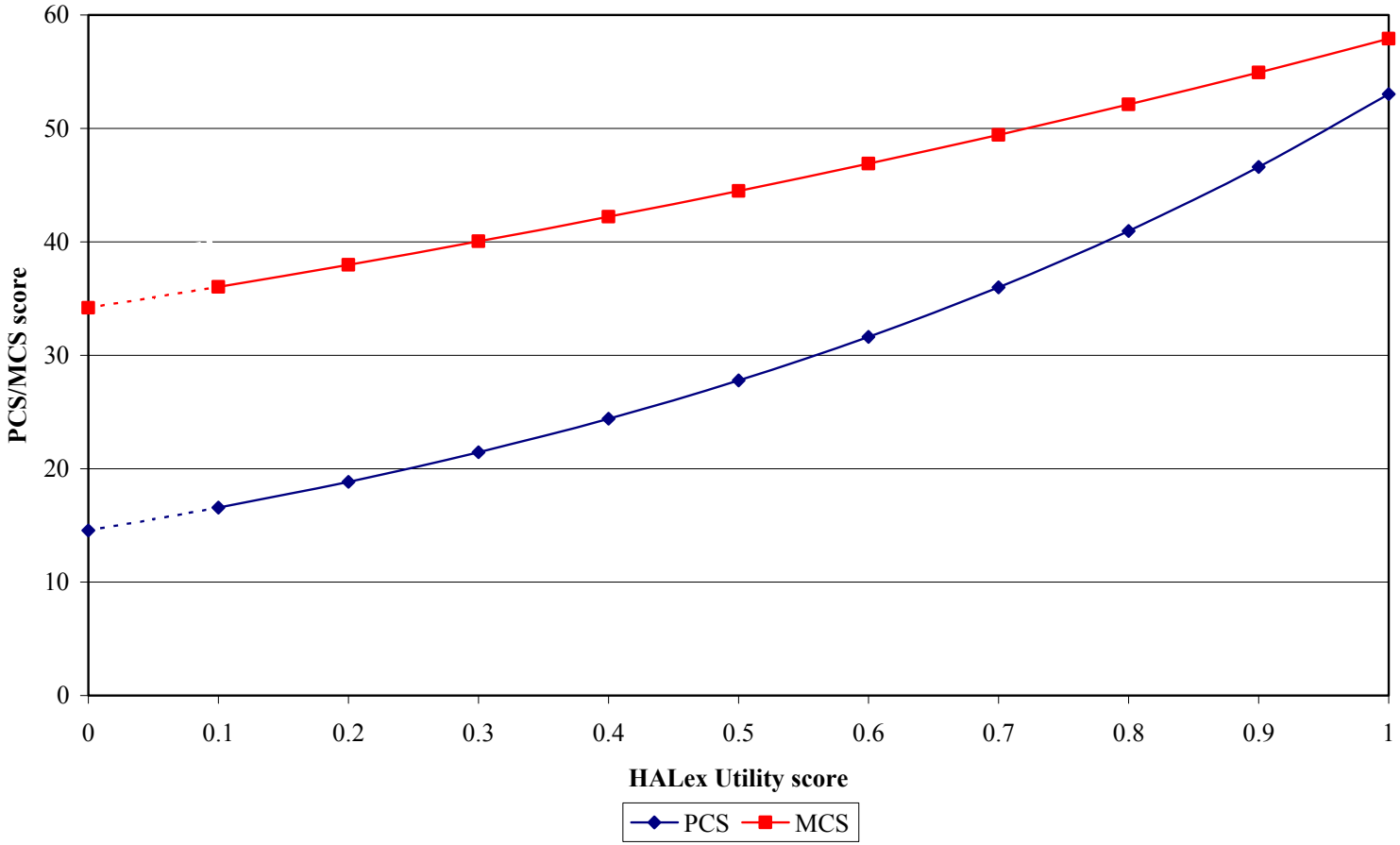
The curvilinear relationships between the utility scores and the PCS and MCS are displayed in Figure 3-1. To determine the component scores equivalent to death, we extrapolated the curves from the observed range (utilities from 0.1 to 1.0) down to a utility score of zero (dotted lines in the figure), which are the points where the regression lines cross the Y-axis. This produced estimates for death of 15 points for the PCS and 34 points for the MCS. In subsequent analyses, we assigned these values for the two-year follow-up scores for the PCS and MCS for all beneficiaries who died between the baseline and follow-up administrations of the HOS survey.

3.3 Comparison of Mean Physical and Mental Health for Medicare Fee-for-Service Beneficiaries Before and After Imputing Health Status Scores for Respondents Who Died Between Baseline and Follow-up

Table 3-2 presents data on mean PCS scores at baseline and follow-up and differences in mean scores after assignment of PCS scores as 15 for those respondents who died between baseline and follow-up, as described above. Its results can be contrasted with those presented in Table 2-7, where PCS scores were set to zero for decedents, following the MOS approach. However, a second change is introduced in

Figure 3-1

Regression Results for PCS and MCS Scores on HALex Utility Scores



SOURCES: RTI International analysis of the baseline Medicare fee-for-service Heatlh Outcomes Survey; Erickson, 1998.

Table 3-2

Mean Baseline and Follow-up PCS Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (PCS Values Imputed as 15.0 for Those Responding at Baseline But Dying Before Follow-up)

	Decedents & Survivors						Survivors					
	PCS Follow-Up Respondents	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Site												
National	544	38.28	34.06	-4.23	0.47	***	89.0%	38.97	36.42	-2.54	0.43	***
PA SGA[3]	559	38.99	34.91	-4.09	0.43	***	88.0	40.16	37.62	-2.54	0.40	***
GA SGA	552	37.48	33.61	-3.87	0.45	***	87.3	38.70	36.31	-2.39	0.44	***
WI SGA	727	37.77	34.05	-3.72	0.37	***	91.7	38.36	35.77	-2.60	0.35	***
AZ SGA	538	39.76	36.10	-3.66	0.44	***	89.4	40.80	38.60	-2.20	0.40	***
WA SGA	553	39.53	35.92	-3.61	0.43	***	89.9	40.59	38.28	-2.31	0.40	***
PGP[4] A	662	37.53	34.11	-3.43	0.40	***	86.7	38.66	37.04	-1.63	0.37	***
PGP B	736	37.77	33.87	-3.90	0.36	***	88.5	38.75	36.34	-2.41	0.34	***
PGP C	654	39.03	35.47	-3.56	0.39	***	89.9	39.84	37.77	-2.07	0.36	***
PGP D	613	39.09	35.91	-3.18	0.41	***	89.6	40.15	38.35	-1.81	0.38	***
Sex												
Male	2,523	39.41	35.40	-4.01	0.21	***	87.0	40.62	38.46	-2.16	0.19	***
Female	3,615	37.82	34.31	-3.51	0.17	***	90.5	38.65	36.35	-2.30	0.16	***
Race												
White	5,779	38.61	34.84	-3.77	0.13	***	88.8	39.61	37.34	-2.27	0.12	***
Black	222	35.42	32.28	-3.14	0.72	***	91.9	35.66	33.81	-1.85	0.68	***
Asian	30	40.47	38.51	-1.96	1.81	ns	96.7	41.04	39.32	-1.71	1.86	ns
Hispanic	33	36.22	33.70	-2.52	1.57	ns	93.9	37.20	34.91	-2.30	1.65	ns
North American Native	9	41.52	39.24	-2.28	4.37	ns	100.0	41.52	39.24	-2.28	4.37	ns
Other	52	37.96	36.53	-1.43	1.06	ns	94.2	39.06	37.85	-1.21	1.11	ns
Unknown	13	32.44	25.58	-6.86	2.29	**	76.9	34.51	28.75	-5.76	2.91	*
Original Reason for Entitlement												
Aged	5,313	39.68	35.71	-3.97	0.14	***	88.8	40.76	38.33	-2.43	0.13	***
Disabled	818	30.67	28.60	-2.07	0.03	***	90.7	31.04	29.99	-1.05	0.30	**
ESRD	3	47.13	47.01	-0.12	5.77	ns	100.0	47.13	47.01	-0.12	5.77	ns
ESRD and Disabled	4	28.57	19.50	-9.07	4.61	ns	75.0	29.95	21.00	-8.95	6.52	ns
Medicaid Status												
No Medicaid	5,565	39.12	35.40	-3.72	0.14	***	89.5	40.06	37.79	-2.27	0.13	***
Medicaid	573	32.21	28.59	-3.62	0.42	***	84.6	33.05	31.05	-1.99	0.41	***

Table 3-2 (Continued)

Mean Baseline and Follow-up PCS Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (PCS Values Imputed as 15.0 for Those Responding at Baseline But Dying Before Follow-up)

	Decedents & Survivors						Survivors					
	PCS Follow-Up Respondents	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Age												
Under 65	470	31.95	30.34	-1.61	0.39	***	94.3	32.20	31.28	-0.92	0.37	**
65-74	2,505	41.96	39.07	-2.89	0.19	***	94.1	42.52	40.58	-1.94	0.17	***
75-84	2,419	37.91	33.66	-4.25	0.22	***	88.3	38.82	36.13	-2.69	0.20	***
85 and Older	744	32.67	26.62	-6.05	0.40	***	71.1	34.31	31.34	-2.97	0.41	***
Household Income												
Less than $10,000	726	32.74	29.17	-3.58	0.36	***	85.4	33.48	31.59	-1.89	0.35	***
$10,000-$19,999	1,036	36.51	32.68	-3.82	0.30	***	88.2	37.38	35.04	-2.34	0.28	***
$20,000-$49,999	1,778	39.83	36.60	-3.23	0.23	***	91.6	40.64	38.57	-2.07	0.21	***
$50,000 or more	595	43.83	41.05	-2.78	0.41	***	92.3	44.68	43.23	-1.45	0.36	***
Not given	483	36.87	33.19	-3.68	0.48	***	85.1	38.31	36.37	-1.94	0.47	***
Missing/No Response	1,520	39.38	34.74	-4.65	0.28	***	87.5	40.35	37.36	-2.99	0.27	***
Education												
Not a HS Graduate	1,383	35.00	31.54	-3.45	0.27	***	86.1	36.14	34.21	-1.93	0.26	***
High School Graduate or GED	1,747	38.67	35.19	-3.48	0.23	***	90.6	39.53	37.30	-2.23	0.21	***
Some College or 2 year degree	1,038	40.01	36.81	-3.21	0.31	***	91.2	40.84	38.91	-1.93	0.29	***
4 year college graduate	431	40.83	37.31	-3.52	0.49	***	91.9	41.39	39.28	-2.11	0.43	***
More than a 4 year college degree	490	41.90	38.75	-3.15	0.44	***	91.2	42.77	41.03	-1.74	0.39	***
Missing/No Response	1,049	38.63	33.34	-5.29	0.36	***	86.0	39.67	36.33	-3.34	0.34	***
Marital Status												
Not Married	2,122	36.74	32.99	-3.75	0.21	***	91.1	37.80	35.54	-2.26	0.20	***
Married	3,001	39.63	36.49	-3.14	0.18	***	87.6	40.47	38.59	-1.88	0.16	***
Missing/No Response	1,015	38.68	33.32	-5.35	0.36	***	85.9	39.72	36.33	-3.39	0.35	***
Home Ownership												
Owned by Beneficiary or Family Member	4,042	39.44	36.05	-3.39	0.15	***	91.0	40.23	38.13	-2.10	0.14	***
Not Owned by Beneficiary or Family Member	924	34.29	30.81	-3.48	0.32	***	84.7	35.60	33.66	-1.94	0.31	***
Missing/No Response	1,172	38.45	33.43	-5.02	0.34	***	85.7	39.56	36.51	-3.05	0.33	***

Table 3-2 (Continued)

Mean Baseline and Follow-up PCS Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (PCS Values Imputed as 15.0 for Those Responding at Baseline But Dying Before Follow-up)

	Decedents & Survivors						Survivors					
	PCS Follow-Up Respondents	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Retirement Community												
Yes	847	37.29	33.29	-4.00	0.36	***	86.0	38.67	36.28	-2.29	0.34	***
No	4,106	38.71	35.49	-3.23	0.15	***	90.8	39.52	37.57	-1.95	0.14	***
Missing/No Response	1,185	38.50	33.29	-5.21	0.33	***	85.2	39.77	36.46	-3.31	0.33	***
Chronic Conditions												
Hypertenstion or high blood pressure	2,674	36.88	33.52	-3.36	0.19	***	89.5	37.72	35.69	-2.03	0.18	***
Angina pectoris or coroary artery disease	942	33.28	29.84	-3.45	0.31	***	84.7	34.41	32.52	-1.89	0.31	***
Congestive heart failure	428	28.93	25.56	-3.37	0.47	***	75.2	30.32	29.04	-1.28	0.49	***
Acute myocardial infarction or heart attack	646	33.03	29.51	-3.52	0.38	***	82.0	34.44	32.69	-1.75	0.38	***
Other heart conditions	1,302	34.09	30.79	-3.31	0.27	***	85.8	35.16	33.40	-1.76	0.26	***
Stroke	506	30.53	26.79	-3.74	0.45	***	77.3	31.94	30.26	-1.68	0.48	***
Emphysema, asthma, or COPD	696	32.02	29.19	-2.83	0.37	***	82.3	33.43	32.24	-1.19	0.37	***
Crohn's disease, ulcerative colitis, or inflamatory bowel disease	366	32.65	30.91	-1.75	0.46	***	90.4	33.44	32.59	-0.85	0.46	*
Arthritis of the hip or knee	2,111	33.19	30.53	-2.66	0.20	***	89.0	33.93	32.45	-1.48	0.20	***
Arthritis of the hand or wrist	1,843	33.97	31.42	-2.55	0.21	***	89.5	34.89	33.35	-1.54	0.20	***
Sciatica	1,334	33.36	31.21	-2.15	0.25	***	90.9	34.04	32.84	-1.20	0.25	***
Diabetes, high blood sugar, or sugar in the urine	879	34.06	30.35	-3.72	0.31	***	86.3	35.11	32.77	-2.34	0.31	***
Any cancer (other than skin cancer)	920	35.75	32.31	-3.44	0.32	***	85.1	37.14	35.34	-1.81	0.31	***
Number of Chronic Conditions Reported[3]												
0	1,562	42.47	37.29	-5.18	0.28	***	88.9	43.47	40.08	-3.39	0.25	***
1	901	44.05	40.51	-3.54	0.36	***	92.3	44.83	42.86	-1.97	0.32	***
2	1,043	41.01	37.34	-3.67	0.31	***	91.9	41.76	39.32	-2.44	0.29	***
3	928	36.98	33.90	-3.08	0.32	***	90.2	37.65	35.78	-1.87	0.29	***
4	684	33.95	31.13	-2.82	0.36	***	86.1	34.76	33.17	-1.59	0.35	***
5	483	30.70	27.83	-2.88	0.43	***	79.5	31.31	29.96	-1.35	0.43	***
6	257	30.06	26.61	-3.45	0.59	***	79.0	31.10	29.55	-1.54	0.61	**
7	146	27.37	25.04	-2.34	0.70	***	75.8	28.09	27.21	-0.88	0.75	ns
8+	134	24.48	22.30	-2.18	0.72	***	72.0	25.36	24.41	-0.96	0.79	ns

Table 3-2 (Continued)

Mean Baseline and Follow-up PCS Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (PCS Values Imputed as 15.0 for Those Responding at Baseline But Dying Before Follow-up)

	Decedents & Survivors						Survivors					
	PCS Follow-Up Respondents	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Health in General												
Excellent	284	54.11	48.51	-5.60	0.60	***	97.5	54.12	49.36	-4.76	0.52	***
Very Good	1,302	48.58	44.03	-4.55	0.29	***	95.5	48.63	45.41	-3.22	0.24	***
Good	2,380	40.74	36.23	-4.52	0.22	***	92.1	41.10	38.06	-3.05	0.20	***
Fair	1,631	29.53	27.28	-2.25	0.24	***	84.9	29.94	29.46	-0.48	0.24	*
Poor	530	22.74	21.21	-1.53	0.36	***	67.9	23.02	24.14	1.12	0.40	***
Missing/No Response	11	31.95	27.06	-4.89	3.45	ns	81.8	32.85	29.74	-3.11	3.66	ns

NOTES:

[1]This value is calculated by the formula 1-(Mean Follow-Up PCS for Survivors/Mean Baseline PCS for all)

[2]Percentage of baseline respondents alive as of the followup survey.

[3]SGA refers to a small geographic area selected for sampling within the given state.

[4]PGP refers to a primary group practice whose members were selected for sampling. One PGP was selected for each state sampled, excluding Georgia.

[5]A zero in this category could also indicate that the recipient did not respond to any of the questions included in this measure.

*p<0.05, **p<0.01, ***p<0.001; ns=not statistically significant at 0.05 level

Output: joinx02e, joinx02f

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Table 3-2. This involved using PCS scores calculated using the SF-36 scoring algorithm whenever possible, instead of matching SF-12 scores with other SF-12 scores if the SF-12 was used for scoring either at baseline or follow-up. The reason for this change was to facilitate the multivariate analysis of factors affecting follow-up scores, which is described in the next chapter. The effect of using the SF-12 would be masked for purposes of the regression analysis if it were consistently matched at baseline and follow-up instead of allowing it to come in at either point only if required to obtain a valid PCS score. As a result, Table 3-2 also provides descriptive statistics for the data set used for the multiple regression analysis presented in the next chapter.

Table 3-2 shows an overall pattern similar to Table 2-7, with the mean difference in PCS scores column showing declining values over time across all of the analytic categories. The declines are clearly smaller in Table 3-2, however, for the decedents and survivors group, reflecting the change from scoring death as 0 to scoring deaths as 15. Some differences in the change scores are also evident for the survivors-only group, but this reflects the change in calculation methods to preferring PCS scores calculated using the SF-36.

