

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.
Project Director

Janet B. Mitchell, Ph.D.
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, Zhooan 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

	<u>Page</u>
1.0 Introduction and Background.....	1-1
1.1 Introduction.....	1-1
1.2 Methodological Approach	1-6
1.3 Organization of Report	1-12
2.0 Description of the Baseline and Follow-up Medicare Fee-for-Service Health Outcomes Survey (HOS).....	2-1
2.1 Background of the Medicare Fee-for-Services Health Outcomes Survey	2-2
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	2-7
2.3 Changes in Sociodemographic and Health Status Characteristics of Respondents at Baseline and Follow-up.....	2-16
2.4 Scoring of PCS and MCS at Baseline and Follow-up	2-22
2.5 Differences in Mean Physical and Mental Health Between Baseline and Follow-up	2-24
2.6 Usual Source of Care at Baseline and Follow-up.....	2-39
2.7 Summary	2-41
3.0 Imputing PCS and MCS Scores for Respondents Who Died Between Baseline and Follow-up.....	3-1
3.1 Introduction.....	3-1
3.2 Associating the Utility Score for Death with PCS and MCS Scores.....	3-2
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	3-6
3.4 Summary	3-17
4.0 Multivariate Statistical Analysis of Predicted versus Actual Health Status at Two-Year Follow-up	4-1
4.1 Overview and Hypotheses	4-1
4.2 Statistical Models for Predicting Follow-up Health Status	4-2
4.3 Specification of the Standard Care Model for Predicting Follow-up Health Status	4-5
4.4 Comparisons of Expected and Actual Two-Year Follow-up in Health Status by Cohort.....	4-17
4.5 Summary	4-26

Table of Contents
(continued)

	<u>Page</u>
5.0 Comparison to the Statistical Methods used by the Health Assessment Laboratory for Analysis of Two-Year Follow-up Health Status for Medicare+Choice Beneficiaries	5-1
5.1 Assigning Respondents to Change Categories	5-1
5.2 Statistical Methods for Analyzing Two-Year Changes in Health Status.....	5-3
5.3 Results of the Statistical Analysis.....	5-6
5.4 Summary	5-19
6.0 Summary of Findings and Discussion	6-1
6.1 Summary of Findings.....	6-1
6.2 Discussion and Policy Implications.....	6-8
References.....	R-1
Appendix	A-1

Table of Tables and Figures

	<u>Page</u>
 Chapter 2	
Table 2-1	Survey Response Rates to Baseline and Follow-up Medicare Fee-For-Service Health Outcomes Survey, by Sample Cohort2-9
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).....2-10
Table 2-3	Survey Response Rates to Baseline and Follow-up Medicare Fee-For-Service Health Outcomes Survey, by Baseline Sociodemographic Characteristics.....2-11
Table 2-4	Demographic and Eligibility Characteristics of Baseline and Follow-up Respondents to the Medicare Fee-for-Service Health Outcomes Survey2-17
Table 2-5	Sociodemographic Characteristics of Baseline and Follow-up Respondents to the Medicare Fee-For-Service Health Outcome Survey.....2-18
Table 2-6	Scoring Method for Baseline and Follow-Up Medicare Fee-For-Service Health Outcomes Survey2-23
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 (follow-up PCS of decedents = 0)2-27
Table 2-8	Difference in Mean MCS Score Between Baseline and Follow-up by Subsample and Baseline Demographic Characteristics.....2-31
Table 2-9	Agreement Between Assignment of Beneficiaries to Physician Group Practices (PGPs) and Beneficiary Perceptions of Doctor Visits.....2-40
 Chapter 3	
Table 3-1	HALex Utility Scores and Sample Size by Activity Limitation and Perceived Health3-5
Figure 3-1	Regression Results for PCS and MCS Scores on HALex Utility Scores ..3-7
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)3-8

Table of Tables and Figures (continued)

Page

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)	3-13
-----------	---	------

Chapter 4

Table 4-1	Regression Model for Predicting PCS Scores at Follow-up	4-11
Table 4-2	Regression Model for Predicting MCS Scores at Follow-up	4-12
Table 4-3	Mean Values of Independent Variables for PCS & MCS Prediction Models by Cohort	4-13
Figure 4-1	Mean Expected Versus Actual Follow-Up PCS and MCS Scores by Cohort	4-21
Table 4-4	Comparisons of Actual and Expected Mean PCS and MCS Follow-up Scores by Cohort	4-23

Chapter 5

Table 5-1	Expected and Actual Outcomes by Cohort Using HAL Software and Fee-for-Service Health Outcomes Study Data	5-7
Table 5-2	Expected versus Actual PCS Outcomes and t Statistics by Cohort Using HAL Software and Fee-for-Service Health Outcomes Study Data	5-10
Table 5-3	Expected versus Actual MCS Outcomes and t Statistics by Cohort Using HAL Software and Fee-for-Service Health Outcomes Study Data	5-13
Table 5-4	Expected versus Actual MCS Outcomes for t Statistics by Cohort with Values Imputed for Death Using HAL Software and Fee-for-Service Health Outcomes Study Data	5-16

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

Cohort	Baseline					Follow-Up				
	Sampled			Baseline	Response	Follow-Up		Death	PCS	MCS
	<u>Beneficiaries</u>	<u>Deaths¹</u>	<u>Eligibles</u>	<u>Respondents</u>	<u>Rate</u>	<u>Respondents</u>	<u>Deaths²</u>	<u>Rate</u>	<u>Rate³</u>	<u>Rate³</u>
	(1)	(2)	(3)=(1)-(2)	(4)	(5)=(4)/(3)	(6)	(7)	(8)=(7)/(4)	(9)=[(6)+(7)]/(4)	(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

¹Deaths 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)

²Deaths 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)

³The 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.

