March 2005

Implementation of Medicare CAHPS® Fee-for-Service Survey

Final Report for the 2003 Survey

Prepared for

Edward S. Sekscenski
Centers for Medicare & Medicaid Services
7500 Security Boulevard
Mail Stop S1-15-03
Baltimore, MD 21244

Prepared by

RTI International
and
RAND

Under subcontract to
Center for Health Systems Research and Analysis
University of Wisconsin-Madison

Contract No. 500-95-0061
RTI Project Number 07903.500

RTI
INTERNATIONAL

IMPLEMENTATION OF MEDICARE CAHPS®
FEE-FOR-SERVICE SURVEY

FINAL REPORT FOR THE 2003 SURVEY

RTI Project Director: Larry N. Campbell

Federal Project Officer: Edward S. Sekscenski

Prepared by

RTI International[*]
and
RAND

Under subcontract to
Center for Health Systems Research and Analysis
University of Wisconsin-Madison

CMS Contract No. 500-95-0061

March 2005

This project was funded by the Centers for Medicare & Medicaid Services under contract no. 500-95-0061. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. RTI assumes responsibility for the accuracy and completeness of the information contained in this report.

---

[*] RTI International is a trade name of Research Triangle Institute.

LIST OF AUTHORS

1. Introduction
    Jeffrey S. Laufenberg, M.S., RTI

2. 2003 Medicare CAHPS® Fee-for-Service Survey Questionnaire
    Judith T. Lynch, B.A., RTI

3. Sample Selection, Weighting, and Data Presentation
    Jeffrey S. Laufenberg, M.S., RTI
    Vincent G. Iannacchione, M.S., RTI
    Lisa A. Carpenter, B.S., RTI
    Marc N. Elliott, Ph.D., RAND

4. Data Collection
    Linda L. Dimitropoulos, Ph.D., RTI
    Carol Prindle, Ph.D., RTI

5. Costs and Benefits of Improving Response Rates in the MFFS Survey
    Larry N. Campbell, M.S., RTI
    Linda L. Dimitropoulos, Ph.D., RTI
    G. Gordon Brown, Ph.D., RTI
    Lisa A. Carpenter, B.S., RTI

6. Case-Mix Adjustment
    Marc N. Elliott, Ph.D., RAND

7. Estimating the Effects of Proxy Respondents in Medicare CAHPS®
    Marc N. Elliott, Ph.D., RAND
    Kelly Chong, UCLA School of Public Health

8. Subgroup Analyses
    Shulamit L. Bernard, Ph.D., RTI
    Erica R. Brody, M.P.H., RTI
    Nathan D. West, M.P.A., RTI

CONTENTS

List of Figures

List of Tables

EXECUTIVE SUMMARY

This report summarizes the methodology and findings of the 2003 Medicare Consumer Assessment of Health Plans Study (CAHPS®) Fee-for-Service (MFFS) Survey conducted for the Centers for Medicare & Medicaid Services (CMS) by RTI, with the assistance of RAND, NCS Pearson, and Discovery Research Group (DRG). The work was performed under subcontract to the Center for Health Systems Research and Analysis at the University of Wisconsin-Madison (UW), as part of UW's contract with CMS. More detailed information on many of the topics presented in this report is available in the individual project reports prepared for the 2003 CAHPS® MFFS survey.

2003 Medicare CAHPS® Fee-for-Service Survey Questionnaire

The CAHPS® surveys were developed using comprehensive reviews of the existing literature, focus groups with consumers, cognitive testing of survey content and question wording, and field testing of preliminary versions of individual items. A set of core items was developed for all consumers, and certain items were targeted for special subpopulations, such as Medicare managed care enrollees. The CAHPS® items include evaluations (ratings) of care and reports of specific experiences with health plans. This combination of global assessments and reports about different aspects of health plan performance also allows users to link global evaluations with specific information to guide quality improvement efforts.

The MFFS questionnaire is based on the CAHPS® core questionnaire for adult, privately insured populations developed as part of the CAHPS® research project sponsored by the Agency for Healthcare Research and Quality (AHRQ). In 1998, the MFFS project team conducted a field test on a sample of fee-for-service Medicare beneficiaries in five states to test field procedures and evaluate the psychometric performance of standard CAHPS® questions within this population. As a result of this field test, the 12-month recall period for health-related experiences was shortened to 6 months. A 6-month recall period is also used for the Medicare CAHPS® Advantage (MA) Survey (previously known as the Medicare CAHPS® Managed Care [MMC] Survey). A report on the findings of the CAHPS® MFFS field test survey is available electronically from CMS. The MFFS survey has been implemented annually in all 50 states, the District of Columbia, and Puerto Rico since fall 2000, with the Virgin Islands added in 2003.

The MFFS project team worked extensively with CMS and the MA project team during the first 7 months of 2003 to coordinate and discuss changes to the questionnaires that would be used in the 2003 Medicare CAHPS® surveys. As a result of those discussions, CMS and the project teams decided to make a number of changes to the 2003 surveys. The most significant changes were the deletion of three questions in the "Personal Doctor or Nurse" section (about health problems affecting day-to-day activities), the addition of some new questions to the series of questions about prescription drugs, and the addition of two new questions about health promotion advice. A more detailed discussion of the changes made to the 2003 MFFS questionnaire is provided in Section 2 of this report. A copy of the 2003 MFFS questionnaire is included in Appendix A.

Sample Selection, Weighting, and Data Presentation

For the 2003 MFFS survey, the MFFS project team selected a sample of 178,650 fee-for-service Medicare beneficiaries from a sampling frame constructed from the July 2003 version of CMS' Enrollment Database (EDB). The frame comprised approximately 32.0 million persons who were continuously enrolled in fee-for-service Medicare for at least 6 months, did not have a representative payee, were over the age of 18, and resided in any of the 50 states, the District of Columbia, Puerto Rico, or the Virgin Islands.

Prior to sample selection, we constructed local geographic areas (geounits), with each local area consisting of one or more counties. Factors considered for grouping counties included geographic contiguity, Medicare Advantage contract areas, and metropolitan statistical area (MSA) and state boundaries. County samples were then aggregated into geographic areas approximating fee-for-service market areas. In the 2000 MFFS survey, the selection allocation was defined to be 600 beneficiaries from each of the 275 geounits in the United States, and 3,000 beneficiaries from Puerto Rico, for a total sample size of 168,000 beneficiaries in 276 geounits. For the 2001 MFFS survey, no change was made to the number of geounits, but the total sample size was increased to 177,950 beneficiaries. For the 2002 survey, no change was made to the number of geounits or to the total (national) MFFS sample size from the 2001 survey. However, we reallocated the sample for the 2002 survey to provide better power for estimates in counties that experienced significant changes in Medicare managed care enrollment and/or counties with insufficient sample owing to higher than expected nonresponse in the prior MFFS survey. For the 2003 MFFS survey, we added one geounit for the Virgin Islands, bringing the total number of geounits in our study design to 277 and our total sample size to 178,650. In addition, our reallocation of sample between the 2002 and 2003 surveys involved selected counties in five "donor" states, which contributed a proportionate amount of their allocated sample in excess of 330 completes from the previous year, to recipient counties in eight states. The states selected as donors of sample were those with the most effective sample size when compared with MA.