Table 3-3 presents similar results for the MCS scores. Here the comparison is to Table 2-8. In Table 2-8, however, only survivors were included, however, since that table replicates the HAL approach, where respondents who died between baseline and follow-up were dropped from the analysis. Once again there are some differences

Table 3-3

Mean Baseline and Follow-up MCS Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (MCS Values Imputed as 34.0 for Those Responding at Baseline But Dying Before Follow-up)

	Decedents & Survivors						Survivors					
	MCS Follow-Up Respondents	Mean Baseline MCS Score	Mean Follow-Up MCS Score	Mean Difference in MCS	Standard Error of the Mean Difference	P	Percent of MCS Follow-Up Respondents[2]	Mean Baseline MCS	Mean Follow-Up MCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Site												
National	544	49.15	48.23	-0.92	0.47	*	89.0%	49.41	49.99	0.59	0.45	ns
PA SGA[3]	559	49.63	47.75	-1.88	0.46	***	88.0	50.21	49.62	-0.59	0.45	ns
GA SGA	552	50.37	48.17	-2.20	0.45	***	87.3	51.09	50.23	-0.86	0.45	ns
WI SGA	727	51.22	50.01	-1.21	0.39	**	91.7	51.44	51.45	0.01	0.38	ns
AZ SGA	538	51.80	49.70	-2.10	0.47	***	89.4	52.14	51.56	-0.57	0.45	ns
WA SGA	553	51.42	50.34	-1.07	0.47	*	89.9	51.73	52.18	0.05	0.43	ns
PGP[4] A	662	49.78	47.48	-2.30	0.43	***	86.7	50.65	49.55	-1.10	0.43	*
PGP B	736	51.50	49.81	-1.69	0.40	***	88.5	51.99	51.87	-0.12	0.37	ns
PGP C	654	52.91	50.52	-2.39	0.41	***	89.9	53.37	52.37	-1.00	0.39	*
PGP D	613	51.65	50.02	-1.62	0.43	***	89.6	52.29	51.89	-0.40	0.42	ns
Sex												
Male	2,523	51.17	49.19	-1.98	0.21	***	87.0	51.81	51.47	-0.34	0.21	ns
Female	3,615	50.85	49.28	-1.57	0.18	***	90.5	51.26	50.89	-0.38	0.17	*
Race												
White	5,779	51.16	49.39	-1.77	0.14	***	88.8	51.68	51.32	-0.36	0.14	**
Black	222	47.66	47.28	-0.38	0.75	ns	91.9	48.17	48.45	0.28	0.75	ns
Asian	30	52.09	46.47	-5.62	2.13	*	96.7	51.86	46.90	-4.96	2.10	*
Hispanic	33	40.33	41.85	1.53	2.18	ns	93.9	39.68	42.36	2.68	2.16	ns
North American Native	9	50.75	45.01	-5.73	5.31	ns	100.0	50.75	45.01	-5.73	5.31	ns
Other	52	50.37	48.62	-1.76	1.64	ns	94.2	50.41	49.51	-0.90	1.66	ns
Unknown	13	55.53	48.71	-6.82	3.28	ns	76.9	57.13	53.12	-4.01	3.42	ns
Original Reason for Entitlement												
Aged	5,313	52.07	50.09	-1.98	0.15	***	88.8	52.64	52.12	-0.52	0.20	***
Disabled	818	43.93	43.71	-0.23	0.42	ns	90.7	44.11	44.70	0.59	0.43	ns
ESRD	3	54.42	60.74	6.32	3.05	ns	100.0	54.42	60.74	6.32	3.05	ns
ESRD and Disabled	4	44.81	46.93	2.12	8.45	ns	75.0	41.67	51.24	9.57	5.65	ns

Table 3-3 (Continued)

Mean Baseline and Follow-up MCS Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (MCS Values Imputed as 34.0 for Those Responding at Baseline But Dying Before Follow-up)

	Decedents & Survivors						Survivors					
	MCS Follow-Up Respondents	Mean Baseline MCS Score	Mean Follow-Up MCS Score	Mean Difference in MCS	Standard Error of the Mean Difference	P	Percent of MCS Follow-Up Respondents[2]	Mean Baseline MCS	Mean Follow-Up MCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Medicaid Status												
No Medicaid	5,565	51.62	49.78	-1.84	0.14	***	89.5	52.12	51.64	-0.48	0.13	***
Medicaid	573	44.79	44.00	-0.79	0.54	ns	84.6	44.91	45.81	0.91	0.56	ns
Age												
Under 65	470	41.55	42.10	0.54	0.55	ns	94.3	41.60	42.59	0.99	0.55	ns
65-74	2,505	52.88	51.72	-1.16	0.20	***	94.1	53.10	52.83	-0.27	0.19	ns
75-84	2,419	51.68	49.48	-2.21	0.22	***	88.3	52.20	51.53	-0.67	0.21	**
85 and Older	744	48.27	44.65	-3.63	0.45	***	71.1	49.62	48.97	-0.64	0.48	ns
Household Income												
Less than $10,000	726	46.32	44.86	-1.46	0.45	**	85.4	46.55	46.72	0.16	0.45	ns
$10,000-$19,999	1,036	50.09	48.60	-1.49	0.34	***	88.2	50.75	50.54	-0.20	0.33	ns
$20,000-$49,999	1,778	52.45	50.93	-1.52	0.24	***	91.6	52.81	52.48	-0.33	0.22	ns
$50,000 or more	595	54.00	52.44	-1.56	0.41	***	92.3	54.27	53.98	-0.29	0.36	ns
Not given	483	49.47	47.44	-2.03	0.51	***	85.1	50.10	49.79	-0.31	0.50	ns
Missing/No Response	1,520	51.41	49.13	-2.28	0.29	***	87.5	51.93	51.13	-0.79	0.28	**
Education												
Not a HS Graduate	1,383	48.62	46.59	-2.03	0.31	***	86.1	49.28	48.62	-0.67	0.31	*
High School Graduate or GED	1,747	51.17	49.98	-1.20	0.25	***	90.6	51.65	51.65	0.00	0.24	ns
Some College or 2 year degree	1,038	52.24	50.50	-1.75	0.32	***	91.2	52.36	52.08	-0.28	0.29	ns
4 year college graduate	431	53.59	51.77	-1.82	0.46	***	91.9	53.73	53.34	-0.39	0.41	ns
More than a 4 year college degree	490	53.65	52.37	-1.28	0.44	**	91.2	54.11	54.14	0.03	0.39	ns
Missing/No Response	1,049	50.23	47.78	-2.45	0.37	***	86.0	50.89	50.02	-0.86	0.38	*
Marital Status												
Not Married	2,122	49.45	48.05	-1.40	0.24	***	91.1	49.90	50.04	0.14	0.24	ns
Married	3,001	52.28	50.56	-1.72	0.18	***	87.6	52.71	52.18	-0.53	0.17	**
Missing/No Response	1,015	50.36	47.84	-2.52	0.38	***	85.9	51.00	50.11	-0.89	0.38	*

Table 3-3 (Continued)

Mean Baseline and Follow-up MCS Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (MCS Values Imputed as 34.0 for Those Responding at Baseline But Dying Before Follow-up)

	Decedents & Survivors						Survivors					
	MCS Follow-Up Respondents	Mean Baseline MCS Score	Mean Follow-Up MCS Score	Mean Difference in MCS	Standard Error of the Mean Difference	P	Percent of MCS Follow-Up Respondents[2]	Mean Baseline MCS	Mean Follow-Up MCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Home Ownership												
Owned by Beneficiary or Family Member	4,042	51.93	50.42	-1.51	0.16	***	91.0	52.31	52.05	-0.26	0.15	ns
Not Owned by Beneficiary or Family Member	924	47.73	46.00	-1.72	0.39	***	84.7	48.21	48.16	-0.05	0.39	ns
Missing/No Response	1,172	50.30	47.73	-2.57	0.35	***	85.7	51.01	50.03	-0.98	0.35	**
Retirement Community												
Yes	847	50.85	48.72	-2.13	0.38	***	86.0	51.40	51.12	-0.28	0.37	ns
No	4,106	51.28	49.79	-1.49	0.16	***	90.8	51.68	51.40	-0.28	0.15	ns
Missing/No Response	1,185	50.04	47.71	-2.33	0.35	***	85.2	50.80	50.09	-0.71	0.35	*
Chronic Conditions												
Hypertenstion or high blood pressure	2,674	50.81	49.28	-1.54	0.20	***	89.5	51.27	51.07	-0.20	0.20	ns
Angina pectoris or coroary artery disease	942	49.42	47.60	-1.82	0.36	***	84.7	50.23	50.05	-0.18	0.36	ns
Congestive heart failure	428	47.69	45.32	-2.37	0.58	***	75.2	48.76	49.05	0.28	0.61	ns
Acute myocardial infarction or heart attack	646	49.49	47.08	-2.41	0.43	***	82.0	50.34	49.94	-0.40	0.42	ns
Other heart conditions	1,302	49.69	47.72	-1.98	0.31	***	85.8	50.17	49.99	-0.18	0.30	ns
Stroke	506	47.55	44.85	-2.70	0.55	***	77.3	48.62	48.05	-0.57	0.59	ns
Emphysema, asthma, or COPD	696	48.42	46.30	-2.12	0.42	***	82.3	49.59	48.94	-0.65	0.44	ns
Crohn's disease, ulcerative colitis, or inflamatory bowel disease	366	47.62	46.50	-1.12	0.54	*	90.4	48.01	47.82	-0.19	0.54	ns
Arthritis of the hip or knee	2,111	50.08	48.61	-1.47	0.23	***	89.0	50.62	50.41	-0.21	0.22	ns
Arthritis of the hand or wrist	1,843	49.79	48.51	-1.27	0.25	***	89.5	50.28	50.22	-0.06	0.24	ns
Sciatica	1,334	49.24	48.03	-1.21	0.30	***	90.9	49.76	49.45	-0.31	0.30	ns
Diabetes, high blood sugar, or sugar in the urine	879	49.56	47.81	-1.75	0.39	***	86.3	50.11	49.99	-0.12	0.38	ns
Any cancer (other than skin cancer)	920	51.05	49.18	-1.87	0.35	***	85.1	51.83	51.83	0.00	0.34	ns

Table 3-3 (Continued)

Mean Baseline and Follow-up MCS Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey (MCS Values Imputed as 34.0 for Those Responding at Baseline But Dying Before Follow-up)

	Decedents & Survivors						Survivors					
	MCS Follow-Up Respondents	Mean Baseline MCS Score	Mean Follow-Up MCS Score	Mean Difference in MCS	Standard Error of the Mean Difference	P	Percent of MCS Follow-Up Respondents[2]	Mean Baseline MCS	Mean Follow-Up MCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Number of Chronic Conditions Reported[5]												
0	1,562	51.23	49.16	-2.07	0.29	***	88.9	51.67	51.06	-0.61	0.28	*
1	901	53.06	51.09	-1.97	0.32	***	92.3	53.41	52.67	-0.74	0.30	*
2	1,043	52.66	51.16	-1.50	0.31	***	91.9	53.05	52.68	-0.37	0.29	ns
3	928	50.87	49.74	-1.14	0.34	***	90.2	51.17	51.30	0.13	0.32	ns
4	684	50.53	48.70	-1.83	0.42	***	86.1	50.96	50.56	-0.40	0.40	ns
5	483	47.46	46.45	-1.01	0.54	ns	79.5	48.13	48.52	0.39	0.55	ns
6	257	48.10	45.60	-2.49	0.72	***	79.0	48.93	48.55	-0.38	0.74	ns
7	146	48.06	45.06	-3.00	0.96	**	75.8	48.30	47.45	-0.85	0.97	ns
8+	134	45.61	43.87	-1.74	0.91	ns	72.0	46.92	46.72	-0.20	1.03	ns
Health in General												
Excellent	284	57.24	56.00	-1.24	0.48	**	97.5	57.18	56.56	-0.62	0.43	ns
Very Good	1,302	56.22	54.30	-1.91	0.25	***	95.5	56.14	52.92	-0.61	0.21	**
Good	2,380	52.90	50.85	-2.04	0.23	***	92.1	52.92	52.31	-0.61	0.21	**
Fair	1,631	46.96	45.30	-1.66	0.29	***	84.9	47.21	47.30	0.09	0.29	ns
Poor	530	38.69	38.24	-0.46	0.52	ns	67.9	38.84	40.23	1.39	0.64	*
Missing/No Response	11	45.05	42.53	-2.52	3.20	ns	81.8	45.03	44.42	-0.60	2.29	ns

NOTES:

[1]This value is calculated by the formula 1-(Mean Follow-Up MCS for Survivors/Mean Baseline MCS for all)

[2]Percentage of baseline respondents alive as of the followup survey.

[3]SGA refers to a small geographic area selected for sampling within the given state.

[4]PGP refers to a primary group practice whose members were selected for sampling. One PGP was selected for each state sampled, excluding Georgia.

[5]A zero in this category could also indicate that the recipient did not respond to any of the questions included in this measure.

*p<0.05, **p<0.01, ***p<0.001; ns=not statistically significant at 0.05 level

Output: joinx02g, joinx02h

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

between the data on survivors only between the tables due to the change to preferring the SF-36 for calculating MCS scores.

Table 3-3 shows a consistent pattern when comparing the results for decedent and survivors versus survivors alone. The change scores are consistently lower when the data imputing the follow-up MCS of 34 for those who died are included in the analysis. Hence dropping those who died from the analysis of MCS change scores, as was done in the HAL approach, will consistently understate declines in mental health status which would be observed when respondents who died between baseline and follow-up are included.

3.4 Summary

In this chapter, we addressed the problem of assigning health status scores to respondents who died between baseline and follow-up. This is one of the key methodological issues in analysis of longitudinal health status outcomes. We first reviewed a range of methods previously used to address this problem by other researchers, which were all found inadequate. We next described our new approach to handling this problem, based on analysis of utility scores, which was developed for this project. We then presented data illustrating the impact of our new method on descriptive statistics for PCS and MCS scores.

We demonstrated how utility scores can be derived using data available in the HOS. We illustrated how utility scores can be used to estimate values for the PCS and

MCS for respondents who died between baseline and follow-up. Our new approach indicated that PCS scores should be imputed as 15 and MCS scores as 34. The PCS value we calculated is much higher than the value of 0 used for this purpose in the MOS. The MCS value we calculated allows decedents to be included in the analysis of MCS change scores, in contrast to the HAL approach, where they were dropped from the analysis.

Applying our imputed values in descriptive analysis of follow-up scores for the PCS and MCS revealed two expected results in comparison with the data presented in the last chapter, which applied the MOS approach. First, the declines over time in mean PCS scores were lessened, reflecting the shift from imputing 0 to imputing 15. Second, the declines in mean MCS scores over time were increased, reflecting the ability to include decedents in that analysis for the first time.

The following chapter applies our new method in multivariate analysis of expected versus actual mean follow-up PCS and MCS scores for each cohort in our FFS sample. Decedents are included in both analyses, using the imputed values described in this chapter.

4 Multivariate Statistical Analysis of Predicted versus Actual Health Status at Two-Year Follow-up

4.1 Overview and Hypotheses

As noted, our goal for the multivariate analysis was to evaluate the performance of four group practices and the health care systems serving FFS beneficiaries in five small geographic areas and in the nation as a whole. Other studies have compared FFS and managed care systems and compared different managed care plans serving M+C beneficiaries (Ware *et al*., 1996; HSAG, 2001). The latter study is described in more detail in Chapter 5.

To evaluate the performance of health care organizations or systems requires a comparison against some standard of reference. One option is to use the prevailing "standard of care" as the reference. This standard should be reflected by average trends in health status (PCS and MCS) found in the beneficiary population as a whole. In national samples of older adults, physical functioning, measured by the PCS, declines over time (Ware *et al*, 1994). Mental functioning, measured by the MCS, shows smaller decreases with age. A similar pattern was also shown in the descriptive statistics for our FFS sample presented in the previous chapter.

For our multivariate analysis, we used our national cohort of beneficiaries to establish "standard care" statistical models for predicting expected physical and mental health status at follow-up. The actual PCS and MCS performance of the individual group

practice and small geographic area cohorts was then compared with the PCS and MCS outcomes predicted by the standard care model developed using the national data. Our standard care models predict follow-up PCS and MCS scores using the baseline PCS or MCS score and other beneficiary and survey characteristics that increase or decrease the slope of the trajectory over the two-year follow-up interval.

To provide a comprehensive assessment, the prediction models also need to account for beneficiaries who completed the survey at baseline, but died before they were able to complete the follow-up survey two years later. As discussed in the last chapter, this is a well-known problem for measuring changes in health status using longitudinal surveys such as the HOS (Diehr, 1995). It is especially salient for studies involving Medicare beneficiaries, who have a higher death rate than the general population. In the last chapter we described a new approach for imputing PCS and MCS scores for respondents who died between baseline and follow-up, using HALex utility scores calculated for the same respondents. That approach is applied here for multivariate prediction of follow-up scores for the first time.

4.2 Statistical Models for Predicting Follow-up Health Status

Our statistical analysis of the raw PCS and MCS scores compared actual follow-up scores with follow-up scores predicted using a regression model which included the baseline scores and other variables. This method for analyzing follow-up scores is known as the "regressor variable" approach, and is commonly used in psychometric

analysis of longitudinal data for scale scores with two waves of data collection (Menard, 1991; Taris, 2000; de Vaus, 2001).

Our analysis involved three steps: (i) developing the regression model for predicting follow-up scores using our national cohort sample; (ii) applying the regression equation to estimate predicted (expected) mean PCS and MCS follow-up scores for the PGP and SGA cohorts; and (iii) evaluating whether the differences between the expected and actual follow-up values for the PGP and SGA cohorts were statistically significant. The first step is described in Section 4.3. Steps two and three are described in Section 4.4 below.

As noted, the predictive models are designed to provide a "standard care" benchmark for assessing the relative PCS and MCS health status of groups of beneficiaries, such as the PGPs or an SGAs in our study. The models produce estimates of the mean PCS and MCS scores that would be expected after two years for a group of beneficiaries with a given set of characteristics at baseline who received the usual types of care delivered by the Medicare FFS system. The explanatory variables in the models adjust the predicted two-year outcomes for the case-mix characteristics of the groups under study.

The standard care regression models were developed using only the national beneficiary cohort sample, under the assumption that the effects of standard care are represented by component score trends averaged over beneficiaries receiving FFS services all across the country. This limited the sample size available for the regression

models, but enabled the predictive models to be developed on a sample independent of the GP and SGA cohorts which were evaluated using the predictive models.

Separate models were derived for physical functioning and mental functioning. The dependent variables in these models are the two-year follow-up PCS and MCS scores computed from either the SF-36 or the SF-12. As noted, the SF-36 was the preferred method for calculating PCS and MCS scores. The SF-12 was only used when a respondent was unwilling to answer all of the questions in the SF-36.

PCS and MCS scores for deaths occurring during the follow-up period were imputed based on the methodology described in the last chapter. The imputed values for death were 15 points for the PCS and 34 points for the MCS. Follow-up scores were available for 544 beneficiaries in the national cohort, including 60 baseline respondents who died before follow-up and had PCS and MCS scores imputed. This sample size was sufficient to develop the predictive models.

Table 3-2 in the previous chapter shows that the mean PCS for this national cohort was 38.3 at baseline and 34.1 at follow-up. Table 3-3 shows that the mean MCS was 49.2 at baseline and 48.2 at follow-up. Thus both PCS and MCS declined for our analytic group of 544, although the decline was slightly larger for the PCS.

4.3 Specification of the Standard Care Model for Predicting Follow-up Health Status

A range of potential explanatory variables measured at baseline and follow-up were tested hierarchically for each model. The first explanatory variable entered in the model was the baseline PCS or MCS. Like most longitudinal data, follow-up FFS outcomes were strongly associated with baseline levels of the component scores. We tested for curvilinear relationships using squared terms for the baseline values, but found that the trends were linear.