⁴SGA refers to a small geographic area selected for sampling within the given state

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

*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% level)

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	Deaths ¹	Eligibles	Baseline	Response	Follow-up	Deaths ²	Death	PCS	MCS
	Beneficiaries	(2)	(3)=(1)-(2)	Respondents	Rate	Respondents	(7)	Rate	Response	Response
	(1)	(2)	(3)=(1)-(2)	(4)	(5)=(4)/(3)	(6)	(7)	(8)=(7)/(4)	(9)=[(6)+(7)]/(4)	(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	2,615*	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

¹Deaths 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)

²Deaths 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)

³The 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**

	<u>Baseline</u>		<u>Follow-up</u>			
	<u>Respondents</u>	<u>Respondents</u>	<u>Deaths¹</u>	<u>Death Rate</u>	<u>PCS Response Rate²</u>	<u>MCS Response Rate²</u>
	(1)	(2)	(3)	(4) = (3)/(1)	(5) = [(2)+(3)]/(1)	(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			
	<u>Respondents</u> (1)	<u>Respondents</u> (2)	<u>Deaths</u> ¹ (3)	<u>Death</u> <u>Rate</u> (4) = (3)/(1)	<u>PCS</u> <u>Response Rate</u> ² (5) = [(2)+(3)]/(1)	<u>MCS</u> <u>Response Rate</u> 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 reported³						
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 do³						
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**

	<u>Baseline</u>		<u>Follow-up</u>			
	<u>Respondents</u> (1)	<u>Respondents</u> (2)	<u>Deaths</u> ¹ (3)	<u>Death</u> <u>Rate</u> (4) = (3)/(1)	<u>PCS</u> <u>Response Rate</u> ² (5) = [(2)+(3)]/(1)	<u>MCS</u> <u>Response Rate</u> 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

¹Deaths 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)

²The 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.

³A 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

¹Includes only respondents to the follow-up Survey. Decedents are excluded.

Hence the response rate is calculated two ways for comparison. The Alternate MCS Response Rate does not include deaths.

²Categories 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

¹Includes only follow-up respondents.

²A 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

Site	Scoring Method ¹							
	SF-12 Only		SF-12/SF-36		SF-36 Only		SF-36/SF-12	
	#	%	#	%	#	%	#	%
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

¹In 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 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	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	Mean Baseline	Mean Follow-Up	Mean Difference	Standard Error of the Mean	Percent of PCS Change Due to PCS=0 for	Percent of PCS Follow-Up	Mean Baseline	Mean Follow-Up	Mean Difference in	Standard Error of the Mean		
	Respondents	PCS Score	PCS Score	in PCS	Difference	P	Decedents ¹	Respondents ²	PCS	PCS	Scores	Difference	P
Chronic Conditions													
Hypertension 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 coronary 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 inflammatory 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 Reported⁵													
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 Do⁵													
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	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:

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

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

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

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

⁵A 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

	<u>MCS Follow-Up Respondents</u>	<u>Mean Baseline MCS Score</u>	<u>Standard Error of the Mean</u>	<u>Mean Follow-Up MCS Score</u>	<u>Standard Error of the Mean</u>	<u>Mean Difference in MCS</u>	<u>Standard Error of the Mean Difference</u>	<u>P</u>
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

	<u>MCS Follow-Up Respondents</u>	<u>Mean Baseline MCS Score</u>	<u>Standard Error of the Mean</u>	<u>Mean Follow-Up MCS Score</u>	<u>Standard Error of the Mean</u>	<u>Mean Difference in MCS</u>	<u>Standard Error of the Mean Difference</u>	<u>P</u>
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

	<u>MCS Follow-Up Respondents</u>	<u>Mean Baseline MCS Score</u>	<u>Standard Error of the Mean</u>	<u>Mean Follow-Up MCS Score</u>	<u>Standard Error of the Mean</u>	<u>Mean Difference in MCS</u>	<u>Standard Error of the Mean Difference</u>	<u>P</u>
Chronic Conditions								
Hypertension or high blood pressure	2,393	51.26	0.21	51.03	0.22	-0.23	0.19	ns
Angina pectoris or coronary 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 inflammatory 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

	<u>MCS Follow-Up Respondents</u>	<u>Mean Baseline MCS Score</u>	<u>Standard Error of the Mean</u>	<u>Mean Follow-Up MCS Score</u>	<u>Standard Error of the Mean</u>	<u>Mean Difference in MCS</u>	<u>Standard Error of the Mean Difference</u>	<u>P</u>
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:

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

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

³A 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 zero¹.

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: $[1 - (\text{ratio of mean difference in scores for survivors to mean difference in scores for survivors and decedents})] * 100$.

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

Assigned to PGP ¹	BASELINE					FOLLOW-UP				
	Respondents ²	'Most Doctor Visits Last Year at This Clinic?' ³			Percent Agreement ⁴	Respondents ⁵	'Most Doctor Visits Last Year at This Clinic?' ³			Percent Agreement ⁴
		Yes	No	Missing			Yes	No	Missing	
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

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

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

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.