We stratified the MFFS population by county and selected a simple random sample from within each county. We then assigned an initial sampling weight to each selected beneficiary as the inverse of the selection probability, reflecting the differential selection rates used to identify beneficiaries from each county. To reduce the potential biasing effects of differential nonresponse, we post-stratified the initial sampling weights of respondents to sum to 337 separate counts of fee-for-service beneficiaries obtained from the October 2003 version of the EDB, which is the approximate midpoint of data collection. The counts included 277 totals for each of the local geounits in the United States, Puerto Rico, and the Virgin Islands, as well as 60 totals formed by the intersection of the age, gender, race, and dual Medicare/Medicaid eligibility factors.

Using responses from the 2003 survey, we evaluated the effect of the MFFS analysis weights on the accuracy of the survey estimates by comparing the mean square errors (MSEs) of weighted estimates to the corresponding MSEs of unweighted estimates. The MSE, defined as the sum of the bias squared and the variance, is used to measure the combined effect of bias and variance on the survey estimates. We assumed that the weighted estimates represent unbiased estimates because of the bias reduction and improved coverage that the weights offer. We estimated the bias associated with the unweighted estimates as the deviation from the corresponding weighted estimate. We used the CAHPS® macro with case-mix adjustment (CMA) to generate both the weighted and unweighted estimates for state estimates of two CAHPS® ratings (Rate Medicare and Rate Health Care) and three CAHPS® composites (Needed Care, Care Quickly, and Good Communication).

The results indicate that the weights are ignorable for many state estimates, especially those for the Good Communication composite. However, the weights are nonignorable for a number of state estimates of the overall ratings of Medicare and health care and the Needed Care and Care Quickly composites. Because all of the root MSEs (square root of the MSE) for the weighted estimates are either equal to or lower than those for the unweighted estimates, we conclude that the weighted analysis of the CAHPS® MFFS survey data can improve the accuracy of state-specific estimates of CAHPS® outcomes without adversely affecting the associated statistical power.

For the 2003 MFFS survey, some variables on the analysis data set were revised, and new variables were added. We revised the race variable to more accurately capture race information from the survey and from supplemental sources of information. Hispanics were added as a racial category in the NEWRACE7 variable. The NEWRACE6 variable combines Asians and Native Hawaiian/Pacific Islanders. The NEWRACE3 variable collapses the NEWRACE7 variable into three categories only: White, African American, and Other.

As begun with the 2002 MFFS in the previous year, we continue to maintain the MFFS trend data file, which is a data file combining the survey results from all 4 years of the MFFS survey (2000 through 2003) thus far. Because there are some differences in the survey instruments used during this time period, we created a crosswalk of survey questions. Users of the trend file should be aware of the differences in the survey instruments across the 4 years. These differences might lead to statistically significant trends that may only be artifacts of the question differences. The trend file assumes that the samples from each of the survey years are independent of each other and, hence, can be combined as one sample.

New for 2003, we are using an extract of the November 1, 2003, Long-Term Care Minimum Data Set (MDS) to identify sample members who had a nursing home stay before, during, and after data collection, among sample members from the 2000 through 2002 CAHPS® MFFS surveys. We have begun to examine the characteristics of this group and are currently assessing the effects on response rate within the MFFS population.

We present details of the sample selection, weighting, and data presentation activities for the 2003 MFFS survey in Section 3.

Data Collection

The MFFS survey is a self-administered mail survey with telephone follow-up of nonrespondents that also offers sample members the option of calling a toll-free number to complete the survey over the telephone. We made four modifications to the 2003 MFFS survey. First, we added 700 beneficiaries from the Virgin Islands to the sample. Second, we sent the third-wave mailing via a special delivery carrier for 2-day delivery, instead of overnight delivery. This change was a cost-saving measure that we implemented as a result of the finding that sample members were more affected by the mode of delivery than the speed of delivery

(Dimitropoulos, Campbell, and Iannacchione, 2003). Third, we mailed an additional thank you/reminder postcard to all survey nonrespondents following the third-wave questionnaire mailing. Finally, to allow the extra postcard enough time to have an effect, we extended the data collection period by 2 weeks.

The data collection period for the 2003 MFFS started with the mailout of the prenotification letter on September 10, 2003, and ended with the close of the telephone follow-up on February 21, 2004. The response rate achieved for the 2003 MFFS was slightly lower than that for the 2002 MFFS. The response rate among eligible sample members was 69.3 percent for the 2003 MFFS, which is 1.3 percent lower than the 2002 MFFS response rate of 70.6 percent. We also noted a decrease in the number of beneficiaries who responded from Puerto Rico, 55.8 percent, which was down from the previous year's rate of 59 percent. Of the 700 Virgin Island beneficiaries who were added to the sample in 2003, only 48.1 percent responded.

We present details of the data collection activities for the 2003 MFFS survey in Section 4.

Costs and Benefits of Improving Response Rates in the MFFS Survey

During this survey year, we conducted research to explore the costs and benefits of increasing the level of effort during the telephone follow-up phase of the MFFS survey. We addressed the following research questions using the 2003 CAHPS® MFFS data:

- Are we penetrating an underrepresented population segment by increasing the level of effort with additional calls?
- At what point do we experience diminishing returns with the number of call attempts sufficient to cap the level of effort?
- What are the cost implications of increasing the level of effort?
- Are the CAHPS® measures affected by the additional responses?

We conducted the analysis on 67,413 valid cases. For each case, a maximum of 12 attempts were made in order to contact a beneficiary and obtain a completed survey. A combination of survival analysis and logistic regression was used to investigate the research questions. Covariates investigated were demographic variables, including race, age, gender, dual eligibility, and disability status.

We found little evidence indicating that segments of the population are being sampled at different rates for the later call attempts versus the earlier attempts. There was no evidence to indicate that late responders differ from early responders. In other words, the demographics of the samples being obtained are not, in general, statistically different when comparing earlier attempts to later attempts. As expected, the total number of responses tends to decline in later call attempts. The mean response probability for the first six attempts is 4.41 percent, whereas the mean for the last six attempts is 3.31 percent ($p=0.0013$). Although this drop is statistically significant, the overall magnitude (1.1 percent) is not large and helps to justify additional attempts.

Some increase in marginal cost per complete is observed; however, it is only a moderate increase. The mean cost for the first six attempts is \$37, whereas the mean for the last six attempts is \$45. This finding implies that, if there is sufficient budget to support additional attempts, there will not be a significant degradation in the cost efficiency of obtaining completed surveys.

Our modeling of the CAHPS® measures, taking into account the same demographic variables of age, race, gender, dual eligibility, and disability status, indicated that the number of attempts had no statistically significant effect on the CAHPS® outcomes.

When choosing the number of attempts to make in the telephone follow-up of the CAHPS® MFFS annual survey, whether response bias is at issue or a concern for diminishing returns and cost, we recommend setting the maximum number of attempts at 12. As long as the budget is sufficient, the maximum number of attempts on this survey should be set no lower than 12, satisfying both scientific and cost concerns.

We present details of this response rate analysis for the 2003 MFFS survey in Section 5.