The second variable to be tested was the length of the time between the baseline and follow-up survey administrations. These intervals ranged from 20 to 28 months, with nearly half of the follow-up surveys completed at 24 months. As expected, this variable was found nonsignificant, and was dropped from the models.

The third set of explanatory variables consisted of three factors – age group, Medicaid status, and an activity limitation score – that were expected to influence the trajectory of change over time. For activity limitations, a sum of activities of daily living (ADL) limitations variable was considered. However, since those data were also used to define values assigned for the dependent variables (PCS and MCS) for respondents who died between baseline and follow-up, we decided to focus on the other options to avoid bias in the regressions which would result if they were also used as independent variables. Since the age group variables were highly significant in both models, we opted for those instead. The Medicaid variable was found to be nonsignificant.

Fourth, we tested a series of dummy variables characterizing survey administration. These included use of proxy respondents, mode of administration (telephone versus mail), and scoring using the SF-12 rather than the SF-36. For each of these variables, separate indicators were coded for the baseline and follow-up administrations. We hypothesized that any administration effects should be of opposite signs at baseline and follow-up, so that they would offset one another if they were present at both time points.

Mode of administration has been found to affect reported PCS and MCS scores in several studies. For example, the Health Assessment Laboratory (HAL) subtracts 1.9 points from PCS scores and 4.5 points from MCS scores in its analysis if the mode of administration was by phone versus by mail (Rogers *et al*., 2000). This adjustment was based on a comparison of scores for beneficiaries eligible for both Medicare and Veterans Administration (VA) health care services, who completed both the HOS survey and a VA-administered SF-36 using phone for one and mail for the other. A somewhat smaller effect was found in an earlier HAL study, in which a randomized trial of phone versus mail administration of the SF-36 in the general U.S. population found no difference in PCS scores and a mean score 2.4 points higher for the MCS (Ware *et al*., 1994).

For these variables, we found both telephone mode of administration and use of proxy respondents to have significant effects for predicting follow-up scores. Variables were included for both baseline and follow-up involvement of these factors to capture the effects of switching between them in different rounds of data collection. Alternate model

specifications were also tested with variables for use of the SF-12 defined at baseline and follow-up. However, the SF-12 and phone administration variables were found to be strongly collinear, since the SF-12 was most often collected by telephone after a respondent had failed to return a mail survey. As a result, the SF-12 variables were dropped from the models.

Finally, we tested a range of dummy variables for specific medical conditions. Included in this set were chronic diseases, vision and hearing impairments, and a depression screening question. While health status trajectories over time are strongly influenced by baseline levels of the PCS and MCS, these trajectories may decline more steeply for people with serious chronic health problems. We therefore tested the independent impact of selected self-reported health conditions on follow-up scores (after first adjusting for baseline levels of the component scores and age-related trends, and the survey administration variables).

Chronic conditions known to affect PCS and MCS scores were identified using published results from the Medical Outcomes Study and general U.S. population studies reported in the SF-36 and SF-12 manuals (Ware, Kosinski & Keller, 1994; 1995). For the PCS, most of the chronic diseases studied in the HOS survey were found to have significant effects in those earlier studies. As a result, they were all tested for significance in our statistical model. In the HOS, respondents indicate whether a doctor has even told them that they have a specific disease. The specific cancer questions were combined with the "any cancer" question. Physical symptom items, such as chest pain,

generally overlap with the chronic disease items in the HOS, such as for angina or coronary artery disease. As a result, we focused on testing the chronic disease variables for the PCS model.

A number of the chronic disease variables were found significant in a series of initial screening regressions, in which each disease variable was tested individually with the baseline PCS or MCS score, age categories, and survey administration variables as the only other covariates. Variables were retained in subsequent specifications if their effect size (coefficient) was greater than 2.0, which is often considered a minimum threshold for clinical or policy significance for changes in PCS or MCS scores. Their effect on the R^2 was also considered. Since these regression models were primarily intended for prediction, and not for explanation, the statistical significance of the independent variables was accorded less consideration. A sample of the preliminary models and alternate specifications tested is included in the Appendix, Table 5 for the PCS models and Table 6 for the MCS models.

At the conclusion of the process, two chronic disease variables were retained in the PCS model. They were the variables for angina or coronary artery disease and for congestive heart failure.

For the MCS, asthma and chronic lung disease were the only HOS survey chronic disease items found to have significant effects in the earlier studies summarized in the SF-36 and SF-12 manuals (Ware *et al*., 1994;1995). Several other diseases, such as depression, and several disabilities, such as visual impairment, were also found to affect

the MCS in those earlier studies, but they are included in the HOS as symptom items, rather than as chronic disease items. As a result, a mix of chronic disease and symptom items were tested for significance in the statistical model for the MCS.

Fewer of the chronic disease or symptom variables were found significant for the MCS in the initial screening regressions, in comparison to the PCS. Depression has been the variable found to affect MCS scores most strongly in previous research. Self-reported depression in the previous years had a significant effect in our initial screening model. It was retained in the final model, with the baseline, age, and survey administration variables, despite having a significance level below the five percent level, since its effect size was greater than 2.0.

It should be noted that there are ongoing debates in the literature over the best statistical methods to use in analysis of longitudinal data for scale scores such as the PCS and MCS. The Health Assessment Laboratory group takes a different approach in their analysis of HOS survey outcomes for M+C beneficiaries. They argue against the use of baseline scores and self-reported diseases and symptoms as covariates, since they are measured or reported with error and thus may bias coefficients.

However, if the objective is to identify conditions that predict health trajectories over time, then there are at least two reasons to believe that patient reports are more valuable than physician diagnoses. First, direct comparisons of physician-evaluated and self-reported health conditions have consistently found that patient reports are better predictors of subsequent morbidity (Ferraro and Su, 2000).

Second, analyses of the SF-36 indicate that patient-reported conditions explain more of the variation in component scores than physician diagnoses. Tracer and comorbid conditions were diagnosed by physicians in the Medical Outcomes Study; a similar set of chronic conditions was reported by adults participating in the National Survey of Functional Health Status. Analyses of these two surveys show that the patient-reported conditions explained a greater percentage of the variation in PCS scores (47 percent vs. 34 percent) and a slightly higher percentage of the variance in MCS scores (30 percent vs. 29 percent) than did the physician reports (Ware *et al*., 1994; Tables 7.11 and 7.12). Self-reports may provide a better indication of the severity of a particular condition.

Tables 4-1 and 4-2 include the final regression models selected after testing the range of candidate variables discussed above. Table 4-3 includes the mean values for each of the independent variables for the national cohort used to estimate these models in its first column. Each model was estimated using ordinary least squares. The R^2 statistics indicate that the PCS model explains about 47 percent of the variance in follow-up PCS scores, while the MCS model explains about 39 percent of the variance in follow-up MCS scores.

As noted, follow-up FFS outcomes were strongly associated with baseline levels. Baseline PCS and MCS were the most highly significant predictors of follow-up scores, with *t* statistics of 17.1 and 12.7, in their respective models. The regression coefficients for the baseline levels were 0.66 for the PCS and 0.54 for the MCS. The fact that these

Table 4-1

Regression Model for Predicting PCS Scores at Follow-up

Variable	Coefficient	*t* statistic	*p* value
Intercept	11.47	6.32	0.00
Baseline PCS	0.66	17.12	0.00
Age < 65	-0.93	-0.58	0.57
Age 75-84	-2.36	-2.39	0.02
Age 85+	-6.63	-4.79	0.00
Baseline Phone[1]	-4.13	-3.21	0.00
Follow-up Phone[2]	4.62	2.39	0.02
Baseline Proxy[3]	-0.79	-0.59	0.56
Follow-up Proxy[4]	2.02	1.68	0.09
Angina/CAD	-3.65	-2.88	0.00
Congestive Heart Failure	-2.29	-1.21	0.23

$R^2 = 0.4719$

[1]Telephone used for survey administration at baseline (survey was not completed solely by mail).
[2]Telephone used for survey administration at follow-up (survey was not completed solely by mail).
[3]Proxy completed the survey on behalf of the beneficiary at baseline.
[4]Proxy completed the survey on behalf of the beneficiary at follow-up.

OUTPUT: joinx02a

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Table 4-2

Regression Model for Predicting MCS Scores at Follow-up

Variable	Coefficient	*t* statistic	*p* value
Intercept	23.57	9.40	0.00
Baseline MCS	0.54	12.68	0.00
Age < 65	-4.59	-2.81	0.01
Age 75-84	-1.88	-1.95	0.05
Age 85+	-2.99	-2.26	0.02
Baseline Phone[1]	-3.52	-2.83	0.00
Follow-up Phone[2]	7.90	4.21	0.00
Baseline Proxy[3]	-3.40	-2.63	0.01
Follow-up Proxy[4]	1.42	1.22	0.22
Depression	-2.01	-1.42	0.16
$R^2 = 0.3859$			

[1]Telephone used for survey administration at baseline (survey was not completed solely by mail).
[2]Telephone used for survey administration at follow-up (survey was not completed solely by mail).
[3]Proxy completed the survey on behalf of the beneficiary at baseline.
[4]Proxy completed the survey on behalf of the beneficiary at follow-up.

OUTPUT: joinx02

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Table 4-3

Mean Values of Independent Variables for PCS & MCS Prediction Models by Cohort

Independent Variables	National Cohort Mean Values	Pennsylvania SGA Mean Values	Georgia SGA Mean Values	Wisconsin SGA Mean Values	Arizona SGA Mean Values	Washington SGA Mean Values
Baseline PCS (PCS model only)	38.28	38.99	37.48	37.77	39.76	39.53
Baseline MCS (MCS model only)	49.15	49.63	50.37	51.22	51.80	51.42
Age < 65	0.09	0.08	0.12	0.09	0.07	0.08
Age 75-84	0.35	0.41	0.38	0.39	0.41	0.39
Age 85+	0.14	0.12	0.12	0.12	0.12	0.12
Baseline Phone	0.14	0.16	0.18	0.13	0.11	0.10
Follow-up Phone	0.06	0.06	0.08	0.06	0.07	0.05
Baseline Proxy	0.15	0.18	0.19	0.13	0.12	0.13
Follow-up Proxy	0.19	0.18	0.21	0.17	0.12	0.14
Angina/CAD (PCS model only)	0.15	0.15	0.16	0.15	0.13	0.14
CHF (PCS model only)	0.06	0.06	0.05	0.08	0.07	0.07
Depression (MCS model only)	0.14	0.16	0.17	0.13	0.11	0.13
Sample Size	544	559	552	727	538	553

Table 4-3 (continued)

Mean Values of Independent Variables for PCS & MCS Prediction Models by Cohort

Independent Variables	PGP A Mean Values	PGP B Mean Values	PGP C Mean Values	PGP D Mean Values
Baseline PCS (PCS model only)	37.53	37.77	39.03	39.09
Baseline MCS (MCS model only)	49.78	51.50	52.91	51.65
Age < 65	0.11	0.08	0.02	0.03
Age 75-84	0.38	0.39	0.42	0.40
Age 85+	0.12	0.13	0.10	0.12
Baseline Phone	0.11	0.11	0.10	0.07
Follow-up Phone	0.06	0.04	0.03	0.06
Baseline Proxy	0.17	0.15	0.07	0.11
Follow-up Proxy	0.17	0.14	0.08	0.12
Angina/CAD (PCS model only)	0.19	0.13	0.19	0.13
CHF (PCS model only)	0.09	0.07	0.08	0.06
Depression (MCS model only)	0.14	0.13	0.09	0.11
Sample Size	662	736	654	613

OUTPUT: joinx02d

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

coefficients are both less than 1.0 indicates that scores tend to fall over time, especially for those on the higher end of the distributions. Follow-up scores are predicted to decline by less than one point for each point of the baseline component score.

The age category variables are dummy variables, coded as 1 if the respondent's age fell into a particular category, and 0 if it did not. The category for ages 65-74 was the omitted category, so the coefficients for these variables show effects relative to that reference category. For the PCS model, coefficients were significant and negative for the two older age groups, indicating that follow-up scores declined more rapidly with age even after adjusting for baseline component scores. The decline of –6.63 points for the Age 85+ groups is a large effect.

For the MCS model, all three of the age category coefficients were negative. The oldest (85+) age category was again highly significant. However, the age 75-84 category was barely significant at the 5 percent level and its coefficient falls slightly below the 2 point threshold for clinical or policy significance. This indicates that mental health also declines with age, especially in the oldest group, even after adjusting for baseline scores, as with the PCS.

However, for the MCS, the younger age category also predicts lower follow-up scores relative to the 65-74 reference age category. It is highly significant and predicts a decline of 4.6 points, indicating this effect has clinical or policy significance. The strength of this effect is likely due to the nature of the under-65 Medicare population, which consists primarily of people with disabilities. Presumably, the higher rate of

mental disabilities among this population is driving the size and significance of this coefficient in the MCS model.

Of the survey administration variables, telephone administration and at least one proxy status variable had effects >2.0 on follow-up PCS and MCS scores. These were both coded as dummy variables, with 1 representing use of the telephone for administration of the survey or a proxy completing the survey for the beneficiary, and 0 representing the other option.

Each of the chronic disease variables was also coded as a dummy variable, with 1 indicating the reported presence of the disease and 0 indicating no reported presence of the disease. In each model the effects were negative, as expected. The effect for angina or CAD was highly significant in the PCS model, at more than the one percent level, and its coefficient is well above the threshold of 2.0. The variable for CHF had an effect size >2.0 in the PCS model, so it was retained despite its lack of statistical significance at the 5 percent level.

The effect for depression was weaker in the MCS model; its coefficient was just above the 2.0 level. The weaker chronic disease effect found for this MCS model may be due to the fact that depression was measured in the HOS survey as a symptom. Future HOS surveys may wish to include depression as a chronic disease question similar to the angina or CAD question, and not only as a symptom variable.

The SF-36 does include some questions related to depression, in its mental health scale (the mental health scale is a depression and anxiety scale). For example, one of the

five questions in that scale asks the respondent "have you felt downhearted and blue" in the past four weeks, with six possible responses ranging from all of the time to none of the time. However, we felt it was important to include a depression variable as a predictor in the MCS model since, as noted, the literature has consistently shown that depression is the disease which most affects MCS measure of mental health status (Ware *et al*., 1994; 1995). The HOS symptom question we selected as an independent variable for the MCS regression model for this study is intended to identify more severe cases. It asks, "In the past year, have you felt depressed or sad much of the time?"

A factor which may have reduced the effects of some of the chronic disease variables in the regression models was the method of coding them. Due to moderately high levels of missing values for most of the chronic disease and symptom variables, missing values were set to 0 under the assumption that beneficiaries did not respond to diseases that did not apply to them. This probably understates the prevalence of these conditions in this sample. Future HOS surveys should consider undertaking additional efforts to boost response rates for these items. Missing values were also coded as zero for the proxy variables and for the follow-up telephone administration variable.

4.4 Comparisons of Expected and Actual Two-Year Follow-up in Health Status by Cohort

After finalizing the standard care models using the national cohort data, we applied the equations to each of the GP and SGA cohorts. We calculated predicted

follow-up scores for the PCS and MCS for each respondent using the constant term and coefficients in the predictive model. We then calculated mean predicted follow-up PCS and MCS scores for each GP and SGA cohort.

Table 4-3 presents the mean values for each cohort for each independent variable used in the standard care models. Baseline PCS and MCS are the continuous variables used in the models. These data show that the cohorts do not differ greatly on the mean values of either variable. The range for the mean baseline PCS is from about 37.5 (Georgia SGA) to 39.8 (Arizona SGA). The range for the mean baseline MCS is from 49.2 (National cohort) to 52.9 (PGP C).

The remaining independent variables in Table 4-3 are all dummy variables, so the means for each cohort are expressed as values between 0 and 1. These data represent the proportion of the respondents with each attribute represented by these variables.

These variables also show a limited range of variation, although there are some cohorts which stand out for some variables. PGP C has the lowest value for the Age 85+ group with 10 percent. All of the other cohorts fall within a range of 12 to 14 percent on that variable. Baseline phone administration was lowest for the PGP D, at 7 percent. The other cohorts all fall within a range of from 10 to 18 percent. Use of proxy respondents at baseline was highest for the Georgia SGA, at 0.19. It was lowest for PGP C, at 7 percent, which is consistent with its smaller percentage of respondents in the oldest, 85+ age group.

Chronic diseases were less prevalent in PGP D. It had the second-lowest proportions for all three variables, with Depression at 11 percent, Angina/CAD at 13 percent, and CHF at 6 percent. Higher values for these variables were more scattered. The PGP A scored highest for Angina/CAD, at 0.19 (with PGP C), and for CHF at 0.09. The Georgia SGA was highest for depression, at 0.17.

The final step in the multivariate statistical analysis involved comparing the actual mean follow-up PCS and MCS values for each of the GP and SGA cohorts with their predicted mean follow-up values. GPs and SGAs performing better or worse than expected could then be identified.

To test the statistical significance of the differences between these actual and expected mean values we used a one-sample *t* test (Jacobsen, 1976). This method is used in situations in which mean values from a sample of respondents are being compared to a criterion value. For our analysis, the actual mean follow-up PCS and MCS values were calculated from the respondents in each cohort, with the expected values calculated from the prediction models as the criterion.