Case-Mix Adjustment

Case-mix adjustment (CMA) attempts to remove from CAHPS® ratings and composites of care response patterns that do not correspond to actual differences in quality of care and are systematically associated with such patient-level characteristics as demographics, socioeconomic status, and health status. The goal of CMA is to estimate the ratings and composites that a plan or collection of MFFS providers would have received if all providers and plans treated the same standardized population of patients (Medicare beneficiaries). This adjustment should make attributions of ratings and composites to MFFS providers and MA plans more appropriate, supporting better decisionmaking by beneficiaries and CMS.

The present study found that the case-mix adjusters employed in 2001 and 2002 MFFS-vs.-MA CMA (age, education, self-rated health status, self-rated mental health status, and proxy respondent status[1]) continue to constitute an effective case-mix model for both comparison purposes. Self-rated health, self-rated mental health, and education were the three most important CMA variables. These findings are consistent with CMA results for 2000, 2001, and 2002.

The average magnitude of the impact of CMA on a *typical* state has remained moderate. The adjustments for a few states are substantial, however. For national estimates, case-mix adjustment has gone from moderate adjustments in favor of MA in 2001 to small adjustments in favor of MA in 2002 to moderate adjustments in favor of MFFS in 2003. A similar pattern exists for case-mix adjustment of state-level comparisons of MA and MFFS (adjustments favor MA in 2001 but favor MFFS by 2003).

Adjustments favoring MA probably correspond to MA having a higher proportion of certain types of negative responders (e.g., the young and the better educated). Adjustments

---

[1] Although proxy respondent status has only a small empirical effect on CMA, it has been included because many stakeholders feel it is important for the face validity of CMA.

favoring MFFS probably correspond to MFFS having a higher proportion of a different segment of negative responders (e.g., the unhealthy). The shift from adjustments favoring MA to adjustments favoring MFFS could mean that selection into MA on the basis of age and education is becoming less pronounced, but that health-based selection is becoming stronger. Future research should investigate trends in MFFS-vs.-MA case-mix demographics.

In comparing MFFS and MA, there was concern that geographic variation in ratings and reports beyond what is captured in a case-mix model might inappropriately influence the comparisons. Thus, to further improve state-level comparisons, we created weights that match MFFS to MA by county in the 43 states (including the District of Columbia) where MA exists. Comparison weights have gone from making moderate adjustments in favor of MA in 2001 to very small adjustments in 2002 and 2003, probably as a result of improved matching in the sampling design.

We present details of our case-mix adjustment activities for the 2003 MFFS survey in Section 6.

Estimating the Effects of Proxy Respondents in Medicare CAHPS®

Many health surveys, such as the MFFS survey, allow the use of proxy respondents or assistance to ensure that the experiences of the most vulnerable seniors are not omitted. The question that naturally arises is whether proxies give responses comparable to what would have been self-reported. To answer this question, we conducted an analysis of the effects of proxy respondents in Medicare CAHPS®. Of MFFS CAHPS® respondents, 5 percent had someone respond for them (proxy respondent), 14 percent received help from a proxy (proxy assistance), and 81 percent received no help (no proxy). Proxy assistance includes someone reading the survey to the respondent, writing responses, or translating. Proxy assistance might still allow for proxy influence, in that the proxy is discussing the items with the respondent. Spouses and life partners comprised 43 percent of all proxies, other family members comprised 46 percent (two-thirds of these being children), and the remaining 11 percent were not family members.

Proxy respondent effects are generally very small for objective report items but can be somewhat larger for global ratings, especially when proxies are not spouses. Proxy assistance effects are very small. Standard CMA can overestimate proxy effects by failing to compare beneficiaries who have comparable health status. Spouses are very accurate proxy respondents.

Proxy effects are likely to have substantial CMA effects only on surveys in which proxy respondents account for a large proportion of response. These effects are likely to be small on the Medicare CAHPS® surveys but might be larger for subreports on the disabled or those in fair to poor health, or for surveys of nursing home residents. The more objective report composites are preferable to the more subjective global ratings when using proxy respondents, because the former are subject to smaller proxy effects than the latter. When available, spouses should serve as proxies. Adjustment should be made for nonspouse proxy respondents if these are a substantial proportion of all responses. Such an adjustment will not be possible in the 2004 Medicare CAHPS® surveys, which lack information on the relationship of the proxy to the beneficiary. Propensity score weighting or matching can improve the estimation of these adjustments somewhat.

We present details of our analyses of the effects of proxy respondents using the 2001 MFFS survey in Section 7.

Subgroup Analyses

One of the key objectives of the Medicare CAHPS® implementation project is to provide information to help beneficiaries decide among health plan options. Data collected through the Medicare CAHPS® MFFS and MA surveys allow beneficiaries who reside in areas in which there is a choice of plans (managed care or fee-for-service) to access data comparing CAHPS® measures for these two populations. In Section 8, we compare and provide results of analyses of data from the 2003 CAHPS® MFFS survey. (The complete survey instrument is provided in Appendix A.)

We performed analyses of Medicare subgroups to gain a better understanding of the differences in health services experience and satisfaction among Medicare beneficiaries by geographic levels (national, regional, and state level), sociodemographic characteristics, health plan options, and health status. The MFFS population is quite heterogeneous in terms of demographic characteristics, region of residence, presence of supplemental insurance (whether with or without prescription drug coverage or Medicaid), and health-related characteristics. These subgroups of the MFFS beneficiaries have vastly different experiences with and expectations of the health care system and, thus, may perceive the quality of and access to services differently.

The analyses presented examine differences across selected subgroups for the most-positive CAHPS® ratings and reports (i.e., “10,” “Always,” “Not a Problem,” or “Yes”). Nine performance indicators (five composite indicators based on reports and four rating indicators) were used from the 2003 CAHPS® MFFS survey:

- Needed Care Composite
- Good Communication Composite
- Care Quickly Composite
- Respectful Treatment Composite
- Medicare Customer Service Composite
- Rate Personal Doctor
- Rate Specialist
- Rate Health Care
- Rate Medicare

Key Findings

Trends From 2000 Through 2003—During the 4-year period (2000 through 2003), between 84 percent and 89 percent of MFFS beneficiaries reported that they always received needed care. The percentage of most-positive responses was slightly lower in 2003 (85 percent) and 2002 (84 percent) than in 2001 (89 percent) and 2000 (87 percent), but remains quite high.

More than half of beneficiaries (58 percent) reported no problems getting care quickly in 2003. The small differences in results of the Care Quickly composite between the 2003, 2002, and 2001 surveys may be the result of changes in the wording of questions that make up this composite indicator.

Most-positive responses for both the Respectful Treatment composite and the Good Communication composite increased slightly in 2003 to 80 percent and 69 percent, after being consistent at 79 percent and 67 percent, respectively, during the first 3 years of the CAHPS® MFFS survey.

Approximately half of beneficiaries rated health care, their specialist, and their personal doctor a “10” during the 4-year survey period. Ratings of the Medicare health plan were slightly lower overall and decreased over the 4-year period. The most noteworthy change in the ratings for 2003 was a 6 percentage point decrease in beneficiaries’ rating the Medicare health plan a “10.” Specifically, only 38 percent of beneficiaries rated Medicare a “10” in 2003, a large deviation from the gradual downward trend in the previous 3 years—47 percent in 2000, 46 percent in 2001, and 44 percent in 2002.

MFFS vs. MA Comparison—Patient experiences with getting care quickly and good communication were very similar for the MFFS and MA beneficiary populations. In contrast, a higher percentage of MFFS beneficiaries provided most-positive responses to the Needed Care composite during the 4 years compared with their peers enrolled in MA. MFFS beneficiaries were also slightly more likely than MA beneficiaries to provide the best ratings for their health care and the Medicare health plan.

Sociodemographic Characteristics

2003 Highlights—For all CAHPS® composites and the rating of the Medicare plan, beneficiaries between 18 and 45 years of age (i.e., disabled beneficiaries) were less likely to provide most-positive responses than all other age groups. For the Needed Care and Respectful Treatment composites and ratings of Medicare and overall health care in 2003, there was at least a 14 percentage point difference between the proportion of most-positive responses reported by beneficiaries 80 years of age or older and those in the youngest age group, 18 to 45 years. There is a general trend with age and these age differences may reflect response tendencies, rather than better care for older beneficiaries.

Black beneficiaries were generally more likely than White beneficiaries or beneficiaries of other races to provide most-positive responses to CAHPS® indicators in 2003. The only exception to this finding was for the Needed Care composite, where 84 percent of White beneficiaries provided the most-positive response compared with 81 percent of Black beneficiaries. Hispanic beneficiaries were more likely than non-Hispanic beneficiaries to provide most-positive responses to the CAHPS® ratings during 2003. We found considerable differences in CAHPS® ratings between Hispanic and non-Hispanic beneficiaries, with 6 to 17 percentage points separating the CAHPS® scores reported by Hispanic compared with non-Hispanic beneficiaries. These race/ethnic patterns are consistent with CAHPS® findings in commercial and Medicaid populations and are thought to reflect differing expectations.

With the exception of the Needed Care composite, there was a consistent tendency for lower ratings and composites to be associated with increasing beneficiary education. There was a greater than 10 percentage point difference in most-positive responses between the most and least educated beneficiaries. For example, 33 percent of beneficiaries with more than a college degree rated their health care a "10" in 2003, whereas almost twice as many beneficiaries (60 percent) with an eighth grade education or less rated their health care a "10."

Findings From 2000 Through 2003—Similar proportions of male and female MFFS beneficiaries provided most-positive responses to the CAHPS® composite indicators, whereas 3 percent to 5 percent more women provided best possible responses to the four CAHPS® ratings during the 4-year period. There was a consistent age effect, with younger beneficiaries less likely to report most-positive experiences than older beneficiaries during all 4 years. CAHPS® indicator scores similarly varied by education during the 4-year period; less educated beneficiaries were consistently more likely to report most-positive experiences than more educated beneficiaries. Blacks were generally more likely to report most-positive ratings of satisfaction with care than were Whites and beneficiaries of other races. For example, slightly over half of Blacks rated their overall health care a "10" across all 4 years compared with only 46 percent to 47 percent of Whites. Hispanics rated Medicare and their overall health care slightly higher than non-Hispanics, but there was generally little to no difference in their composite scores during the 4-year period.

Health Status

2003 Highlights—For most indicators, beneficiaries reporting excellent health were more likely to give most-positive responses than those reporting poor health. In 2003, there was a 15 percentage point difference between the proportion of the healthiest beneficiaries and sickest beneficiaries providing most-positive responses for the Needed Care composite (80 percent vs. 65 percent for general health perception; 76 percent vs. 61 percent for mental health perception).

The self-reported presence of a chronic illness had no impact on several composites but had a minor impact on the Needed Care and Medicare Customer Service composites, both of which had fewer most-positive responses from beneficiaries with a chronic illness than from beneficiaries with no chronic illness. This finding suggests that issues related to access to care and dealing with Medicare may be more problematic for those who are likely to use more health care services. A smaller proportion of beneficiaries with a chronic illness provided most-positive ratings of "10" for all four indicators reflecting ratings of beneficiary satisfaction, compared with beneficiaries who did not report having a chronic illness. In 2003, only 36 percent of chronically ill beneficiaries rated Medicare a "10" compared with 43 percent of beneficiaries with no chronic illness. These findings are adjusted for self-rated health. Beneficiaries who had been hospitalized overnight at least once during the past year were more likely to rate their specialists a "10";

however, hospitalization during the past year had little effect on other CAHPS® scores during the 4 years. In 2003, disability seemed to have no effect on the five CAHPS® composites; however, disabled beneficiaries gave a slightly higher percentage of “10” ratings for their personal doctor, specialist, health care, and the Medicare plan.

Findings From 2000 Through 2003—Across all indicators, we found a strong association between self-reported health status and CAHPS® scores; beneficiaries reporting better general and mental health status were more likely to provide most-positive responses for each CAHPS® indicator during the 4 years.

Access to Care

2003 Highlights—Beneficiaries who were dually eligible for Medicare and Medicaid reported greater satisfaction and better experiences than those who reported having additional insurance and prescription drug coverage; this was the case for the Good Communication and Medicare Customer Service composites and all ratings in 2003. In particular, over half of dually eligible beneficiaries rated Medicare a “10” compared with 35 percent of beneficiaries who had additional insurance and prescription drug coverage. The additional insurance and prescription drug coverage is self-reported. Dual eligibility status is obtained from the CMS database.

Beneficiaries who reported having a personal doctor were more likely to provide most-positive responses, generally by at least 10 percentage points, for the five CAHPS® composites and their ratings of health care and specialists in 2003. Ratings of Medicare were not affected by whether beneficiaries reported having a personal doctor or not.

Findings From 2000 Through 2003—Beneficiaries who reported having additional insurance but no prescription drug benefits were less satisfied and reported worse experiences than those with additional insurance and prescription drug benefits during each of the 4 years.

Additional Analyses

Supplemental Analysis Using Claims Data: Relationship Between CAHPS® Scores and Hospital Encounters for Ambulatory Care Sensitive Conditions—The objective of this analysis was to examine whether the incidence of hospital encounters for ambulatory care sensitive conditions (ACSCs) is associated with self-reports of poor access to care or dissatisfaction with health care services among MFFS beneficiaries. Our results showed that 6.1 percent of MFFS beneficiaries experience at least one inpatient admission, observation stay, or emergency room visit for an ACSC during the 12 months prior to the MFFS survey. Although beneficiaries with an ACSC were modestly more likely to report problems getting needed care, and this difference was statistically significant, the mean difference in this CAHPS® composite was very small. Overall, there were virtually no differences in mean CAHPS® scores between beneficiaries who did or did not experience an ACSC encounter. Because this analysis used a composite measure of different ACSCs, we cannot know if a relationship exists between CAHPS® scores and the incidence of hospitalizations for specific conditions. Therefore, we recommend that additional analyses be conducted to examine whether there is a relationship between CAHPS® measures and ACSCs for specific conditions.