The one sample *t* test was calculated in four steps. First, the difference scores were calculated by subtracting the expected mean follow-up scores from the actual mean follow-up scores for each GP and SGA cohort, for both the PCS and MCS. Second, a standard error was calculated by dividing the standard deviation of the actual PCS and MCS scores for each cohort by the square root of the sample size. Third, *t* values were calculated by dividing the difference scores by the standard error. Fourth, *p* values were

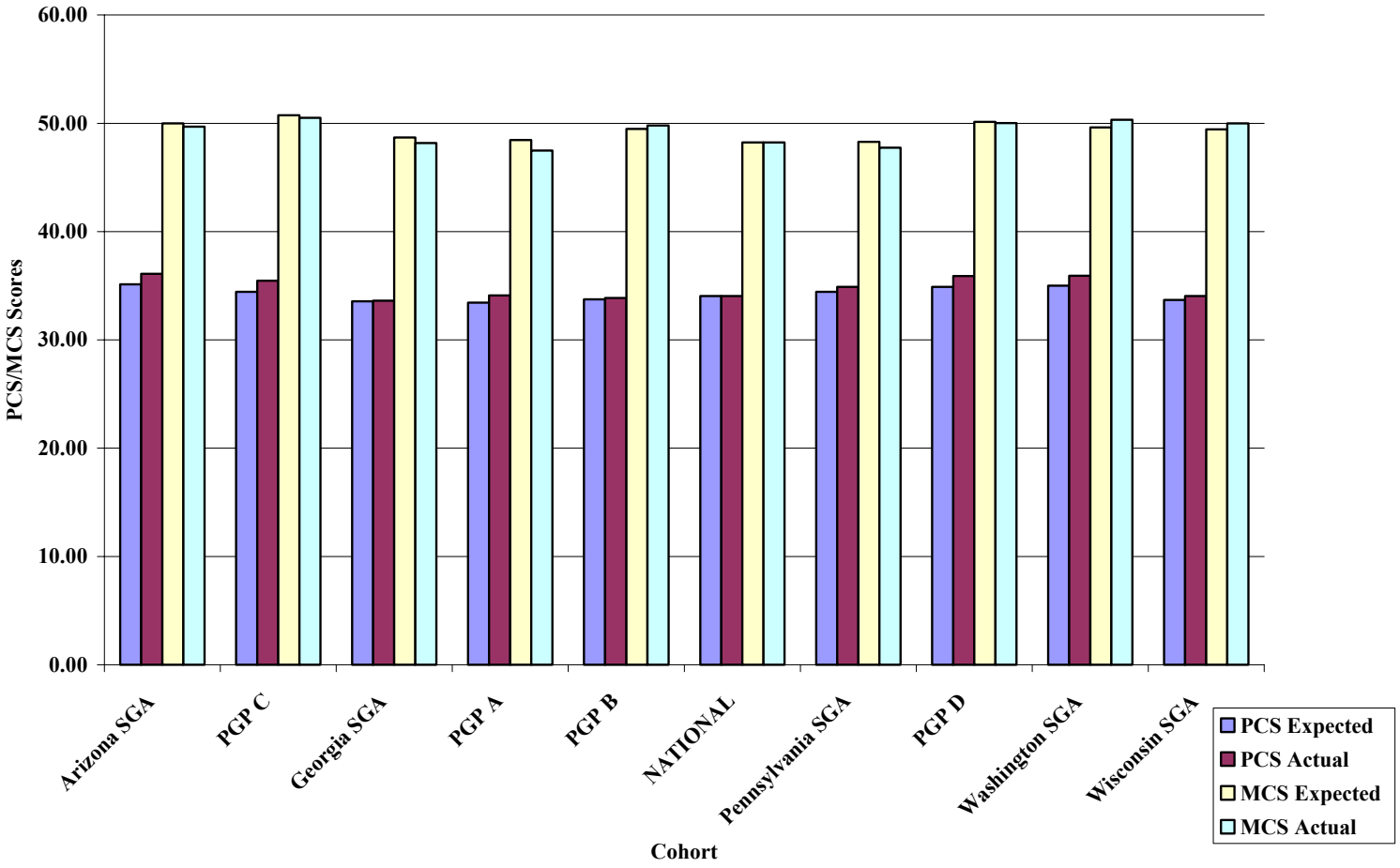
identified for each of the *t* statistics using a standard table of statistical significance for *t* values at different sample sizes.

The *p* values identify whether or not the differences found between the actual and predicted mean PCS and MCS values for each cohort are statistically different from zero. They enabled us to identify which differences might be due to chance alone. However, since our sample sizes were fairly large for detecting statistical significance, there is a danger that even very small differences in means will be found significant. As a result, we supplemented the statistical assessment with consideration of the clinical or policy significance of the size of differences in actual versus predicted mean values found for any given cohort. As noted, for the PCS and MCS, the usual standard is that a difference of 2.0 points in mean scores for groups of respondents is the minimum level which is expected to have clinical or policy significance.

Expected mean PCS and MCS follow-up scores for each FFS cohort were computed by applying the coefficients in the standard care regression equations from Tables 4-1 and 4-2 to the beneficiaries in each cohort, and then averaging across beneficiaries. The expected means were then compared with the actual mean values for PCS and MCS for each cohort. Figure 4-1 depicts the results of the statistical analysis graphically. It plots the expected and actual PCS and MCS scores for each cohort in a bar graph. The bars representing the actual scores for both the PCS and MCS are all about the same height as the bars representing the expected scores. Only small deviations were found from the expected levels.

Figure 4-1

Mean Expected Versus Actual Follow-Up PCS and MCS Scores by Cohort



SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Study.

Table 4-4 presents the detailed statistical data underlying the results presented graphically in Figure 4-1. It presents the mean expected scores and compares them with the mean actual follow-up scores for each of the PGPs and SGAs. In general, the FFS cohorts were similar in terms of mean values for the predictor variables in the regression equations (Table 4-3), so that the mean expected follow-up PCS and MCS scores were also similar across the different cohorts.

The national cohort is also included in Table 4-4 for reference. Actual and predicted scores were identical for this cohort, as expected, since it was the data source used to develop the regression model.

The largest negative difference found in Table 4-4 was for –0.98 points, for MCS scores for PGP A. Using the one-sample t-test, this is also the only statistically significant negative difference among the comparisons in Table 4-4 ($t = -2.14$, $p < 0.05$). These data indicate that this PGP performed worse than expected for MCS outcomes. However, the effect size is small, and the absolute difference is less than the commonly used clinical and policy threshold of 2.0 points.

The largest positive difference found in Table 4-4 was for 1.04 points, for PCS scores for PGP C. This was the only statistically significant positive difference among the comparisons in Table 4-4 ($t = 1.99$, $p < 0.05$). These data indicate that this PGP performed better than expected for PCS outcomes. However, the effect size is again small, with the difference less than the commonly used clinical and policy threshold of 2.0 points.

Table 4-4

Comparisons of Actual and Expected Mean PCS and MCS Follow-up Scores by Cohort

Cohort	Variable at Follow-Up	N	Mean	Std. Dev.	Std. Err.	*t* Statistic	p value for Ho: Diff. = 0
National	Actual PCS Score	544	34.06	13.60			
	Expected PCS Score	544	34.06	9.34			
	Diff. = Actual PCS - Expected PCS	544	0.00	9.88	0.58	0.00	**p > .20**
	Actual MCS Score	544	48.23	12.23			
	Expected MCS Score	544	48.23	7.60			
	Diff. = Actual MCS - Expected MCS	544	0.00	9.58	0.52	0.00	**p > .20**
Pennsylvania SGA	Actual PCS Score	559	34.91	13.54			
	Expected PCS Score	559	34.43	8.63			
	Diff. = Actual PCS - Expected PCS	559	0.48	9.53	0.57	0.84	**p > .20**
	Actual MCS Score	559	47.75	12.01			
	Expected MCS Score	559	48.27	7.58			
	Diff. = Actual MCS - Expected MCS	559	-0.52	9.86	0.51	-1.03	**p > .20**
Georgia SGA	Actual PCS Score	552	33.61	13.57			
	Expected PCS Score	552	33.56	9.81			
	Diff. = Actual PCS - Expected PCS	552	0.05	9.83	0.58	0.09	**p > .20**
	Actual MCS Score	552	48.17	12.20			
	Expected MCS Score	552	48.69	7.80			
	Diff. = Actual MCS - Expected MCS	552	-0.52	9.81	0.52	-1.00	**p > .20**
Wisconsin SGA	Actual PCS Score	727	34.05	13.01			
	Expected PCS Score	727	33.69	8.99			
	Diff. = Actual PCS - Expected PCS	727	0.36	9.24	0.48	0.75	**p > .20**
	Actual MCS Score	727	50.01	11.49			
	Expected MCS Score	727	49.43	7.22			
	Diff. = Actual MCS - Expected MCS	727	0.58	9.53	0.43	1.36	**p < .20**

Table 4-4 (continued)

Comparisons of Actual and Expected Mean PCS and MCS Follow-up Scores by Cohort

Cohort	Variable at Follow-Up	N	Mean	Std. Dev.	Std. Err.	*t* Statistic	p value for Ho: Diff. = 0
Arizona SGA	Actual PCS Score	538	36.10	13.99			
	Expected PCS Score	538	35.12	9.67			
	Diff. = Actual PCS - Expected PCS	538	0.98	9.87	0.60	1.63	**p < .20**
	Actual MCS Score	538	49.70	11.63			
	Expected MCS Score	538	50.00	6.97			
	Diff. = Actual MCS - Expected MCS	538	-0.29	9.90	0.50	-0.59	**p > .20**
Washingon SGA	Actual PCS Score	553	35.92	13.66			
	Expected PCS Score	553	35.00	9.36			
	Diff. = Actual PCS - Expected PCS	553	0.92	9.55	0.58	1.58	**p < .20**
	Actual MCS Score	553	50.34	11.38			
	Expected MCS Score	553	49.62	6.98			
	Diff. = Actual MCS - Expected MCS	553	0.72	9.60	0.48	1.49	**p < .20**
PGP A	Actual PCS Score	662	34.11	13.60			
	Expected PCS Score	662	33.43	9.67			
	Diff. = Actual PCS - Expected PCS	662	0.68	9.54	0.53	1.28	**p > .20**
	Actual MCS Score	662	47.48	11.71			
	Expected MCS Score	662	48.46	7.51			
	Diff. = Actual MCS - Expected MCS	662	-0.98	9.60	0.45	-2.14	**p < .05****
PGP B	Actual PCS Score	736	33.87	13.35			
	Expected PCS Score	736	33.74	9.57			
	Diff. = Actual PCS - Expected PCS	736	0.14	9.14	0.49	0.27	**p > .20**
	Actual MCS Score	736	49.81	10.93			
	Expected MCS Score	736	49.48	6.92			
	Diff. = Actual MCS - Expected MCS	736	0.32	9.45	0.40	0.80	**p > .20**

Table 4-4 (continued)

Comparisons of Actual and Expected Mean PCS and MCS Follow-up Scores by Cohort

Cohort	Variable at Follow-Up	N	Mean	Std. Dev.	Std. Err.	*t* Statistic	p value for Ho: Diff. = 0
PGP C	Actual PCS Score	654	35.47	13.33			
	Expected PCS Score	654	34.43	9.24			
	Diff. = Actual PCS - Expected PCS	654	1.04	9.50	0.52	1.99	**p < .05****
	Actual MCS Score	654	50.52	10.54			
	Expected MCS Score	654	50.75	6.13			
	Diff. = Actual MCS - Expected MCS	654	-0.23	9.41	0.41	-0.55	**p > .20**
PGP D	Actual PCS Score	613	35.91	13.38			
	Expected PCS Score	613	34.90	9.17			
	Diff. = Actual PCS - Expected PCS	613	1.01	9.39	0.54	1.87	**p < .10***
	Actual MCS Score	613	50.02	11.39			
	Expected MCS Score	613	50.14	7.05			
	Diff. = Actual MCS - Expected MCS	613	-0.12	9.45	0.46	-0.25	**p > .20**

*p < .10
**p < .05
***p < .01

OUTPUT: joinx02k

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

One other difference was significant at the 10 percent level. This was for PGP D, with a difference of +1.01 points on the PCS scale. This difference just failed to achieve statistical significance at the 5 percent level, however, with a *t* statistic of 1.87. Moreover, the effect size is again below the threshold of 2.0.

4.5 Summary

In this chapter, we compared expected versus actual follow-up PCS and MCS scores, using a multivariate statistical model. This constitutes our basic test of the ability of health status scores to discriminate between high- and low-performing PGPs and SGAs.

We first developed statistical models for predicting expected PCS and MCS follow-up scores, using our national cohort to represent a "standard care" model. We next applied those statistical models to calculate mean expected PCS and MCS scores for each cohort, and compared them with actual mean values achieved by each cohort.

We found statistically significant differences between expected and actual mean PCS scores for two PGPs. We also found significant differences between expected and actual mean MCS scores for one PGP. None of the differences were greater than 2.0 points, however, which is the usual standard for considering differences in PCS and MCS scores to have clinical or policy significance.

We found no statistically significant differences between expected and actual scores among the SGA cohorts. This may be due to the diversity of FFS providers and beneficiaries that inhabit a given SGA.

The lack of clinically significant differences in our results indicate that health status scores, at least as measured at the aggregate level through the PCS and MCS, may have limited value for identifying high- and low-performing PGPs or SGAs. Ware and colleagues have suggested a longer follow-up period may be required to identify differences which have clinical or policy significance; they indicate four years may be more appropriate than the two year period used in this study (Ware *et al*., 1996; 2001). That approach may be worth considering.

Alternatively, these results may indicate that other approaches should be pursued to evaluate the performance of PGPs and FFS providers serving Medicare beneficiaries in a given SGA. It may be, for example, that a focus on quality of care for specific diseases or specific population subgroups will better discriminate between PGPs and SGAs.

5 Comparison to the Statistical Methods used by the Health Assessment Laboratory for Analysis of Two-Year Follow-up Health Status for Medicare+Choice Beneficiaries

The Health Assessment Laboratory (HAL) group took a different approach for analyzing two-year changes in PCS and MCS in its study of M+C beneficiaries. We replicated their methodology to provide a comparison to our statistical methods (described in the last two chapters). This comparison of methodologies will enable a better assessment of future options for statistical analysis of longitudinal FFS health status data.

The method HAL used in its M+C study is described in several recent reports (HSAG, 2001; Rogers *et al*., 2000; Rogers *et al*., 2001; Ware *et al*., 2001). An earlier version of the HAL methodology was applied in the MOS (Ware *et al*., 1996). Our focus in this chapter is on the more recent, 2001 version of the HAL methods.

5.1 Assigning Respondents to Change Categories

As noted, the HAL approach begins by sorting the respondents into three mutually exclusive categories, depending on whether or not they were "better", "the same", or "worse" when comparing their baseline PCS and MCS scores to their follow-up scores. Respondents were assigned to these three groups based on the standard error of measurement (SEM), a measure of the reliability of each component score.

Respondents were assigned to those categories based on whether or not the observed changes fell within ±1.41 times the measurement error for a difference of two values. The cut-off for defining a change in PCS as "better" was 5.66 points. The MCS cut-off was 6.72 points. Negative changes of the same magnitudes were used to define a change in PCS or MCS as "worse."

In our replication of the HAL methodology, we used the same cut-off points as their 2001 HOS study, since they also analyzed a Medicare sample and our primary goal was to compare the assessment of GPs and SGAs provided by their statistical methods with the assessment provided by ours. We thus held the cut-off points and other parameters in their predictive models constant to ensure that differences in the results of the two analyses were based on the statistical methods used, and not on differences in the parameters.

This means, for example, for the PCS cut-offs for our FFS data, that only changes between baseline and follow-up of ≥5.66 resulted in assignment of the respondent to the "better" category for analysis using the HAL methodology. Anything between -5.66 and +5.66 resulted in assignment to the "same" category. Anything ≤-5.66 resulted in assignment to the "worse" category. Similar calculations were used to assign respondents to the three categories for changes in MCS scores for our FFS data, although the cut-off point was based on the 6.72 figure.

5.2 Statistical Methods for Analyzing Two-Year Changes in Health Status

The methods used by the HAL group in their 2001 HOS study are described in their most recent reports and in a SAS software program provided to us by them (HSAG, 2001; Sinclair & Gardek, 2001; Rogers *et al*., 2001). For our study, we replicated the HAL methodology by applying the HAL SAS software to our FFS HOS data. The steps in the analysis were as follows:

1. The primary outcomes for comparisons were "alive and PCS same or better", "MCS same or better", "alive and PCS better", and "MCS better". In essence, two of the three possible outcomes (better/same/worse) were merged for each analysis in order to dichotomize the outcome categories (either better and same versus worse or better versus same and worse).

2. The approach for analyzing PCS results included two logistic regression models, one for death and one for PCS scores. This enabled the expected probability of "alive and PCS same or better" to be calculated by combining two models in a conditional probability relationship:

   P(alive and PCS same or better) = P(alive) x P(PCS same or better given alive)

   The HAL group argues that there are several reasons for this approach to the analysis. First, deaths between baseline and follow-up are handled in a straightforward way, by implicitly including them in the "worse" outcome category. Second, the logistic regression equations can be separated for death (a model which includes extensive case-mix control variables and is estimated using all baseline respondents in the sample) and for the expected change in PCS outcome categories (a model estimated using demographic characteristics for respondents with both baseline and follow-up data). Third, the result is a percentage-based outcome, which may be easier to understand and interpret.

For MCS results, respondents who died between baseline and follow-up were simply excluded from the analysis. Thus the death model was not included in the MCS equation. The analysis focused on estimating a one-part model instead of the two-part model used for the PCS analysis:

$$P(\text{MCS same or better})$$

For simplicity, the rest of this description of the HAL methodology focuses on the PCS methods, but the same general approach applies to the MCS outcomes (after dropping the probability of death model).

3. Next, the expected death rate was estimated for each beneficiary within a FFS cohort using the entire baseline sample. This logistic regression model included demographics, chronic disease variables, and other case-mix control data.

4. A model predicting the expected "PCS same or better" rate was next estimated, using only demographic data since HAL argues that only strictly exogenous variables should be used for this analysis. This model was estimated using beneficiaries with both baseline and follow-up responses.

5. The results of the models from steps 3 and 4 were then used to calculate the average expected death rates and the average expected PCS same or better rates at follow-up for the beneficiaries in each of the 10 cohorts in our FFS sample. These expected rates were then combined in the two-part model by multiplying them together.

6. After the expected rates had been estimated, the actual death rates and actual PCS same or better rates for those alive at follow-up were calculated for each of our FFS cohorts. These actual rates were also multiplied together.

7. Deviation scores were next calculated by subtracting the expected values for the combined "alive and PCS same or better" rates from the corresponding actual values for each of our FFS cohorts (or for each M+C plan).

8. Standard errors were next calculated for the deviation scores. These were used to calculate *t* statistics, by dividing the deviation scores by the standard errors.

9. Finally, using the *t* statistics, FFS cohorts whose actual "alive and PCS same or better" scores were significantly above (or below) their expected values could be identified.

In the M+C study, an overall ANOVA F-test of M+C plan differences was also performed for both the PCS and MCS (HSAG, 2001; Rogers *et al*., 2001). Tests of individual plan differences between actual and expected values, using the *t* statistics described in steps 8 and 9 above, were not conducted unless the overall F-test was statistically significant for identifying differences among M+C plans.

In the M+C study results, the F test for the PCS scores was not statistically significant, but the F test for MCS scores was statistically significant. As a result, the M+C results indicated that there were no differences among the 188 M+C plans for two-year changes in PCS outcomes. In contrast, the F test showed a significant difference among the M+C plans for two-year changes in MCS outcomes (HSAG, 2001; Rogers *et al*., 2001).

With only 9 GP and SGA cohorts involved in our FFS analysis, the number of FFS groups is much smaller than the 188 plans in the M+C analysis. As a result, we focused our replication of the HAL statistical methods on the *t* statistics used to compare the differences between cohorts' expected and actual rates of outcomes in the "PCS alive

and same or better" and the other dichotomous outcome categories for PCS and MCS scores.