Supplemental Analysis Using Claims Data: Relationship Between CAHPS® Scores and Claims-Based Quality Measure for Beneficiaries With Diabetes—In this study, using the individual as the unit of analysis, we examined the relationship between measures of patient satisfaction and claims-based clinical measures of (a) the quality of preventive care delivered to MFFS beneficiaries with diabetes and (b) complications from diabetes. We hypothesized that CAHPS® measures of patient experience (i.e., composite measures of getting needed care, getting care quickly, and good communication with providers) were more likely to be associated with clinical measures of quality than attitudes represented by CAHPS® ratings and patient experiences with Medicare and provider customer service.

Overall, we found that patient satisfaction measures were positively associated with a subset of the clinical measures of health care quality under study. Specifically, beneficiaries with diabetes who received recommended eye exams and at least one of the recommended physiological measures were more satisfied with their health care than their peers who did not receive these preventive care services. In contrast, the clinical measures of self-care and disease progression were not associated with the CAHPS® indicators.

However, the significant differences found among persons who received recommended eye exams and physiological tests represent very small differences in average CAHPS® scores. Therefore, these statistically significant differences do not necessarily reflect *meaningful* differences in patient satisfaction.

Depression Among 2000 and 2001 MFFS Beneficiaries—A third analysis begun this year uses the linked MFFS survey and beneficiary claims data to identify beneficiaries diagnosed with depression. To date, we have constructed definitions of depression using ICD-9 codes. Future research will examine the relationship between depression diagnosis and the mental health component scores derived from the SF-12 questions included on the CAHPS® MFFS survey.

We present details of the subgroup analyses for the 2003 MFFS survey in Section 8.

SECTION 1
INTRODUCTION

by Jeffrey S. Laufenberg, M.S., RTI

The Centers for Medicare & Medicaid Services (CMS) currently conducts three Consumer Assessment of Health Plans Study (CAHPS®) surveys of the Medicare population.

1. The Medicare CAHPS® Fee-for-Service (MFFS) Survey: a survey of Medicare beneficiaries who are enrolled in Original Medicare, also referred to as fee-for-service Medicare.

2. The Medicare CAHPS®Advantage (MA) Survey (previously known as the Medicare CAHPS® Managed Care [MMC] Survey): a survey of Medicare beneficiaries currently enrolled in a Medicare managed care plan.

3. The Medicare CAHPS® Disenrollment Assessment Survey: a survey of Medicare beneficiaries who recently left or disenrolled from a Medicare managed care plan.

The surveys collect information on an annual basis to fulfill a requirement of Congress (under the Balanced Budget Act of 1997) to provide information to Medicare beneficiaries on the quality of health services provided through the Original Medicare plan and to compare this information with similar information collected from beneficiaries enrolled in Medicare managed care health plans. Information from all three surveys is used in the Medicare Personal Plan Finder web page that is available to Medicare beneficiaries on the Medicare web site (www.medicare.gov/MPPF/DefaultVersion/home.asp). The Personal Plan Finder is designed to help beneficiaries make more informed decisions when choosing a Medicare health plan.

The 2003 MA survey was the seventh implementation of what had previously been known as the MMC survey. The Disenrollment Assessment and MFFS surveys were conducted for the first time in fall 2000. The data from the MA and Disenrollment Assessment surveys are combined and analyzed together so that results for any given managed care plan reflect the experiences both of individuals who stayed in the plan and of those who voluntarily disenrolled during the 6 months preceding data collection. In fall 2000, CMS funded the national implementation of the MFFS survey, thereby providing the data to construct CAHPS® ratings and composites for both the fee-for-service and managed care populations.

The 2000 through 2003 CAHPS® MFFS surveys were conducted for CMS by RTI, with the assistance of RAND, NCS Pearson, and Discovery Research Group (DRG). The work was performed under subcontract to the Center for Health Systems Research and Analysis at the University of Wisconsin-Madison (UW), as part of UW's contract with CMS (Contract Number 500-95-0061).

This report provides a summary of the methodology and findings of the 2003 CAHPS® MFFS survey as well as some trend analysis of the 2000 through 2003 survey results. More detailed information on the topics presented here is available in the individual reports developed for the 2003 survey. Similar reports are also available for the years 2000, 2001, and 2002 CAHPS® MFFS surveys.

SECTION 2
2003 MEDICARE CAHPS® FEE-FOR-SERVICE SURVEY QUESTIONNAIRE

by Judith T. Lynch, B.A., RTI

One of the purposes of conducting the three Medicare CAHPS® surveys is to collect and provide comparative information to Medicare beneficiaries about Medicare managed care plans and Original Medicare. Therefore, it is important that the questionnaires used in each of the three surveys—that is, the Medicare CAHPS® Fee-for-Service (MFFS) Survey, the Medicare CAHPS® Advantage (MA) Survey (previously known as the Medicare CAHPS® Managed Care [MMC] Survey), and the Medicare CAHPS® Disenrollment Assessment Survey—be as similar as possible to one another. The MFFS project team worked with CMS and the other two Medicare CAHPS® project teams in January through July 2003 to discuss potential changes to the questionnaires to be used in the three Medicare CAHPS® surveys in the upcoming year as a result of analytic, consumer reporting, or policy issues. In addition, the project teams reviewed any changes recommended by the Agency for Healthcare Research and Quality (AHRQ) to selected core CAHPS® questions. However, most of the discussions during the project coordination meetings for the 2003 surveys were to address differences in wording for some questions on the questionnaire and to suggest strategies for reconciling the differences for future implementation of the Medicare CAHPS® surveys. The changes made to the questionnaire for the 2003 survey implementation are described in the following paragraphs. A copy of the 2003 MFFS survey questionnaire is provided in Appendix A.

2.1 Changes to Prescription Drug Questions

Some of the more significant changes made to the questionnaires used in the 2003 implementation of the three Medicare CAHPS® surveys resulted from an analysis of responses to the prescription drug questions—specifically, a comparison of responses to the prescription questions included in the 2001 and 2002 survey years. Each of the three Medicare CAHPS® surveys contains questions about respondents' experience with getting prescription medicines. In the 2000 surveys, the prescription drug questions were as follows:

- **In the last 6 months, did you get <u>any</u> new prescription medicine or refill a prescription?** "Yes" and "No." A skip instruction appeared beside the "No" response directing respondents to skip out of the follow-up series of questions.

- **In the last 6 months, how much of a problem, if any, was it to get your prescription medicines?** "A big problem," "A small problem," "Not a problem," and "I didn't get any new prescriptions or refills in the last 6 months." This last "not applicable" answer choice was included for respondents who failed to follow the skip instruction for the "No" response to the preceding question.

- **In the last 6 months, how often did you get the prescription medicine you needed?** "Never," "Sometimes," "Usually," "Always," and "I didn't get any new prescriptions or refills in the last 6 months."

During the project coordination meetings among CMS and the three Medicare CAHPS® teams held before the 2001 surveys were implemented, team members expressed concern that the screening question might cause us to miss sample members who did not get a new prescription or refill a prescription but who had problems getting the prescription medicines that they needed. Therefore, for the 2001 surveys, the screening question was deleted and the other prescription drug questions were changed to the following:

- **In the last 6 months, how often did you get the prescription medicine you needed?** "Never," "Sometimes," "Usually," "Always," and "I didn't need any prescription medicines in the last 6 months."