5.3 Results of the Statistical Analysis

Table 5-1 presents two-year change category rates for the PCS and MCS, calculated by applying the HAL software. For each of the 10 cohorts, expected and actual death rates are calculated, and the dichotomous outcomes from the better/same/worse change categories are also presented. For example, for the national cohort, about 9.3 percent of respondents actually died between baseline and follow-up (over the two-year period from 1998 to 2000). This is compared to an expected death rate of 8.9 percent.

In the national cohort, the actual percentage of respondents in Table 5-1 with outcomes of "PCS same or better" was about 67 percent of the sample. This was a bit below the expected rate for that cohort of about 70 percent. The actual percentage of outcomes in the "PCS better" category was much lower, however, at only about 17 percent. That was, however, above the expected level of 15 percent for that outcome. These data indicate that the use of the SEM-based cut-offs to define "better" and "worse" outcomes means that most of the respondents fall into the two-year change category of "same."

Table 5-1

Expected and Actual Outcomes by Cohort Using HAL Software and Fee-for-Service Health Outcomes Study Data

Cohort	Expected Death Rate	Actual Death Rate	Expected PCS Same or Better	Actual PCS Same or Better	Expected PCS Better	Actual PCS Better
NATIONAL	8.87%	9.31%	70.22%	66.59%	15.18%	16.71%
Pennsylvania SGA	8.78	8.62	69.85	69.93	15.22	13.36
Georgia SGA	9.79	10.00	70.25	69.36	15.17	16.15
Wisconsin SGA	8.44	6.82	69.99	69.73	15.33	13.21
Arizona SGA	8.27	8.78	70.61	70.79	14.65	15.73
Washington SGA	8.95	8.16	70.67	69.18	14.81	14.86
PGP A	9.93	11.34	69.96	70.04	15.09	17.06
PGP B	9.48	10.75	69.99	69.19	15.28	14.65
PGP C	7.31	8.05	70.59	69.46	14.72	15.88
PGP D	8.38	8.54	70.84	73.09	14.65	15.48

Table 5-1 (continued)

Expected and Actual Outcomes by Cohort Using HAL Software and Fee-for-Service Health Outcomes Study Data

Cohort	Expected MCS Same or Better	Actual MCS Same or Better	Expected MCS Better	Actual MCS Better
NATIONAL	77.11%	81.90%	15.61%	21.58%
Pennsylvania SGA	76.94	80.40	15.69	18.49
Georgia SGA	77.25	76.25	15.41	15.91
Wisconsin SGA	76.59	82.11	15.82	16.89
Arizona SGA	78.55	82.02	14.23	17.53
Washington SGA	78.17	83.37	14.62	19.73
PGP A	77.29	77.78	15.32	16.87
PGP B	76.72	83.17	15.36	17.00
PGP C	78.48	80.63	14.13	16.41
PGP D	78.67	80.15	14.15	16.79

OUTPUT: hal05

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Study; Sinclair and Gandek, 2001.

For most of the outcome comparisons in Table 5-1, the actual and expected rates are fairly close. There are some outcomes that show more sizable differences, however, such as for "MCS better" for the national cohort. In that case, the actual rate was about 22 percent, the expected rate about 16 percent, and the difference about 6 percentage points. Analysis of the statistical significance of these differences are presented in the next three tables.

Table 5-2 presents the steps for calculating the *t* statistics for evaluating the statistical significance of differences between actual and expected two-year change category outcomes for PCS scores for each FFS cohort. The figures in Table 5-2 are expressed as proportions resulting from the two-part model used to estimate PCS outcomes.

The first five columns in Table 5-2 include data for the outcome "alive and PCS same or better." Columns 1 and 2 include the actual and expected values for this outcome. Column 3 includes the deviation, or the difference calculated by subtracting the expected value from the actual value. Column 4 includes the standard error for the deviation figure. Column 5 includes the *t* statistic, calculated by dividing the deviation value by its standard error. Columns 6-10 include the same data elements calculated for the other PCS outcome variable, "alive and PCS better."

Table 5-2

Expected versus Actual PCS Outcomes and *t* Statistics by Cohort Using HAL Software and Fee-for-Service Health Outcomes Study Data

Cohort	Actual Alive and PCS Same or Better	Expected Alive and PCS Same or Better	Deviation Alive and PCS Same or Better	Standard Error for Deviation Alive and PCS Same or Better	*t* Statistic for Alive and PCS Same or Better
NATIONAL	0.604	0.640	-0.036	0.022	-1.651
Pennsylvania SGA	0.639	0.637	0.002	0.021	0.083
Georgia SGA	0.624	0.634	-0.009	0.022	-0.431
Wisconsin SGA	0.650	0.641	0.009	0.019	0.480
Arizona SGA	0.646	0.648	-0.002	0.021	-0.096
Washington SGA	0.635	0.643	-0.008	0.021	-0.380
PGP A	0.621	0.630	-0.009	0.020	-0.454
PGP B	0.618	0.634	-0.016	0.019	-0.856
PGP C	0.639	0.654	-0.016	0.019	-0.829
PGP D	0.669	0.649	0.020	0.020	0.984

Table 5-2 (continued)

Expected versus Actual PCS Outcomes and *t* Statistics by Cohort Using HAL Software and Fee-for-Service Health Outcomes Study Data

Cohort	Actual Alive PCS and Better	Expected Alive PCS and Better	Deviation Alive PCS and Better	Standard Error for Deviation Alive PCS and Better	*t* Statistic for Alive PCS and Better
NATIONAL	0.152	0.138	0.013	0.016	0.827
Pennsylvania SGA	0.122	0.139	-0.017	0.016	-1.076
Georgia SGA	0.145	0.137	0.009	0.016	0.537
Wisconsin SGA	0.123	0.140	-0.017	0.014	-1.267
Arizona SGA	0.143	0.134	0.009	0.015	0.585
Washington SGA	0.136	0.135	0.002	0.015	0.102
PGP A	0.151	0.136	0.015	0.014	1.062
PGP B	0.131	0.138	-0.008	0.013	-0.564
PGP C	0.146	0.136	0.010	0.014	0.696
PGP D	0.145	0.134	0.011	0.014	0.745

**p<.05
***p<.01

OUTPUT: hal05

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Study; Sinclair and Gandek, 2001.

The results in Table 5-2 indicate that none of the differences between actual and expected PCS outcomes are statistically significant. That is inconsistent with the results found using our statistical approach, presented above in Table 4-4. We found a significant increase in PCS outcomes for PGP C. (As noted, however, the effect size was below the usual threshold for clinical or policy significance of 2.0 points).

The results in Table 5-2 are consistent with the results found in HAL's M+C study, which was conducted over the same two-year period from 1998 to 2000. As noted, that study found no statistically significant differences for PCS outcomes between the 188 M+C plans analyzed.

One hypothesis regarding the general lack of significant PCS differences across both the M+C and FFS studies is that longer time periods may be needed to detect significant changes in PCS health status. The HAL group has suggested that four years would be a better follow-up period for this type of analysis than two years. Four years was the follow-up period used in the MOS (Ware *et al*., 1996).

Table 5-3 presents results for MCS outcomes in the same format as was used for the PCS outcomes in Table 5-2. This table shows that in seven cases, the differences between actual and expected MCS outcomes were statistically significant. These include four cohorts with statistically significant results for the "MCS same or better" outcome and three cohorts with statistically significant results for the "MCS better" outcome. In each significant case the change was positive, with actual outcomes exceeding expected

Table 5-3

Expected versus Actual MCS Outcomes and *t* Statistics by Cohort Using HAL Software and Fee-for-Service Health Outcomes Study Data

	Actual MCS Same or Better	Expected MCS Same or Better	Deviation MCS Same or Better	Standard Error for Deviation MCS Same or Better	*t* Statistic for MCS Same or Better
NATIONAL	0.819	0.771	0.048	0.020	**2.367** ***
Pennsylvania SGA	0.804	0.769	0.035	0.020	1.742
Georgia SGA	0.762	0.773	-0.010	0.020	-0.491
Wisconsin SGA	0.821	0.766	0.055	0.017	**3.185** ***
Arizona SGA	0.820	0.786	0.035	0.019	1.783
Washington SGA	0.834	0.782	0.052	0.019	**2.673** ***
PGP A	0.778	0.773	0.005	0.019	0.260
PGP B	0.832	0.767	0.064	0.017	**3.715** ***
PGP C	0.806	0.785	0.021	0.017	1.251
PGP D	0.802	0.787	0.015	0.018	0.827

Table 5-3 (continued)

Expected versus Actual MCS Outcomes and *t* Statistics by Cohort Using HAL Software and Fee-for-Service Health Outcomes Study Data

	Actual MCS Better	Expected MCS Better	Deviation MCS Better	Standard Error for Deviation MCS Better	*t* Statistic for MCS Better
NATIONAL	0.216	0.156	0.060	0.017	**3.412** ***
Pennsylvania SGA	0.185	0.157	0.028	0.017	1.630
Georgia SGA	0.159	0.154	0.005	0.018	0.288
Wisconsin SGA	0.169	0.158	0.011	0.015	0.715
Arizona SGA	0.175	0.142	0.033	0.017	**1.992** **
Washington SGA	0.197	0.146	0.051	0.017	**3.077** ***
PGP A	0.169	0.153	0.015	0.016	0.964
PGP B	0.170	0.154	0.016	0.015	1.114
PGP C	0.164	0.141	0.023	0.015	1.560
PGP D	0.168	0.142	0.026	0.015	1.736

**p<.05

***p<.01

OUTPUT: hal05

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Study; Sinclair and Gandek, 2001.

outcomes. These results are consistent, in terms of statistical significance, with HAL's study of M+C plans, which found a number of significant differences between plans on MCS scores. The M+C results had some positive and some negative differences, however.

However, the results in Table 5-3 are very different from the results found using our statistical methods, as reported above in Table 4-4. In our statistical analysis, only one cohort had a significant result (PGP A), and its result was negative, with the actual MCS outcome lower than the expected MCS outcome. Results for that PGP were not significantly different in the analysis in Table 5-3 using the HAL software. In contrast, five cohorts had significant results in this analysis using the HAL software (and two of these cohorts had significant results for both MCS outcomes); but none of those cohorts had significant results in our analysis, presented in Table 4-4.

As noted, one of the main differences between the HAL methodology for analyzing MCS outcomes and our method is the treatment of respondents who died between baseline and follow-up. HAL's analysis of MCS outcomes excludes respondents who died between baseline and follow-up. As described in Chapters 3 and 4, our analysis accounted for those deaths by imputing MCS values for those respondents.

Table 5-4 tests the sensitivity of the HAL results to the approach for handling deaths of respondents. It presents the same MCS outcomes as Table 5-3, using the same HAL analytic software. The only difference is that the data set used for Table 5-4 includes the values we imputed for MCS scores for respondents who died between

Table 5-4

Expected versus Actual MCS Outcomes for *t* Statstics by Cohort with Values Imputed for Death Using HAL Software and Fee-for-Service Health Outcomes Study Data

	Actual MCS Same or Better	Expected MCS Same or Better	Deviation MCS Same or Better	Standard Error for Deviation MCS Same or Better	*t* Statistic for MCS Same or Better
NATIONAL	0.763	0.767	0.004	0.019	-0.205
Pennsylvania SGA	0.763	0.768	0.039	0.191	-0.205
Georgia SGA	0.710	0.768	-0.059	0.019	**- 3.052** ***
Wisconsin SGA	0.773	0.761	0.012	0.017	0.730
Arizona SGA	0.754	0.782	-0.273	0.018	-1.480
Washington SGA	0.778	0.778	0.000	0.019	0.005
PGP A	0.732	0.769	-0.037	0.017	**- 2.111** **
PGP B	0.765	0.761	0.004	0.016	0.226
PGP C	0.754	0.782	-0.029	0.016	-0.176
PGP D	0.751	0.783	-0.032	0.017	-1.894

Table 5-4 (continued)

Expected versus Actual MCS Outcomes for *t* Statstics by Cohort with Values Imputed for Death Using HAL Software and Fee-for-Service Health Outcomes Study Data

	Actual MCS Better	Expected MCS Better	Deviation MCS Better	Standard Error for Deviation MCS Better	*t* Statistic for MCS Better
NATIONAL	0.198	0.158	0.040	0.016	**2.443** **
Pennsylvania SGA	0.173	0.158	0.015	0.016	0.923
Georgia SGA	0.166	-0.004	0.014	-0.272	0.000
Wisconsin SGA	0.156	0.160	-0.004	0.014	-0.272
Arizona SGA	0.158	0.243	0.014	0.016	0.911
Washington SGA	0.182	0.147	0.034	0.016	**2.156** **
PGP A	0.159	0.155	0.004	0.015	0.244
PGP B	0.152	0.156	-0.004	0.014	-0.291
PGP C	0.152	0.142	0.010	0.014	0.731
PGP D	0.155	0.143	0.012	0.014	0.855

**p<.05
***p<.01

OUTPUT: hal08

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Study; Sinclair and Gandek, 2001.

baseline and follow-up. The data set used for Table 5-3 did not include any of the imputed data, to make the results more consistent with the HAL approach.

Table 5-4 reveals that the statistical results found using the HAL methodology are sensitive to the assumptions about deaths for MCS outcomes. Table 5-3 showed seven significant differences between actual and expected MCS outcomes, and all were positive. Table 5-4 shows four significant differences between actual and expected MCS outcomes, and two are negative. Moreover, only two of the seven significant results in Table 5-3 remained significant in Table 5-4. Finally, Table 5-4 shows a negative and significant result for MCS outcomes for PGP A, which is consistent with the negative and significant result found for that PGP in our statistical analysis presented in Table 4-4. MCS outcomes for that cohort were nonsignificant in Table 5-3.

In sum, the differential treatment of deaths between the statistical analysis of PCS and MCS outcomes in the HAL approach raises concerns. The differences in results between Tables 5-3 and 5-4 are due solely to changing the treatment of deaths by adding the imputed values for deaths in the data set used to generate Table 5-4. Ignoring deaths for MCS outcomes appears to bias the HAL statistical methodology toward more positive findings.

5.4 Summary

In this chapter, we compared expected versus actual follow-up PCS and MCS categorical outcomes, using a statistical methodology developed by HAL for their HOS study of M+C plans. This constitutes an alternate approach for testing the ability of health status scores to discriminate between high- and low-performing PGPs and SGAs.

We first assigned our FFS respondents to change categories at follow-up, depending on whether their PCS and MCS scores were "better," "the same," or "worse" compared with their baseline scores. We used the same cut-off points to define those categories as HAL used in their M+C analysis.

We next applied SAS software provided us by HAL to replicate their statistical methodology. Their approach uses coefficients derived from logistic regression analysis to estimate the expected probability of dichotomous outcomes, such as "PCS same or better" and "MCS same or better." The PCS analysis is extended to incorporate a two-part model, which also estimates the probability of death using coefficients derived from logistic regression analysis. It thus estimated the probability of "PCS alive and same or better." Deaths between baseline and follow-up were excluded from the MCS analysis. A broad range of independent variables is included in the models predicting the likelihood of death. A much more limited range of variables is used in the models predicting categorical change in PCS and MCS.

Expected probabilities for the categorical outcomes were then compared with the actual percentages of respondents falling into each category for each cohort. The differences between the expected and actual values were assessed for statistical significance using *t* statistics.

We found no statistically significant differences between expected and actual PCS scores. This was consistent with HAL's findings for PCS outcomes in its HOS study of M+C plans. However, this was somewhat inconsistent with the findings using our statistical methods in Chapter 4. We found one difference between actual and expected PCS scores to be significant (for PGP C). However, that result was below the 2.0 point threshold for clinical or policy significance.

In contrast, we found seven statistically significant results for MCS outcomes using the HAL software and our FFS HOS data. They included four significant results for the "MCS same or better" outcome and three significant results for the "MCS better" outcome. Each was positive, indicating the cohorts performed better than expected. This was somewhat consistent with HAL's findings for MCS outcomes in its HOS study of M+C plans, where a range of significant results were also found (although they included a mix of positive and negative results). These results are inconsistent with the findings using our statistical approach from Chapter 4, however; we found only one significant result and it was negative.

Our final analysis in this chapter recalculated the MCS analysis using the HAL software, this time with MCS values imputed for deaths between baseline and follow-up.

We found differences in the results compared with the prior MCS analysis. When values were imputed for MCS for deaths, four *t* statistics were significant, and two were negative. This contrasts with the seven positive and significant results found using the standard HAL approach (which, as noted, excludes deaths from the analysis of MCS outcomes). Moreover, only two of those seven significant results remained significant when values were imputed for death. As a result, it appears that excluding deaths from the analysis of MCS outcomes may bias the HAL method toward more positive and more statistically significant results.

In sum, there are several differences between HAL's statistical approach and ours. These include the treatment of deaths between baseline and follow-up, the use of change categories (better/same/worse) instead of retaining the continuous measurements used for the PCS and MCS, and the range of independent variables used in the statistical modeling efforts. The HAL approach used logistic regression analysis for statistical modeling since it studied dichotomous outcomes, and extended that approach to a two-part model for PCS that incorporated the likelihood of death between baseline and follow-up. We used ordinary least squares regression since our outcomes were continuous variables.

6 Summary of Findings and Discussion

6.1 Summary of Findings

The primary goal of this study was to assess the validity and feasibility of using longitudinal estimates of self-reported health status for cohorts of Medicare beneficiaries to evaluate the care provided to FFS beneficiaries by physician group practices or by FFS providers in small geographic areas. Our aim was to explore new methods which might be used by CMS to differentiate high-performing and low-performing providers serving its beneficiaries. Health status data could be used to complement other quality monitoring efforts, including those based on report cards, patient satisfaction, and other approaches. For this study, we evaluated the performance of four large multi-specialty group practices (PGPs) and the health care systems serving Medicare FFS beneficiaries in five small geographic areas (SGAs) using two alternative methods of analyzing follow-up scores of health status.

This analysis builds upon previous work evaluating the feasibility of implementing performance measurement in Medicare FFS that has been conducted by Health Economics Research for the Centers for Medicare & Medicaid Services (CMS). Other studies have compared FFS and managed care systems and compared different managed care plans serving M+C beneficiaries (Ware *et al*., 1996; HSAG, 2001). However, validity and feasibility issues related to using self-reported health status, as

derived from the HOS, for a nonenrolled, FFS population for performance measurement have not previously been studied.

6.1.1 Summary of Descriptive Comparisons between Baseline and Follow-up Respondents

In our descriptive analyses, we explored four selected issues related to the validity and feasibility of using health status follow-up scores in Medicare FFS for performance measurement. First, we were interested in determining whether there were systematic differences in respondents to the baseline HOS versus the follow-up HOS. We analyzed this issue using response rates and the distributions of respondents across strata of interest. Second, we were interested in determining whether there were any systematic differences in the completeness of survey responses, since the scoring of the PCS and MCS measures are highly dependent on item response. Third, we were interested in exploring differences in mean PCS and MCS scores between baseline and follow-up and directly examining the effect of setting the PCS score to zero at follow-up for decedents. Fourth, we were interested in exploring the degree of retention of baseline respondents at follow-up by the four physician group practices.

Our analysis revealed no systematic differences in response rates or material differences in characteristics of follow-up respondents that were unexpected. The overall follow-up response rate was an impressive 92 percent. We found no significant differences in response rates across the national sample, four group practices, and five geographic areas. Further, we observed a high rate of retention of beneficiaries by the

four physician group practices. Over 80 percent of beneficiaries remained with their usual source of care provider between baseline and follow-up. This suggests that it is feasible to use a longitudinal survey method in Medicare fee-for-service.

We did not observe any systematic differences in the item completion rate of the survey instrument between baseline and follow-up. The same scoring method could be used for all but a handful of respondents, thus we believe any observed differences in mean PCS and MCS scores between the two time periods would be a reflection of true changes in average physical and mental health and not an artifact of changes in scoring methods.

As expected, we observed aging of the cohort and some degradation of physical health as assessed through an increase in the number of chronic conditions and activities of daily living limitations. We also observed a reduction in the proportion of respondents that rated their health as excellent or who received a PCS score above 50. We did not observe much change in average mental health status over the two-year period.

We did observe a statistically and clinically significant decline in average physical health scores of roughly 5 percentage points at follow-up. This pattern held for all ten sampling strata. There was some variation in the degree of physical health decline between men and women, across racial cohorts, and across the age groups. The very old (85+) experienced the largest decline, 10 points on average between the two years. Beneficiaries with selected chronic conditions, such as CHF, also experienced larger average declines in physical health than those without these diseases. And, not

surprisingly, we observed those with the highest baseline PCS scores showing the largest average decline in health over the two-year period relative to those with the lowest PCS scores. This is the regression to the mean phenomenon that has been observed in other studies of changes in health status.

The observed differences in mean PCS scores is heavily influenced by setting the PCS to zero for decedents. Restricting our analysis to only survivors, we observed a decline in average PCS of roughly two points; a clinically significant difference. Many of the observed statistical and clinical differences that we observed across the strata were a function of a disproportionate rate of death across strata. Removing decedents from the follow-up analysis resulted in minimal differences in mean follow-up PCS scores across the strata.

Removing all decedents from the follow-up health status score estimation appeared to underestimate declines in PCS, on average, while setting death to zero appeared to overestimate the PCS decline, on average. This raises an issue of the fairness of comparisons using differences in mean scores across different health care systems or providers in Medicare FFS. Providers or health systems may be systematically penalized if, at baseline, they have a pool of patients with a higher than average risk of mortality.

6.1.2 Summary of Comparison of Expected and Follow-up Health Status Methods

Given the limitations of comparing mean scores between two time periods demonstrated in the descriptive analyses, we next critically evaluated the performance of

the four PGPs and the health care systems serving FFS beneficiaries in five SGAs using two methodological approaches, one which we developed and one developed by HAL. Each predicts mean PCS and MCS follow-up scores for cohorts of beneficiaries as a function of baseline characteristics of the cohort, and then compares expected with actual mean follow-up scores.

One of the differences between the two methods is how follow-up physical and mental health status is estimated for baseline respondents who die before the follow-up survey is fielded. A second difference is use of individual change score calculations in the HAL method to estimate categorical outcomes (better, same or worse), at follow-up prior to comparing predicted and actual follow-up health status. Our principal focus, however, was on the ability of each method to discriminate between better and worse performing PGPs and SGAs.

Our estimation method used utilities to impute PCS and MCS values for decedents. Further, our method used a "standard of care" modeling approach whereby we predicted follow-up health status scores using the national cohort. We then compared the performance of PGPs and SGAs relative to the expected follow-up PCS and MCS scores calculated using the standard of care model. Our standard of care models predict follow-up PCS and MCS scores using the baseline PCS or MCS score and a limited number of other beneficiary and survey characteristics that increase or decrease the slope of the trajectory over the two year period, thus controlling for differences across cohorts. To provide a comprehensive assessment, the follow-up health status prediction models

were estimated for all baseline respondents who responded on follow-up and those who died before follow-up. We did this by imputing PCS and MCS scores for decedents.

Our results showed that there were no clinically significant differences in actual performance relative to the expected performance for any of our PGPs and SGAs for either PCS or MCS. We considered a 2 point difference to be a clinically significant difference. For PCS, each of the nine performance cohorts all did slightly better than predicted using the national random sample's experience as a benchmark, but none had more than a 1 point difference. Two of the physician group practices had actual follow-up PCS scores that were statistically higher than predicted follow-up scores, using a statistical test of significance at the 0.10 level. Each of these practices had an average PCS score that was about 1 point above predicted, but this was less than a clinically significant difference.

There was more variation in the comparison of mean predicted and actual MCS scores; again however, none of the differences were clinically meaningful. Three of the SGAs and three of the PGPs had actual mean MCS scores lower than predicted; the rest higher. One PGP's lower than expected mean MCS score was statistically significant.

The HAL methodology was developed to evaluate the performance of M+C health plans. It builds off an approach developed for the Medical Outcomes Study (MOS) for comparing outcomes between managed care and fee-for-service (Ware *et al*., 1996). The HAL approach excludes decedents from the MCS analysis but retains them for the PCS analysis. The two primary outcomes for comparison are rates of beneficiaries

(1) alive and PCS same or better and (2) MCS same or better, as they were developing a measure to indicate whether a health plan was maintaining or improving the health of its members.

Using this method, none of the differences between actual and expected rates of beneficiaries alive and PCS same or better or beneficiaries alive and PCS better were statistically significant. This is consistent with the results found in HAL's Medicare+Choice study, which was conducted over the same two-year period from 1998 to 2000. As noted, that study found no statistically significant differences for PCS outcomes among 188 M+C plans.

For MCS scores, the HAL method yielded seven sets of statistically significant differences when decedents were excluded. All of these results were positive in terms of performance. The national sample, two SGAs, and one PGP had better than expected rates of beneficiaries the same or better at follow-up. The national sample and two SGAs were also found to have statistically higher than expected rates of beneficiaries who were better at follow-up. The HAL M+C study found more variation in performance among the 188 health plans analyzed; 13 plans were identified as better and 15 plans were identified as worse than the national average for MCS.

However, when we imputed a value of MCS for decedents, the results using the HAL method changed. Rather than seven outcomes having better performance than expected, we observed one SGA and one PGP having worse performance than expected, when assessing rate of beneficiaries with outcomes "MCS same or better". The national

FFS sample and one PGP continued to outperform the predicted rate of beneficiaries with outcomes of "MCS better".

6.2 Methodological Issues

This report compared the performance of selected PGP and SGA cohorts of Medicare FFS providers using two alternative methodologies that are new to the research and policy community. As noted, we found no substantive differences between cohorts for either PCS or MCS outcomes using our RTI standard care method. Our results agree generally with those found using the HAL method, given the differences in handling decedents for MCS analysis. However, is it accurate to say that there are no substantive differences among these FFS cohorts after two years? There are several possible explanations for our null findings.

First, the RTI and HAL statistical models may have both produced inaccurate expected levels for the comparisons. The expected values were generally similar for the different cohorts, but this reflected the fact that the cohorts were similar in terms of beneficiary characteristics to begin with. Both models explained substantial proportions of the variation in follow-up component scores in the national sample, and are capable of producing different expected values for cohorts with different case mixes. Moreover, even if risk adjustments were ignored and we simply compared cohort scores to the national averages, the absolute differences would still be small.

Second, exclusion of nonrespondents may have biased the results. The follow-up response rate in our total analytic sample was quite high, over 90 percent, but the baseline response rate was only 68 percent. A prior analysis of nonresponse in the FFS HOS using secondary data (Khatusky, 2001) indicated that nonrespondents were only slightly less healthy than respondents. Including nonrespondents would therefore lower both the expected and actual component scores for a particular cohort, but probably would not affect the comparisons by much unless the unobserved follow-up outcomes for nonrespondents were considerably lower than those predicted by the standard care model.

Third, our procedures for imputing PCS and MCS scores for respondents who died between baseline and follow-up may have been inaccurate. Our method is innovative, but as such it should be further tested and validated in studies conducted by other researchers using a range of other data sets and utility measures. Changing the imputed values, however, would have similar effects on both expected and actual scores, so this is unlikely to alter the overall pattern of findings.

Fourth, the group practices that agreed to participate in our study may have been a self-selected sample of high performing groups. They may have already been performing at a high level at baseline, and thus had difficulty showing major improvement at follow-up, controlling for baseline performance.

A fifth possibility is that PCS and MCS scores are too insensitive to detect true differences in cohort performance, even when genuine differences exist for the aggregate measures we were studying. This also seems unlikely since the SF-36 has previously

been shown to be responsive to adverse events and to the effects of many different medical and surgical interventions. Our statistical tests had sufficient power to detect component score differences as small as one point.

A sixth possibility is that the findings are accurate. The results for group practices and small geographic areas were aggregated over large numbers of physicians and beneficiaries, many of whom were healthy or suffer from only minor health problems. When aggregated at this level, differences among cohorts may be negligible. Facility-level analyses of nursing homes (Porell and Caro, 1998) and hospitals (Iezonni, 1997) have also failed to find consistent differences over time in performance measures.

6.3 Policy Implications

Our results may indicate that health status scores, at least as measured through the PCS and MCS, have limited value for identifying high- and low-performing PGPs or SGAs. On the other hand, our results could also indicate that there were in fact no significant differences in overall health status performance among the PGPs and SGAs we analyzed. Additional studies, with other large samples of Medicare beneficiaries, are needed to further investigate these questions.

PCS and MCS scores are likely to be more sensitive to specific interventions, clinical trials, or specific diseases than to changes in a general population. Health care interventions may have significant effects on subpopulations that "wash out" in population averages. For example, if effective health care improves the health status

scores of the 10 percent of the population that is most ill by 2 points, but has no effect on the health status scores of the healthier 90 percent, then the average health status of the population will improve by only 0.2 points.

Variations in preventive services and "well patient" care are unlikely to impact broad measures of health status. Recognizing this, health status performance measurement might instead focus on vulnerable groups that are are the most likely to suffer declining health. These could include the chronically or severely ill, beneficiaries known to have specific chronic diseases, and frail, old, previously hospitalized or other beneficiaries at high risk of health status decline. Effects of health care providers on health status outcomes are more likely to be detected in those populations than in random samples of beneficiaries that include predominantly healthy beneficiaries.

A related question is whether the primary focus of measuring and reporting health status outcomes should be for public accountability of providers serving Medicare beneficiaries, or whether the goals of this effort should be expanded to include continuous quality improvement (CQI). In our study, both of the analytic methods we used focused on identifying "outliers," or providers performing significantly above or below standard or average care. Thus those methods emphasize public accountability of providers, although they could also have the effect of promoting quality improvement among the providers identified as outliers and others concerned about their reported performance. Public release of quality performance data has been shown in some cases

to motivate hospitals, physician groups, and individual providers to take concrete steps to improve their quality of care (Shahian *et al.*, 2001; Galvin & Milstein, 2002).

However, CQI programs generally take a different approach, utilizing quality data for confidential peer review, benchmarking results against "best practices," and fostering collaborative efforts at provider education and systems improvements (Shahian *et al.*, 2001). Identifying outliers is deemphasized in favor of broader efforts to raise the quality performance of all providers.

CQI programs have generally focused on process measures of quality (e.g., frequency of lab tests for diabetics) instead of outcomes, however, since the former are considered more under the control of providers, occur more frequently (facilitating measurement and feedback of data to providers), and usually require shorter time periods for evaluation of effects (Eddy, 1998). Outcomes, such as health status scores or mortality rates, can in some cases be applied in CQI efforts, however. Cardiac surgery mortality rates, for example, have been reported publicly for provider accountability, but also applied in CQI efforts (Shahian *et al.*, 2001). Those quality improvement efforts, however, benefited from the focus on a single, clearly defined surgical procedure, relatively clear links between provider performance and mortality outcomes (although debate continues on that issue), and relatively wide variations in measured outcomes between the best and worst performing providers. The PCS and MCS outcomes measured in our study do not fit those criteria for CQI efforts, since they were applied to a broad mix of patients and very little variation was found among the groups of providers

we studied. However, if PCS and MCS outcomes were measured for selected subpopulations, such as severely ill patients, then these outcomes might also serve for CQI efforts. This need not supplant public reporting of PGP or SGA outcomes to identify outliers in response to public demands for provider accountability, but could complement those efforts. Indeed, pursuing both public reporting for accountability and CQI programs is an approach recently recommended for measurement and analysis of cardiac surgery outcomes (Shahian *et al.*, 2001).

Ware and colleagues have suggested a longer follow-up period may be required to identify differences in PCS or MCS outcomes which have clinical or policy significance; they suggest four years may be more appropriate than the two year period used in this study (Ware *et al*., 1996; 2001). That approach may be worth considering. A longer period may be required for differences in clinical quality of care to be reflected in average health status outcomes for large groups of beneficiaries. However, longer time periods would also mean increased measurement problems due to higher rates of death between baseline and follow-up.

Moreover, that approach raises the question of in what time frame would policy makers *expect* to see meaningful provider performance differences emerge. Studies that require more than two years to complete may extend beyond the time horizon faced by most policy makers, given their need to respond to the expectations of beneficiaries and other stakeholders regarding provider accountability.

Techniques for analyzing changes in health status over time remain an area of active concern for policymakers, and an area of active investigation and development in health services research. For example, the RTI standard care models should also be estimated and tested for other large, representative beneficiary samples, and for high risk subpopulations as well. The two methods presented here provide useful perspectives for future efforts to assess and monitor the quality of care provided to Medicare beneficiaries. However, they need to be further evaluated and contrasted with other approaches.

References

Bradley CH, J Kroll, M Holmes-Rovner: "The Health and Activities Limitation Index in Patients with Acute Myocardial Infarction." *J Clinical Epidemiology* 53:555-562, 2000.

De Vaus D: Research Design in Social Research. Thousand Oaks, California: Sage Publications, 2001.

Diehr P, D Patrick, S Hedrick, *et al*.: "Including Deaths When Measuring Health Status Over Time." *Medical Care* 33(4 Suppl): AS164-AS172, 1995.

Diehr P., D Patrick, J Spertus., *et al*:. "Transforming Self-rated Health and the SF-36 Scales to Include Death and Improve Interpretability." *Medical Care* 39(7):670-680, 2001.

Eddy D. (1998) Performance measurement: problems and solutions. *Health Affairs* 17(4): 7-25.

Erickson P.: "Evaluation of a Population-based Measure of Quality of Life: the Health and Activity Limitation Index." (HALex). *Quality of Life Research* 7:101-114, 1998.

Ferraro KF, YP Su. "Physician-evaluated and Self-reported Morbidity for Predicting Disability." *American Journal of Public Health*, 2000:90:103-108.

Gage B., S Miller, V Mor, *et al*.: "*Synthesis and Analysis of Medicare's Hospice Benefit*." Reports 1-6. Prepared for the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Cambridge, Massachusetts: The MEDSTAT Group, 2000.

Galvin R. and A. Milstein. (2002) Large employers' new strategies in health care. *New England Journal of Medicine* 347(12): 939-942.

HSAG. *Medicare Health Outcomes Survey, Performance Measurement Report, Cohort I: 1998 – 2000 [Draft]*. Prepared for the Centers for Medicare & Medicaid Services. Phoenix, Arizona: Health Services Advisory Group, 2001.

Iezzoni L: Risk Adjustment for Measuring Healthcare Outcomes. Second Edition. Chicago: Health Administration Press, 1997.

Jacobson PE. Introduction to Statistical Measures for the Social and Behavioral Sciences. Hinsdale, IL: Dryden Press, 1976.

Khatutsky G., N McCall, G Pope, K Smith. *Health Status Nonresponse Bias in the Medicare Fee-for-Service Health Outcomes Survey.* Prepared for the Health Care

Financing Administration under Master Order Contract No. 500-95-0058, Task Order No. 2. Waltham, Massachusetts: Health Economics Research, Inc., 2001.

McCall N., G Pope., K Adamache, *et al*:. *Research and Analytic Support For Implementing Performance Measurement in Medicare Fee for Service: First Annual Report.* Prepared for the Health Care Financing Administration under Master Order Contract No. 500-95-0058, Task Order No. 2. Waltham, Massachusetts: Health Economics Research, Inc., 1998.

McCall N., G Pope, M Griggs, *et al*.: *Research and Analytic Support For Implementing Performance Measurement in Medicare Fee for Service: Second Annual Report.* Prepared for the Health Care Financing Administration under Master Order Contract No. 500-95-0058, Task Order No. 2. Waltham, Massachusetts: Health Economics Research, Inc., 2000.

Menard S.: Longitudinal Research. Thousand Oaks, California: Sage Publications, 1991.

Petitti D.: Meta-Analysis, Decision Analysis, and Cost-Effectiveness Analysis: Methods for Quantitative Synthesis in Medicine. Second Edition. New York: Oxford University Press, 2000.

Pope G, M Griggs, N McCall: *Comparison of the Health Status of Medicare Fee-for-Service and Managed Care Enrollees Using the Health Outcomes Survey*. Prepared for the Health Care Financing Administration under Master Order Contract No. 500-95-0058, Task Order No. 2. Waltham, Massachusetts: Health Economics Research, Inc., 2001.

Porell F, FG Caro: "Facility-level Outcome Performance Measures for Nursing Homes." *The Gerontologist*; 38:665-683, 1998.

Rogers W, J Ware, B Gandek, M. Bayliss: *Examining Longitudinal Outcomes in the Medicare Health Outcomes Survey: Preliminary Recommendations*. Final Technical Report. Prepared for the Health Care Financing Administration and the National Committee for Quality Assurance. Boston: Health Assessment Lab and The Health Institute, 2000.

Rogers W, J Ware, B Gandek: *Measuring Outcomes in the Medicare Health Outcomes Survey: I. Sampling and Methods of Estimating Outcomes*. Prepared for the Health Care Financing Administration and the National Committee for Quality Assurance. Technical Expert Panel Meeting: May 8, 2001. Boston: Health Assessment Lab and The Health Institute, 2001.

Shahian D., Normand S., Torchiana D., Lewis S., Pastore J., Kuntz R. and P. Dreyer. (2001) Cardiac surgery report cards: comprehensive review and statistical critique. *Annals of Thoracic Surgery* 72: 2155-2168.

Sinclair J, B Gandek: *Program: c1csmix1.sas*. Boston: Health Assessment Laboratory, 2001.

Taris T: A Primer in Longitudinal Data Analysis. Thousand Oaks, California: Sage Publications, 2000.

Ware J, M Kosinski, S Keller: *"*SF-36 Physical and Mental Health Summary Scales: A User's Manual." Boston: Health Assessment Lab, New England Medical Center, 1994.