- **In the last 6 months, how much of a problem, if any, was it to get the prescription medicine you needed?** "A big problem," "A small problem," "Not a problem," and "I didn't need any prescription medicines in the last 6 months."

In fall 2002, the MA project team conducted some analyses to assess the impact of these question changes on response patterns to the prescription drug questions. The findings from those analyses are summarized in the following excerpt from a report that the team prepared and submitted to CMS in December 2002 (Cleary, Zaborski, and Zasklavsky, 2002):

> The approach to determining whether a person got any prescriptions was different in 2000 and 2001. One motivation for changing the question in 2000 was the belief that some people who needed a prescription might not have gotten one, and thus, one should not screen them out from subsequent questions. Although this is possible, the lack of an explicit question about obtaining prescription in 2001 apparently resulted in about 10 percent of the sample who did not get a prescription answering questions intended for those who needed a prescription.
>
> These results are consistent with previous CAHPS® research showing that a "does not apply" response category in a substantive question is a very inaccurate way of determining the persons who should be answering a specific question.

As a result of these findings, CMS and the three Medicare CAHPS® project teams agreed to change the series of questions about prescription drugs for the 2003 surveys to be similar to those included in the 2000 surveys, including a screening question to identify and screen out people who did not get a new prescription or refill a prescription during the reference period. The prescription drug questions included in the 2003 Medicare CAHPS® surveys were as follows:

- **Question 43. In the last 6 months, did you need <u>any</u> new prescription medicines or need to refill a prescription?** "Yes" and "No." If "No," skip to Question 46.

- **Question 44. In the last 6 months, how much of a problem, if any, was it to get the prescription medicine you needed?** "A big problem," "A small problem," "Not a problem," and "I didn't need any prescription medicines in the last 6 months."

- **Question 45. Of the times when you needed prescription medicines in the last 6 months, how often were you able to get the medicine?** "Never," "Sometimes," "Usually," "Always," and "I didn't need any prescription medicines in the last 6 months."

The MFFS project team also added a question about prescription drug discount cards in an attempt to help respondents differentiate between discount drug cards and actual prescription medicine insurance coverage. In addition, the placement of the question about insurance that pays for prescription drugs was changed: Question 47 in the 2003 questionnaire is a revised version of Question 3 in the 2002 questionnaire. The preamble to this question was changed to "*Medicare does not pay for prescription medicines unless they need to be injected by a doctor or you are in the hospital.*" The actual question was changed to read, "*Not including Medicare, do you have any* <u>*other*</u> *health insurance that pays at least some of the costs of medicines prescribed by doctors and other health providers?*" The final change to the series of questions about prescription drugs in the CAHPS® MFFS questions was the addition of a question to determine how much of the cost of prescription medicines is covered by other health insurance.

2.2 Other Changes to the 2003 Questionnaire

Other changes were made to the questionnaire prior to the 2003 survey to make the question wording and placement of questions in the three Medicare CAHPS® questionnaires more comparable or to respond to AHRQ recommendations. These changes are summarized below:

- In Question 1, the phrase "Some people who are on Medicare also have" was changed to "Some people *who have* Medicare also have…." The wording was changed because "being on" Medicare may have a negative connotation for some sample members.

- The answer choices to the question about the number of months sample members have been going to their personal doctor or nurse (Question 5 in the 2003 survey) were changed to be consistent with the units of time that are used in the CAHPS® core questions.

- The preamble "How would you rate your personal doctor or nurse now?" was deleted from the rating of personal doctor question, which is Question 6 in the 2003 survey.

- The questions about needing and making visits for conditions that needed care right away (Questions 19–22 in the 2003 survey) were moved so that they appear before the questions about sample members' getting an appointment for health care as soon as they wanted.

- The definition of wait time was deleted from the question about waiting time in a doctor's office. In addition, this question (Question 29 in the 2003 survey) was changed to read, "*In the last 6 months, how often were you taken to the exam room* <u>*within 15 minutes*</u> *of your appointment?*"

- The preamble "How would you rate all the health care you got in the last 6 months from all doctors and other health providers?" was removed from Question 36.

- The definition of paperwork was dropped from the question about paperwork, which is Question 53 in the 2003 survey. This change, which was recommended by the core

CAHPS® team, was made because the definition was long and confusing and the meaning of the question was not clear to respondents.

- The preamble “How would you rate all of your experience with Medicare?” was deleted from the rating of Medicare question (Question 55 in the 2003 questionnaire).
- Several questions were deleted from the “About You” section. These include the questions about having a mammogram in the last 12 months, having a Pap smear in the last 12 months, and having a prostate screening or PSA test in the last 12 months.

The placement of many of the questions in the “About You” section (Questions 56–91) was changed to ensure consistency between the placement of questions in the 2003 CAHPS® MFFS and MA questionnaires. The project teams decided that questions about physical and emotional health that are common to both surveys should appear first in this section, followed by questions that are unique to each survey. The following is a summary of question placement changes between the 2002 and 2003 MFFS questionnaires:

1. The rating of overall health question (old Question 59) was moved to become the first question in this section (now Question 56).
2. The question comparing health to 1 year ago is now the second question in this section (old Question 60, now Question 57).
3. Old Question 54 (patient in hospital overnight in last 12 months) is now Question 58.
4. The rating of mental health question is now the fourth question in this section (old Question 58, now Question 59).
5. Questions about chronic conditions (old Questions 55–57) are now Questions 60–62.
6. The question about needing help with personal care needs because of an impairment or health problem (old Question 73) is now Question 63.
7. The question about needing help with routine needs because of an impairment or health problem (old Question 74) is now Question 64.
8. The question about having a condition that interferes with independence (old Question 72) is now Question 65.
9. Questions about health limitations (old Questions 61 and 62) are now Questions 66 and 67.
10. The rating of overall health question was moved to appear as the first question in this section and therefore is not placed with other SF-12 questions. The remaining SF-12 questions (old Questions 61–71) appear now as Questions 68–76.

11. The question about rating of life overall (old Question 53, now Question 86) now appears after the question about race.
12. Questions about getting flu and pneumonia shots (old Questions 79 and 80) are now Questions 77 and 78.

SECTION 3
SAMPLE SELECTION, WEIGHTING, AND DATA PRESENTATION

by Jeffrey S. Laufenberg, M.S., RTI; Vincent G. Iannacchione, M.S., RTI;
Lisa A. Carpenter, B.S., RTI; and Marc N. Elliott, Ph.D., RAND

3.1 Sample Selection and Weighting

In this section, we summarize our procedures for assigning design-consistent weights to Medicare CAHPS® Fee-for-Service (MFFS) Survey respondents for use in the analysis and presentation of MFFS data. We review the sample frame applicable to all 4 years of the survey thus far and contrast the current year's sampling methodology with those in previous years to highlight its advantages. To examine the effects of applying our weights, we provide a detailed, state-by-state comparison of the unweighted and weighted proportions of beneficiaries reporting the highest scores for each of the five CAHPS® indicators on the CMS web site.[1]

Analysis weights enable design-consistent estimation of population parameters by scaling the disproportionalities between the sample and the population. For the MFFS survey, the weights may be viewed as inflation factors that account for the number of beneficiaries in the target population that a sample member represents. The initial component of an analysis weight for a sample member is the selection probability that is specified by the sample design. Typically, adjustments are then made to the weights to compensate for potential biases attributable to differential response and coverage among sample members.

We begin with a description of the MFFS sampling frame and its coverage of the MFFS target population. Next, we compare the selection allocation of this year's MFFS survey sample with those of previous years. Then, we describe how the initial sampling weights were assigned to sample members and how the initial weights of sample respondents were post-stratified to specified MFFS population counts. Finally, we present state-level proportions (weighted and unweighted) for five CAHPS® indicators—Care Quickly composite, Good Communication composite, Needed Care composite, Rate Medicare, and Rate Health Care—to graphically depict the effects of the weights on our case-mix adjusted estimates.

3.1.1 MFFS Sampling Frame

The sample of 178,650 beneficiaries selected for the 2003 MFFS was drawn from a sampling frame constructed from the July 2003 version of CMS' Enrollment Database (EDB). The frame comprised almost 32.0 million persons who were enrolled in fee-for-service Medicare for at least the prior 6 months, did not have a representative payee, were over the age of 18, and resided in the United States, Puerto Rico, or the Virgin Islands. The frame included 4,002 beneficiaries who were initially sampled but subsequently died before or during data collection and were therefore determined to be *ineligible* for the survey.

---

[1] Medicare Personal Plan Finder (MPPF) in http://www.medicare.gov.

The frame also included beneficiaries who did not speak English or Spanish[1] and beneficiaries who were mentally or physically incompetent and without access to a proxy. We classified these individuals as survey *eligible* to be consistent with the Medicare CAHPS® Advantage (MA) Survey. Please refer to ***Table 3.1*** for the eligibility status of the 2003 MFFS sample. Unlike the CAHPS® MA survey, we classified beneficiaries who were institutionalized during data collection as eligible for the survey.[2]

Table 3.1
Eligibility status of 2003 MFFS sample members

		Sample members	
		Number	Percent
Eligible	Completed questionnaire	120,974	67.7
	Proxy respondent	7,900	4.4
	Received help with survey	13,614	7.6
	Beneficiary respondent	99,460	55.7
	Refused	13,982	7.8
	Mental/physical barrier	1,999	1.1
	Language barrier	714	0.4
	Other nonrespondents		
	Deliverable address and phone	34,407	19.4
	Deliverable address and no phone	2,169	1.2
	Undeliverable address and phone	12	0.0
	Undeliverable address and no phone	391	0.2
	Total eligible sample members	174,648	97.8
Ineligible	Deceased	4,002	2.2
	Total sample	178,650	100.0

3.1.2 Historical MFFS Selection Allocation

In the 2000 MFFS survey, the selection allocation was defined to be 600 beneficiaries from each of the 275 geographic units (hereafter, “geounits”) in the United States, and 3,000 beneficiaries from Puerto Rico, for a total sample size of 168,000 beneficiaries in 276 geounits. For the 2001 MFFS survey, no change was made to the number of geounits, but the total sample size was increased to 177,950 beneficiaries. For the 2002 survey, no change was made to the number of geounits or to the total (national) MFFS sample size from the 2001 survey, but we did introduce a reallocation of the sample.

---

[1] The questionnaire was only administered in English and Spanish.

[2] Institutionalized MA beneficiaries were identified on the Group Health Plan file prior to the selection of the MA sample. MFFS beneficiaries are not included on this file, and institutionalized status is not maintained on the EDB.

3.1.3 2003 MFFS Selection Allocation

For the 2003 MFFS survey, we added one geounit for the Virgin Islands, bringing the total number of geounits in our study design to 277 and our total sample size to 178,650. In addition, we introduced a reallocation of sample between 2002 and 2003 to provide better power for estimates in counties that experienced significant managed care retreat and counties with insufficient sample sizes owing to higher than expected nonresponse in the previous year. In particular, selected counties in five "donor" states (California, Florida, New York, Ohio, and Pennsylvania) with the greatest effective sample size when compared with MA contributed a proportionate amount of their allocated sample, in excess of 330 completes from the previous year, to recipient counties in eight states: Iowa, Kansas, Kentucky, New Mexico, North Dakota, Rhode Island, Virginia, and West Virginia.

3.1.4 MFFS Target Population

Ideally, the sampling frame would include all members of the MFFS *target population*, the entire population of MFFS-eligible beneficiaries to which we make inference. However, changes in the composition of the MFFS population during the 5 months between sample selection and the end of data collection made complete coverage of the target population impossible. Therefore, after consultation with the CMS Project Officer, we defined the MFFS target population to include all survey-eligible beneficiaries as of October 2003. This target date was chosen for two reasons: (1) it corresponds with the peak of data collection; and (2) it is about the same amount of time after sample selection (approximately 2 months) as the target date for the CAHPS® MA survey.

There were more than 32.1 million fee-for-service Medicare beneficiaries on the October 2003 version of the EDB who satisfied the eligibility requirements for the survey.

3.1.5 Initial Sampling Weights

We stratified by county to vary the sampling rates of beneficiaries selected for the MFFS survey to achieve the design goals of the study (Elliott et al., 2000). In general, beneficiaries in rural counties and less populous states were sampled at higher rates than those in urban counties and populous states. The median county-level sampling rate was one selection per 176 fee-for-service beneficiaries, with beneficiaries in 50 percent of the counties receiving between one selection per 146 beneficiaries and one selection per 203 beneficiaries. The highest county-level sampling rate was one selection per 2.3 beneficiaries (Pierce County, ND), while the lowest rate was one selection per 874 beneficiaries (Barbour County, WV). As a result of this differential selection, the sample distribution produced by the stratified selection of fee-for-service beneficiaries is not proportional to the population distribution from which the sample was selected.

We assigned an initial sampling weight to each selected beneficiary as the inverse of the selection probability to reflect the differential selection rates of beneficiaries from each state or county. For example, beneficiaries selected from the county having the highest sampling rate were assigned an initial weight of 2.3, compared with an initial weight of 874 for beneficiaries selected from the county having the lowest sampling rate. This variability in the sampling weights induces unequal weighting effects that inflate the variances of the sample estimates.

Table 3.2, on the following page, summarizes the unequal weighting effects attributable to the initial sampling weights for various MFFS subpopulations.

3.1.6 Post-Stratification of the Initial Sampling Weights

The response rates for the 2003 MFFS survey varied considerably with respect to race (Whites higher than Blacks and Other/Unknown race), age (younger seniors higher than the very old), dual eligibility, and region (Midwest higher than others) (see ***Table 4.2*** in Section 4, Data Collection). As a result, the respondent distribution is composed of too few Blacks and Other/Unknown, too few dual eligibles, and too many beneficiaries from the Midwest as compared to the original sample distribution. These differential response-rate patterns combined with differential answer patterns to the survey represent a potential for nonresponse bias.