Ware J, M Kosinski, S Keller: SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. Second Edition. Boston: Health Assessment Lab, New England Medical Center, 1995.

Ware J, M Bayliss, W Rogers, M Kosinski, A Tarlov: "Differences in 4-Year Health Outcomes for Elderly and Poor, Chronically Ill Patients Treated in HMO and Fee-for-Service Systems: Results from the Medical Outcomes Study." *Journal of the American Medical Association* 276(13): 1039-1047, 1996.

Ware J, W Rogers, B Gandek: *Medicare HOS Briefing: Results from Cohort 1*. Presented to the Health Care Financing Administration. Boston: Health Assessment Lab and The Health Institute, 2001.

Ware J, M Kosinski: SF-36 Physical & Mental Health Summary Scales: A Manual for Users of Version 1, Second Edition. Lincoln, Rhode Island: QualityMetric, Inc., 2001.

Appendix Tables

Appendix - Table 1

PCS Change Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey, Using SF-12 Scores Only (followup PCS of decedents = 0)

	Decedents & Survivors							Survivors					
	PCS Follow-Up Respondents	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Change Due to PCS=0 for Decedents[1]	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
ALL	1,262	38.75	32.05	-6.70	0.41	***	58.5	87.0	39.62	36.84	-2.78	0.30	***
Site													
National	149	37.80	31.45	-6.35	1.24	***	63.5	87.2	38.37	36.05	-2.32	0.94	*
PA SGA[3]	134	40.46	32.21	-8.24	1.22	***	43.5	87.3	41.55	36.89	-4.66	0.97	***
GA SGA	136	38.87	32.30	-6.57	1.26	***	55.5	87.5	39.84	36.92	-2.92	0.99	**
WI SGA	131	38.61	30.92	-7.69	1.30	***	59.0	84.7	39.64	36.49	-3.16	0.93	***
AZ SGA	92	39.33	34.36	-4.96	1.16	***	23.5	95.7	39.72	35.93	-3.80	1.03	***
WA SGA	128	38.79	32.85	-5.94	1.19	***	57.5	87.5	40.07	37.55	-2.52	0.90	**
A PGP[4]	139	38.34	32.20	-6.14	1.34	***	86.2	84.9	38.78	37.93	-0.85	0.86	ns
B PGP	123	37.90	32.07	-5.83	1.17	***	53.8	88.6	38.88	36.19	-2.69	0.91	**
C PGP	108	38.87	30.99	-7.88	1.46	***	67.0	83.3	39.79	37.19	-2.60	1.00	*
D PGP	122	38.82	31.74	-7.08	1.38	***	64.7	85.2	39.74	37.23	-2.50	1.01	*
Sex													
Male	453	39.73	31.78	-7.95	0.71	***	61.0	83.7	41.08	37.98	-3.10	0.51	***
Female	809	38.21	32.21	-6.00	0.49	***	56.5	88.9	38.85	36.24	-2.61	0.37	***
Race													
White	1,160	38.85	31.95	-6.90	0.42	***	57.9	86.7	39.75	36.84	-2.91	0.31	***
Black	69	37.22	32.90	-4.32	1.89	*	78.7	91.3	36.96	36.04	-0.92	1.36	ns
Asian	6	44.28	44.51	0.23	6.43	ns	na	83.3	33.18	41.20	8.01	3.23	ns
Hispanic	7	29.71	29.43	-0.29	5.90	ns	na	85.7	44.28	44.51	0.23	6.43	ns
North American Native	1	53.81	33.71	-20.10	na	na	na	100.0	53.81	33.71	-20.10	na	na
Other	15	38.98	34.66	-4.32	2.36	ns	20.3	93.3	40.58	37.13	-3.45	2.35	ns
Unknown	4	39.38	23.32	-16.05	3.89	*	16.9	75.0	44.44	31.10	-13.34	3.95	ns
Original Reason for Entitlement													
Aged	1,039	40.18	33.00	-7.18	0.46	***	56.7	86.3	41.33	38.22	-3.11	0.34	***
Disabled	220	32.17	27.84	-4.33	0.87	***	71.2	90.5	32.02	30.78	-1.25	0.58	*
ESRD	0	na	na	na	na	na	0.0	0.0	na	na	na	na	na
ESRD and Disabled	3	26.54	13.80	-12.74	6.82	ns	na	66.7	27.59	20.70	-6.89	6.09	ns

Appendix - Table 1 (continued)

PCS Change Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey, Using SF-12 Scores Only (followup PCS of decedents = 0)

	Decedents & Survivors							Survivors					
	PCS Follow-Up Respondents n	Mean Baseline PCS Score	Mean Follow-Up PCS Score	Mean Difference in PCS	Standard Error of the Mean Difference	P	Percent of PCS Change Due to PCS=0 for Decedents[1]	Percent of PCS Follow-Up Respondents[2]	Mean Baseline PCS	Mean Follow-Up PCS	Mean Difference in Scores	Standard Error of the Mean Difference	P
Medicaid Status													
No Medicaid	1,098	39.61	32.84	-6.77	0.43	***	56.0	87.6	40.46	37.48	-2.98	0.32	***
Medicaid	164	33.03	26.82	-6.21	1.14	***	78.4	82.9	33.68	32.34	-1.34	0.84	ns
Age													
Under 65	121	33.08	30.11	-2.96	1.05	**	64.5	94.2	33.01	31.96	-1.05	0.79	ns
65-74	446	41.27	36.62	-4.65	0.59	***	48.6	92.8	41.84	39.45	-2.39	0.44	***
75-84	494	39.69	32.74	-6.96	0.66	***	49.7	88.7	40.42	36.92	-3.50	0.52	***
85 and Older	201	34.27	21.41	-12.86	1.21	***	76.0	65.7	35.69	32.60	-3.09	0.92	**
Household Income													
Less than $10,000	71	34.71	31.84	-2.86	1.58	ns	96.5	91.5	34.88	34.78	-0.10	1.23	ns
$10,000-$19,999	69	35.41	31.64	-3.77	1.45	*	57.4	91.3	36.26	34.66	-1.61	1.24	ns
$20,000-$49,999	114	40.45	34.48	-5.97	1.36	***	68.9	87.7	41.17	39.31	-1.86	0.88	*
$50,000 or more	40	43.09	41.08	-2.01	2.04	ns	68.3	97.5	42.77	42.14	-0.64	1.55	ns
Missing/No Response	968	38.91	31.44	-7.47	0.47	***	56.0	85.8	39.91	36.62	-3.29	0.35	***
Education													
Not a HS Graduate	128	36.85	30.16	-6.69	1.28	***	58.3	86.7	37.58	34.78	-2.79	0.96	**
High School Graduate or GEI	128	37.56	34.71	-2.85	1.06	**	73.4	93.0	38.09	37.34	-0.76	0.85	ns
Some College or 2 year degre	67	40.53	35.49	-5.04	1.67	**	47.4	94.0	40.40	37.74	-2.65	1.24	*
4 year college graduate	27	37.88	28.64	-9.24	3.55	*	76.4	81.5	37.33	35.15	-2.18	2.02	ns
More than a 4 year college deg	33	43.54	41.60	-1.94	1.37	ns	70.6	93.9	44.85	44.28	-0.57	1.04	ns
Missing/No Response	879	38.91	31.42	-7.49	0.50	***	57.1	85.6	39.95	36.73	-3.21	0.37	***
Marital Status													
Not Married	186	38.07	33.35	-4.72	0.89	***	46.8	98.9	38.58	36.07	-2.51	0.72	***
Married	199	38.95	33.66	-5.29	1.06	***	74.4	82.4	39.53	38.17	-1.35	0.71	ns
Missing/No Response	877	38.87	31.42	-7.45	0.50	***	57.6	85.5	39.89	36.74	-3.15	0.37	***
Home Ownership													
Owned by Beneficiary or Fam	269	39.16	34.91	-4.24	0.83	***	65.2	91.8	39.50	38.02	-1.48	0.61	*
Not Owned by Beneficiary or	96	36.79	30.04	-6.75	1.33	***	54.6	86.5	37.81	34.74	-3.07	1.00	**
Missing/No Response	897	38.84	31.41	-7.43	0.49	***	57.4	85.6	39.85	36.69	-3.16	0.37	***
Retirement Community													
Yes	70	36.49	29.72	-6.77	1.58	***	51.8	88.6	36.82	33.56	-3.26	1.11	**
No	293	38.84	34.34	-4.50	0.81	***	68.1	90.8	39.26	37.83	-1.43	0.58	*

Missing/No Response	899	38.90	31.49	-7.41	0.49	***	56.8	85.7	39.97	36.77	-3.20	0.37	***
Chronic Conditions													
Hypertenstion or high blood p	198	37.24	33.04	-4.20	0.92	***	53.8	92.9	37.50	35.56	-1.94	0.73	**
Angina pectoris or coroary art	62	33.64	27.82	-5.83	1.77	**	65.3	87.1	33.96	31.94	-2.02	1.30	ns
Congestive heart failure	35	30.25	27.84	-2.41	2.03	ns	131.2	88.6	30.68	31.43	0.75	1.50	ns
Acute myocardial infarction o	41	30.45	22.89	-7.56	2.57	**	92.5	75.6	30.84	30.28	-0.56	1.92	ns
Other heart conditions	105	34.29	29.14	-5.15	1.41	***	81.2	86.7	34.59	33.62	-0.97	1.00	ns
Stroke	41	32.60	23.22	-9.38	2.69	**	84.0	73.2	33.23	31.73	-1.50	2.03	ns
Emphysema, asthma, or COPI	43	30.70	24.88	-5.82	2.38	*	84.2	83.7	30.64	29.72	-0.92	1.84	ns
Crohn's disease, ulcerative colitis, or inflamatory bowel disea	28	31.63	28.21	-3.42	2.22	ns	76.3	89.3	32.41	31.60	-0.81	1.86	ns
Arthritis of the hip or knee	158	33.35	30.06	-3.29	0.92	***	58.0	93.0	33.69	32.31	-1.38	0.77	ns
Arthritis of the hand or wrist	138	34.98	31.31	-3.67	1.02	***	49.8	94.2	35.08	33.24	-1.84	0.82	*
Sciatica	100	33.63	30.75	-2.88	1.18	*	53.1	94.0	34.06	32.71	-1.35	1.07	ns
Diabetes, high blood sugar, or	58	36.06	31.49	-4.57	1.68	**	51.6	91.4	36.67	34.46	-2.21	1.43	ns
Any cancer (other than skin ca	68	34.65	29.13	-5.52	1.54	***	77.0	83.8	36.02	34.76	-1.27	1.13	ns
Health in General													
Excellent	68	53.50	45.87	-7.63	1.44	***	17.0	97.1	53.60	47.26	-6.34	1.16	***
Very Good	234	49.90	42.47	-7.43	0.92	***	36.6	94.0	49.88	45.17	-4.70	0.62	***
Good	446	41.56	33.42	-8.14	0.68	***	42.4	89.7	41.95	37.27	-4.68	0.51	***
Fair	345	31.65	26.99	-4.66	0.77	***	97.9	84.9	31.88	31.78	-0.10	0.56	ns
Poor	169	24.48	18.80	-5.69	1.17	***	145.4	70.4	24.12	26.69	2.58	0.78	**
Baseline PCS Score													
0-20	88	17.97	17.98	0.01	1.40	ns	-47776.9	72.7	17.97	24.72	6.75	1.03	***
21-30	293	25.73	22.42	-3.31	0.75	***	157.6	80.9	25.81	27.72	1.91	0.49	***
31-40	292	35.18	29.53	-5.66	0.86	***	87.0	85.6	35.22	34.49	-0.74	0.58	ns
41-50	285	45.73	36.51	-9.22	0.83	***	33.2	92.3	45.73	39.57	-6.16	0.58	***
51+	303	54.38	43.71	-10.67	0.84	***	28.5	93.4	54.43	46.80	-7.63	0.56	***

NOTES:

[1]This value is calculated by the formula 1-(Mean Follow-Up PCS for Survivors/Mean Baseline PCS for all)

[2]Percentage of baseline respondents alive as of the followup survey.

[3]SGA refers to a small geographic area selected for sampling within the given state.

[4]PGP refers to a primary group practice whose members were selected for sampling. One PGP was selected for each state sampled, excluding Georgia.

*p<0.05, **p<0.01, ***p<0.001; ns=not statistically significant at 0.05 level

Output: a09_sf12

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Appendix - Table 2

MCS Change Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up
Medicare Fee-For-Service Health Outcomes Survey, Using SF-12 Scores Only

	MCS Follow-Up Respondents	Mean Baseline MCS Score	Standard Error of the Mean	Mean Follow-Up MCS Score	Standard Error of the Mean	Mean Difference in MCS	Standard Error of the Mean Difference	P
ALL	5,401	50.90	0.14	50.38	0.14	-0.52	0.13	***
Site								
National	473	48.89	0.51	49.17	0.51	0.27	0.46	ns
PA SGA[1]	488	49.58	0.48	48.86	0.50	-0.72	0.45	ns
GA SGA	480	50.61	0.48	49.40	0.51	-1.21	0.45	**
WI SGA	662	50.98	0.40	50.64	0.41	-0.34	0.37	ns
AZ SGA	474	51.72	0.45	51.07	0.47	-0.66	0.45	ns
WA SGA	494	50.91	0.44	51.31	0.46	0.40	0.43	ns
A PGP[2]	569	50.07	0.43	48.85	0.45	-1.22	0.43	**
B PGP	647	51.26	0.39	51.04	0.38	-0.22	0.38	ns
C PGP	573	52.96	0.38	51.77	0.39	-1.19	0.39	**
D PGP	541	51.50	0.43	51.24	0.44	-0.26	0.42	ns
Sex								
Male	2,172	51.38	0.21	50.89	0.22	-0.49	0.21	*
Female	3,229	50.57	0.18	50.04	0.19	-0.54	0.17	**
Race								
White	5,075	51.11	0.14	50.58	0.14	-0.52	0.14	***
Black	202	47.41	0.76	47.15	0.82	-0.26	0.76	ns
Asian	29	50.48	1.76	46.76	1.85	-3.72	2.19	ns
Hispanic	30	39.72	2.09	42.96	2.12	3.23	2.31	ns
North American Native	8	48.73	3.25	43.54	4.96	-5.19	5.83	ns
Other	47	49.92	1.80	49.45	1.52	-0.47	1.70	ns
Unknown	10	56.17	1.74	53.30	2.98	-2.87	3.17	ns
Original Reason for Entitlement								
Aged	4,661	51.99	0.14	51.35	0.14	-0.64	0.14	***
Disabled	734	44.01	0.45	44.19	0.44	0.19	0.43	ns
ESRD	3	53.41	3.65	60.26	0.86	6.85	3.12	ns
ESRD and Disabled	3	40.33	6.84	48.84	3.82	8.51	4.73	ns
Medicaid Status								
No Medicaid	4,922	51.52	0.14	50.88	0.14	-0.64	0.14	***
Medicaid	479	44.49	0.52	45.19	0.54	0.69	0.54	ns
Age								
Under 65	438	41.76	0.57	42.37	0.58	0.61	0.53	ns
65-74	2,338	52.51	0.19	52.12	0.20	-0.39	0.19	*
75-84	2,107	51.47	0.21	50.70	0.22	-0.77	0.21	***
85 and Older	518	49.01	0.45	47.97	0.46	-1.04	0.46	*
Household Income								
Less than $10,000	611	45.81	0.46	45.99	0.46	0.19	0.43	ns
$10,000-$19,999	903	50.07	0.35	49.63	0.36	-0.43	0.33	ns
$20,000-$49,999	1,615	52.20	0.23	51.73	0.24	-0.47	0.23	*
$50,000 or more	542	53.53	0.36	53.20	0.36	-0.33	0.35	ns
Missing/No Response	1,730	51.09	0.25	50.17	0.25	-0.92	0.25	***

Appendix - Table 2

MCS Change Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up
Medicare Fee-For-Service Health Outcomes Survey, Using SF-12 Scores Only

	MCS Follow-Up Respondents	Mean Baseline MCS Score	Standard Error of the Mean	Mean Follow-Up MCS Score	Standard Error of the Mean	Mean Difference in MCS	Standard Error of the Mean Difference	P
Education								
Not a HS Graduate	1,170	48.52	0.31	47.87	0.32	-0.65	0.31	*
High School Graduate or GED	1,569	51.02	0.25	50.91	0.26	-0.10	0.24	ns
Some College or 2 year degree	935	51.73	0.32	51.26	0.33	-0.47	0.29	ns
4 year college graduate	390	53.01	0.47	52.52	0.48	-0.50	0.41	ns
More than a 4 year college degree	442	53.45	0.40	53.33	0.40	-0.13	0.39	ns
Missing/No Response	895	50.74	0.36	49.41	0.37	-1.33	0.39	***
Marital Status								
Not Married	1,832	49.15	0.25	49.21	0.25	0.06	0.24	ns
Married	2,704	52.09	0.18	51.46	0.19	-0.64	0.17	***
Missing/No Response	865	50.86	0.37	49.48	0.37	-1.38	0.39	***
Home Ownership								
Owned by Beneficiary or Family Member	3,632	51.63	0.16	51.25	0.16	-0.37	0.15	*
Not Owned by Beneficiary or Family Member	776	47.58	0.41	47.45	0.41	-0.14	0.38	ns
Missing/No Response	993	50.81	0.34	49.46	0.35	-1.35	0.36	***
Retirement Community								
Yes	721	50.82	0.38	50.29	0.38	-0.52	0.36	ns
No	3,681	50.99	0.16	50.63	0.17	-0.36	0.15	*
Missing/No Response	999	50.60	0.34	49.52	0.35	-1.08	0.36	**
Chronic Conditions								
Hypertenstion or high blood pressure	2,362	50.56	0.21	50.30	0.21	-0.26	0.20	ns
Angina pectoris or coroary artery disease	785	49.41	0.36	49.05	0.38	-0.36	0.37	ns
Congestive heart failure	318	47.75	0.58	48.19	0.62	0.44	0.60	ns
Acute myocardial infarction or heart attack	520	49.58	0.44	49.31	0.45	-0.27	0.43	ns
Other heart conditions	1,103	49.57	0.32	49.10	0.33	-0.47	0.30	ns
Stroke	383	47.95	0.53	47.43	0.57	-0.52	0.58	ns
Emphysema, asthma, or COPD	564	48.76	0.46	48.36	0.47	-0.39	0.43	ns
Crohn's disease, ulcerative colitis, or inflamatory bowel disease	323	47.47	0.62	47.09	0.63	-0.38	0.55	ns
Arthritis of the hip or knee	1,856	49.77	0.24	49.51	0.25	-0.26	0.22	ns
Arthritis of the hand or wrist	1,629	49.55	0.26	49.44	0.27	-0.10	0.24	ns
Sciatica	1,199	49.02	0.31	48.75	0.32	-0.27	0.30	ns

Appendix - Table 2

MCS Change Scores by Subsample and Demographic Characteristics for Baseline and Follow-Up
Medicare Fee-For-Service Health Outcomes Survey, Using SF-12 Scores Only

	MCS Follow-Up Respondents	Mean Baseline MCS Score	Standard Error of the Mean	Mean Follow-Up MCS Score	Standard Error of the Mean	Mean Difference in MCS	Standard Error of the Mean Difference	P
Diabetes, high blood sugar, or sugar in the urine	749	49.22	0.38	48.90	0.40	-0.31	0.38	ns
Any cancer (other than skin cancer)	773	50.99	0.36	50.99	0.36	0.00	0.34	ns
Health in General								
Excellent	275	56.39	0.39	55.52	0.42	-0.86	0.46	ns
Very Good	1,233	55.50	0.19	54.55	0.22	-0.95	0.23	***
Good	2,169	52.40	0.20	51.65	0.21	-0.76	0.21	***
Fair	1,369	46.46	0.28	46.31	0.29	-0.15	0.29	ns
Poor	355	38.57	0.59	39.83	0.60	1.26	0.63	*
Baseline MCS Score								
0-30	254	25.90	0.24	36.03	0.72	10.13	0.74	***
31-40	735	36.26	0.10	42.07	0.37	5.80	0.37	***
41-50	1,106	45.98	0.09	47.47	0.28	1.49	0.29	***
51-60	2,563	56.29	0.05	53.74	0.16	-2.56	0.15	***
61+	743	62.61	0.07	56.24	0.31	-6.37	0.32	***

NOTES:

[1] SGA refers to a small geographic area selected for sampling within the given state.

[2] PGP refers to a primary group practice whose members were selected for sampling. One PGP was selected for each state sampled, excludi Georgia.

*p<0.05, **p<0.01, ***p<0.001; ns=not statistically significant at 0.05 level

Output: a11_sf12

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Appendix - Table 3

Consistency of Condition Reports and PCS Outcomes

Hypertenstion or high blood pressure Kappa= 0.75

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	37.75	35.73	2,073	37.29	35.94	202
No in 1998	40.41	37.55	341	41.44	39.72	1,761

Angina pectoris or coronary artery disease Kappa= 0.72

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	34.46	32.28	596	34.13	33.69	143
No in 1998	36.17	33.08	219	40.88	38.96	3,336

Congestive heart failure Kappa= 0.59

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	29.79	27.86	214	32.60	32.41	78
No in 1998	32.96	28.17	184	40.60	38.74	3,794

Acute myocardial infarction or heart attack Kappa= 0.73

	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	34.34	32.22	399	34.88	35.19	102
No in 1998	35.04	31.58	148	40.44	38.52	3,620

Other heart conditions Kappa= 0.60

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	35.02	33.09	755	35.50	34.38	284
No in 1998	38.46	35.57	370	41.15	39.28	2,907

Stroke Kappa= 0.72

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	31.85	29.57	297	33.25	33.98	66
No in 1998	34.33	30.51	133	40.39	38.47	3,824

Emphysema, asthma, or COPD Kappa= 0.71

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	33.37	31.96	434	34.20	33.79	106
No in 1998	38.40	33.88	186	40.51	38.54	3,599

Appendix - Table 3 (continued)

Consistency of Condition Reports and PCS Outcomes

Crohn's disease, ulcerative colitis, or inflamatory bowel disease Kappa= 0.54

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	33.39	32.37	175	33.49	32.67	137
No in 1998	35.63	32.29	123	40.16	38.16	3,844

Arthritis of the hip or knee Kappa= 0.63

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	33.14	31.42	1,449	37.45	37.24	326
No in 1998	39.77	35.61	456	43.99	42.01	2,140

Arthritis of the hand or wrist Kappa= 0.60

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	34.36	32.49	1,194	36.39	36.40	368
No in 1998	38.99	35.58	441	42.71	40.63	2,348

Sciatica Kappa= 0.56

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	33.58	32.19	782	35.05	34.52	355
No in 1998	37.38	32.94	387	42.13	40.14	2,787

Diabetes, high blood sugar, or sugar in the urine Kappa= 0.80

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	34.84	32.19	634	37.22	35.93	91
No in 1998	37.31	34.41	165	40.44	38.60	3,483

Any cancer (other than skin cancer) Kappa= 0.80

	Yes in 2000			No in 2000		
	PCS 1998	PCS 2000	No. in 2000	PCS 1998	PCS 2000	No. in 2000
Yes in 1998	37.74	35.49	654	34.40	34.92	92
No in 1998	40.00	36.04	170	39.86	37.99	3,461

Output: a14mean and a15kappa

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Appendix - Table 4

Mean PCS and MCS Scores by Mode of Survey Administration

	N Obs	Mean	Std Dev	Minimum	Maximum
Baseline Survey					
Physical Component Score (PCS)					
Mail	4,828	39.22	12.16	6.75	66.05
Phone	637	41.11	12.09	12.08	62.55
Mental Component Score (MCS)					
Mail	4,828	51.36	10.59	1.46	74.51
Phone	637	52.39	10.19	12.62	70.63
Follow-up Survey					
Physical Component Score (PCS)					
Mail	5,125	37.17	12.28	7.30	67.60
Phone	340	37.61	12.25	6.93	59.19
Mental Component Score (MCS)					
Mail	5,125	51.14	10.87	9.29	74.27
Phone	340	50.81	10.88	16.38	71.95

Output: joinx02c, joinx02e

SOURCE: RTI International analysis of the baseline and follow-up Medicare fee-for-service Health Outcomes Survey.

Appendix - Table 5

Preliminary Regression Models Used to Develop The Final PCS Regression Model in Table 4-1

Variable	Coefficient	*t* statistic	*p* value
Intercept	10.23	6.17	0.00
Baseline PCS	0.67	18.30	0.00
Age < 65	-0.90	-0.55	0.58
Age 75-84	-2.49	-2.48	0.01
Age 85+	-6.64	-4.76	0.00

$R^2 = 0.4453$

OUTPUT: joinx01b

Variable	Coefficient	*t* statistic	*p* value
Intercept	9.70	5.56	0.00
Baseline PCS	0.69	18.23	0.00
Age < 65	-0.75	-4.58	0.65
Age 75-84	-2.30	-2.30	0.02
Age 85+	-6.38	-4.55	0.00
Baseline SF12	-1.58	-1.23	0.22
Follow-up SF12	2.72	0.80	0.43
Baseline Phone	-2.28	-1.40	0.16
Follow-up Phone	2.49	0.75	0.45
Baseline Proxy	-1.00	-0.74	0.46
Follow-up Proxy	2.33	1.92	0.05

$R^2 = 0.4616$

OUTPUT: joinx01b

Appendix - Table 5 (continued)

Preliminary Regression Models Used to Develop
The Final PCS Regression Model in Table 4-1

Variable	Coefficient	*t* statistic	*p* value
Intercept	9.78	5.61	0.00
Baseline PCS	0.68	18.21	0.00
Age < 65	-0.63	-0.39	0.70
Age 75-84	-2.31	-2.30	0.02
Age 85+	-6.40	-4.56	0.00
Baseline SF12	-2.67	-2.61	0.01
Follow-up Phone	4.15	2.16	0.03
Baseline Proxy	-1.11	-0.82	0.41
Follow-up Proxy	2.32	1.92	0.06

$R^2 = 0.4589$

OUTPUT: joinx01b

Variable	Coefficient	*t* statistic	*p* value
Intercept	9.59	5.53	0.00
Baseline PCS	0.68	18.22	0.00
Age < 65	-1.06	-0.65	0.52
Age 75-84	-2.41	-2.42	0.02
Age 85+	-6.53	-4.67	0.00
Follow-up SF12	4.75	2.38	0.02
Baseline Phone	-3.69	-2.64	0.01
Baseline Proxy	-0.82	-0.61	0.54
Follow-up Proxy	2.29	1.89	0.06

$R^2 = 0.4594$

OUTPUT: joinx01b

Appendix - Table 5 (continued)

Preliminary Regression Models Used to Develop The Final PCS Regression Model in Table 4-1

Variable	Coefficient	*t* statistic	*p* value
Intercept	9.50	5.46	0.00
Baseline PCS	0.69	18.27	0.00
Age < 65	-0.91	-0.55	0.58
Age 75-84	-2.37	-2.37	0.02
Age 85+	-6.52	-4.66	0.00
Baseline Phone	-3.49	-2.71	0.01
Follow-up Phone	4.64	2.38	0.02
Baseline Proxy	-0.88	-0.65	0.51
Follow-up Proxy	2.34	1.93	0.05

$R^2 = 0.4594$

OUTPUT: joinx02

Variable	Coefficient	*t* statistic	*p* value
Intercept	11.08	6.20	0.00
Baseline PCS	0.67	17.63	0.00
Age < 65	-0.87	-0.54	0.59
Age 75-84	-2.38	-2.41	0.02
Age 85+	-6.63	-4.79	0.00
Baseline Phone	-4.02	-3.13	0.00
Follow-up Phone	4.46	2.31	0.02
Baseline Proxy	-0.97	-0.73	0.47
Follow-up Proxy	2.09	1.74	0.08
Angina/CAD	-4.07	-3.34	0.00

$R^2 = 0.4705$

OUTPUT: joinx02

Appendix - Table 5 (continued)

Preliminary Regression Models Used to Develop The Final PCS Regression Model in Table 4-1

Variable	Coefficient	*t* statistic	*p* value
Intercept	10.43	5.82	0.00
Baseline PCS	0.67	17.40	0.00
Age < 65	-1.01	-0.62	0.54
Age 75-84	-2.34	-2.34	0.02
Age 85+	-6.54	-4.69	0.00
Baseline Phone	2.23	-2.93	0.00
Follow-up Phone	4.88	2.51	0.01
Baseline Proxy	-0.60	-0.44	0.66
Follow-up Proxy	2.18	1.80	0.07
CHF	-3.78	-2.06	0.04

$R^2 = 0.4637$

OUTPUT: joinx02

Variable	Coefficient	*t* statistic	*p* value
Intercept	10.10	5.71	0.00
Baseline PCS	0.68	17.99	0.00
Age < 65	-0.79	-0.49	0.63
Age 75-84	-2.40	-2.40	0.02
Age 85+	-6.52	-4.68	0.00
Baseline Phone	-3.78	-2.92	0.00
Follow-up Phone	4.80	2.47	0.01
Baseline Proxy	-0.88	-0.66	0.51
Follow-up Proxy	2.27	1.88	0.06
AMI	-2.51	-1.78	0.08

$R^2 = 0.4626$

OUTPUT: joinx02

Appendix - Table 5 (continued)

Preliminary Regression Models Used to Develop The Final PCS Regression Model in Table 4-1

Variable	Coefficient	*t* statistic	*p* value
Intercept	10.45	5.80	0.00
Baseline PCS	0.67	17.75	0.00
Age < 65	-0.87	-0.53	0.59
Age 75-84	-2.56	-2.56	0.01
Age 85+	-6.86	-4.88	0.00
Baseline Phone	-3.83	-2.96	0.00
Follow-up Phone	4.63	2.38	0.02
Baseline Proxy	-0.83	-0.62	0.54
Follow-up Proxy	2.37	1.96	0.05
Diabetes	-2.51	-1.95	0.05

$R^2 = 0.4633$

OUTPUT: joinx02

Variable	Coefficient	*t* statistic	*p* value
Intercept	11.09	6.19	0.00
Baseline PCS	0.67	17.60	0.00
Age < 65	-0.86	-0.53	0.60
Age 75-84	-2.38	-2.41	0.02
Age 85+	-6.63	-4.78	0.00
Baseline Phone	-4.03	-3.13	0.00
Follow-up Phone	4.47	2.31	0.02
Baseline Proxy	-0.97	-0.73	0.47
Follow-up Proxy	2.10	1.74	0.08
Angina/CAD	-4.00	-2.81	0.01
AMI	-0.16	-0.10	0.92

$R^2 = 0.4705$

OUTPUT: joinx02a

Appendix - Table 5 (continued)

Preliminary Regression Models Used to Develop The Final PCS Regression Model in Table 4-1

Variable	Coefficient	*t* statistic	*p* value
Intercept	11.62	6.35	0.00
Baseline PCS	0.66	17.31	0.00
Age < 65	-0.85	-0.52	0.60
Age 75-84	-2.52	-2.53	0.01
Age 85+	-6.86	-4.92	0.00
Baseline Phone	-4.21	-3.26	0.00
Follow-up Phone	4.46	2.31	0.02
Baseline Proxy	-0.93	-0.70	0.49
Follow-up Proxy	2.14	1.78	0.08
Angina/CAD	-3.75	-3.02	0.00
Diabetes	-1.76	-1.35	0.18

$R^2 = 0.4723$

OUTPUT: joinx02a

Appendix - Table 6

Preliminary Regression Models Used to Develop The Final MCS Regression Model in Table 4-2

Variable	Coefficient	*t* statistic	*p* value
Intercept	21.91	10.27	0.00
Baseline MCS	0.57	14.50	0.00
Age < 65	-4.95	-2.98	0.00
Age 75-84	-2.13	-2.19	0.03
Age 85+	-3.58	-2.70	0.01

$R^2 = 0.3506$

OUTPUT: joinx01b

Variable	Coefficient	*t* statistic	*p* value
Intercept	21.81	10.05	0.00
Baseline MCS	0.57	14.64	0.00
Age < 65	-4.61	-2.80	0.01
Age 75-84	-1.75	-1.82	0.07
Age 85+	-2.84	-2.15	0.03
Baseline SF12	-0.42	-0.34	0.73
Follow-up SF12	1.03	0.31	0.75
Baseline Phone	-2.94	-1.88	0.06
Follow-up Phone	7.11	2.22	0.03
Baseline Proxy	-3.70	-2.88	0.00
Follow-up Proxy	1.52	1.30	0.19

$R^2 = 0.3838$

OUTPUT: joinx01b

Appendix - Table 6 (continued)

Preliminary Regression Models Used to Develop The Final MCS Regression Model in Table 4-2

Variable	Coefficient	*t* statistic	*p* value
Intercept	22.04	10.15	0.00
Baseline PCS	0.57	14.53	0.00
Age < 65	-4.56	-2.77	0.01
Age 75-84	-1.76	-1.82	0.07
Age 85+	-2.85	-2.14	0.03
Baseline SF12	-1.85	-1.88	0.06
Follow-up Phone	7.27	3.93	0.00
Baseline Proxy	-3.80	-2.96	0.00
Follow-up Proxy	1.48	1.27	0.20

$R^2 = 0.3796$

OUTPUT: joinx01b

Variable	Coefficient	*t* statistic	*p* value
Intercept	21.99	10.18	0.00
Baseline PCS	0.57	14.57	0.00
Age < 65	-4.91	-2.99	0.00
Age 75-84	-1.87	-1.95	0.05
Age 85+	-2.89	-2.18	0.03
Follow-up SF12	6.91	3.58	0.00
Baseline Phone	-2.92	-2.38	0.02
Baseline Proxy	-3.57	-2.78	0.01
Follow-up Proxy	1.46	1.25	0.21

$R^2 = 0.3779$

OUTPUT: joinx01b

Appendix - Table 6 (continued)

Preliminary Regression Models Used to Develop
The Final MCS Regression Model in Table 4-2

Variable	Coefficient	*t* statistic	*p* value
Intercept	21.73	10.09	0.00
Baseline MCS	0.57	14.71	0.00
Age < 65	-4.64	-2.84	0.00
Age 75-84	-1.76	-1.84	0.07
Age 85+	-2.88	-2.18	0.03
Baseline Phone	-3.27	-2.65	0.01
Follow-up Phone	7.92	4.22	0.00
Baseline Proxy	-3.67	-2.87	0.00
Follow-up Proxy	1.52	1.31	0.19

$R^2 = 0.3835$

OUTPUT: joinx02

Variable	Coefficient	*t* statistic	*p* value
Intercept	22.32	10.19	0.00
Baseline MCS	0.57	14.54	0.00
Age < 65	-4.64	-2.84	0.00
Age 75-84	-1.80	-1.88	0.06
Age 85+	-2.89	-2.19	0.03
Baseline Phone	-3.61	-2.88	0.00
Follow-up Phone	8.10	4.31	0.00
Baseline Proxy	-3.74	-2.93	0.00
Follow-up Proxy	1.56	1.34	0.18
Other Heart Disease	-1.56	-1.43	0.15

$R^2 = 0.3859$

OUTPUT: joinx02

Appendix - Table 6 (continued)

Preliminary Regression Models Used to Develop The Final MCS Regression Model in Table 4-2

Variable	Coefficient	*t* statistic	*p* value
Intercept	22.34	10.18	0.00
Baseline MCS	0.57	14.44	0.00
Age < 65	-4.87	-2.96	0.00
Age 75-84	-1.89	-1.97	0.05
Age 85+	-3.01	-2.28	0.02
Baseline Phone	-3.46	-2.79	0.01
Follow-up Phone	8.04	4.29	0.00
Baseline Proxy	-3.64	-2.85	0.00
Follow-up Proxy	1.59	1.37	0.17
COPD/Lung Disease	-2.02	-1.39	0.17

$R^2 = 0.3858$

OUTPUT: joinx02

Variable	Coefficient	*t* statistic	*p* value
Intercept	24.03	9.51	0.00
Baseline MCS	0.05	12.19	0.00
Age < 65	-4.60	-2.82	0.01
Age 75-84	-1.91	-1.99	0.05
Age 85+	-2.99	-2.26	0.02
Baseline Phone	-3.84	-3.03	0.00
Follow-up Phone	8.07	4.30	0.00
Baseline Proxy	-3.48	-2.70	0.01
Follow-up Proxy	1.46	1.26	0.21
Other Heart Disease	-1.48	-1.36	0.17
Depression	-1.91	-1.35	0.18

$R^2 = 0.3880$

OUTPUT: joinx02a

Appendix - Table 6 (continued)

Preliminary Regression Models Used to Develop The Final MCS Regression Model in Table 4-2

Variable	Coefficient	*t* statistic	*p* value
Intercept	24.04	9.50	0.00
Baseline MCS	0.54	12.13	0.00
Age < 65	-4.81	-2.93	0.00
Age 75-84	-1.99	-2.07	0.04
Age 85+	-3.10	-2.34	0.02
Baseline Phone	-3.70	-2.95	0.00
Follow-up Phone	8.01	4.27	0.00
Baseline Proxy	-3.39	-2.63	0.01
Follow-up Proxy	1.49	1.28	0.20
COPD/Lung Disease	-1.19	-1.31	0.19
Depression	-1.91	-1.35	0.18

$R^2 = 0.3878$

OUTPUT: joinx02a