To reduce the potential biasing effects of differential nonresponse, we post-stratified the initial sampling weights of respondents to 337 separate counts of the number of eligible fee-for-service beneficiaries obtained from the October 2003 version of the EDB. (The EDB counts exclude eligible beneficiaries who died between October 2003 and February 2004, the end of data collection.) The counts include totals for each of the 277 geounits in the United States, Puerto Rico, and Virgin Islands, as well as 60 totals formed by the intersection of the following demographic variables:

- Age Category (5): under 65, 65–69, 70–74, 75–79, 80 and older
- Gender (2): Male, Female
- Race (3): White, Black, Other/Unknown
- Dual Eligibility (2): Yes, No

Readers should note that the 277 geographic counts “cut across” the 60 demographic counts in that each sample member belongs to both a geographic cell and a demographic cell. Therefore, we used a generalized exponential model (GEM) (Singh and Folsom, 2000) to ensure that the adjusted weights sum to all 337 counts while imposing bounds on the adjustment factors so that extreme weights could be controlled. GEM is a generalization of the well-known logit method of Deville and Sarndal (1992), and the usual raking method can be obtained as a special case.

As a result of the above, the post-stratified weights of eligible respondents sum to 32.1 million—the number of survey-eligible beneficiaries on the October 2003 version of the EDB.

Table 3.2 summarizes the post-stratification adjustments we made to the sampling weights. At the national level, on average, an adjustment factor of 1.50 (32,143/21,409) was applied to the initial sampling weights of eligible respondents. Because of differential nonresponse, however, the adjustment factors varied substantially across the post-strata. For example, an average adjustment factor of 1.69 (2,067/1,223) was applied to beneficiaries in New York to compensate for the low (59.4 percent) response rate there. At the other extreme, respondents in North Dakota (which had a 76.6 percent response rate) only required an average adjustment of 1.27.

Table 3.2
2003 MFFS sampling weight post-stratification summary

		Initial sampling weights			Post-stratified weights		
Post-stratum	Number of respondents	Weight sum (000s)	Unequal weighting effect	Effective sample size	Weight sum (000s)	Unequal weighting effect	Effective sample size
Overall							
USA, PR, and VI	120,974	21,409	1.22	99,551	32,143	1.31	92,134
Dual eligibility							
No	106,731	18,853	1.21	88,319	27,629	1.29	82,915
Yes	14,243	2,556	1.26	11,271	4,514	1.40	10,151
Age							
Under 65	13,101	2,297	1.21	10,865	4,151	1.28	10,201
65-69	27,769	4,916	1.22	22,766	7,293	1.31	21,174
70-74	27,454	4,847	1.22	22,543	6,802	1.31	20,885
75-79	23,807	4,230	1.22	19,566	5,958	1.30	18,289
80 +	28,843	5,120	1.21	23,815	7,940	1.31	22,065
Race							
White	106,842	18,808	1.20	89,298	27,374	1.26	84,861
Black	9,283	1,661	1.29	7,185	2,953	1.38	6,716
Other	4,849	940	1.42	3,415	1,816	1.60	3,040
Gender							
Male	52,198	9,224	1.21	42,982	13,947	1.32	39,524
Female	68,776	12,185	1.22	56,570	18,197	1.31	52,619
CMS region							
I. CT, ME, MA, NH, RI, VT	8,402	1,123	1.29	6,520	1,727	1.40	6,006
II. NJ, NY, PR	10,209	2,111	1.17	8,713	3,538	1.25	8,200
III. DE, DC, MD, PA, VA, WV	14,155	2,251	1.22	11,650	3,357	1.22	11,559
IV. IL, NC, GA, SC, FL, KY, MS, TN	25,097	4,659	1.08	23,294	6,929	1.10	22,873
V. IL, IN, MI, MN, OH, WI	21,736	4,221	1.20	18,122	6,080	1.32	16,477
VI. AR, LA, NM, OK, TX	13,424	2,400	1.17	11,486	3,573	1.22	10,981
VII. IA, KS, MO, NE	7,864	1,220	1.15	6,840	1,697	1.15	6,809
VIII. CO, MT, ND, SD, UT, WY	5,856	690	1.31	4,486	921	1.34	4,364
IX. AZ, CA, HI, NV	9,005	1,973	1.37	6,550	3,222	1.57	5,720
X. AK, ID, OR, WA	4,889	755	1.11	4,407	1,087	1.15	4,261
States							
Alabama	2,481	422	1.01	2,466	598	1.02	2,428
Alaska	510	32	1.00	510	43	1.01	503
Arizona	1,342	318	1.04	1,288	480	1.06	1,267
Arkansas	1,690	297	1.03	1,647	414	1.04	1,620
California	6,021	1,466	1.37	4,393	2,460	1.56	3,849
Colorado	1,307	232	1.01	1,290	326	1.03	1,264
Connecticut	1,511	279	1.03	1,466	448	1.07	1,418
Delaware	1,253	77	1.00	1,247	110	1.03	1,222
District of Columbia	676	35	1.00	676	62	1.02	664
Florida	6,448	1,439	1.04	6,228	2,192	1.05	6,141
Georgia	3,053	552	1.02	2,989	845	1.03	2,951
Hawaii	817	70	1.33	614	105	1.37	595

(continued)

Table 3.2
(continued)

		Initial sampling weights			Post-stratified weights		
Post-stratum	Number of respondents	Weight sum (000s)	Unequal weighting effect	Effective sample size	Weight sum (000s)	Unequal weighting effect	Effective sample size
States							
Idaho	913	111	1.00	913	148	1.01	905
Illinois	3,803	935	1.51	2,519	1,430	1.74	2,189
Indiana	3,314	551	1.00	3,302	780	1.02	3,263
Iowa	1,811	322	1.21	1,502	425	1.22	1,485
Kansas	1,633	259	1.15	1,419	349	1.16	1,413
Kentucky	2,268	400	1.16	1,963	571	1.17	1,945
Louisiana	1,891	307	1.05	1,803	493	1.06	1,780
Maine	1,761	152	1.02	1,731	209	1.03	1,708
Maryland	2,099	395	1.05	1,998	597	1.06	1,989
Massachusetts	2,634	450	1.34	1,969	718	1.42	1,859
Michigan	4,553	898	1.13	4,013	1,293	1.17	3,894
Minnesota	2,244	398	1.00	2,237	534	1.02	2,210
Mississippi	2,061	262	1.28	1,615	388	1.29	1,594
Missouri	2,992	475	1.04	2,872	697	1.06	2,816
Montana	527	98	1.00	527	131	1.01	523
Nebraska	1,428	164	1.21	1,183	227	1.19	1,199
Nevada	825	119	1.07	775	177	1.10	749
New Hampshire	1,354	115	1.03	1,312	164	1.04	1,303
New Jersey	3,538	644	1.17	3,018	1,034	1.19	2,984
New Mexico	1,408	133	1.14	1,231	192	1.15	1,224
New York	6,001	1,223	1.14	5,246	2,067	1.23	4,874
North Carolina	4,006	712	1.09	3,668	1,052	1.11	3,599
North Dakota	1,531	74	1.83	837	94	1.83	835
Ohio	4,603	919	1.02	4,530	1,344	1.04	4,427
Oklahoma	1,705	309	1.00	1,701	437	1.02	1,679
Oregon	1,254	212	1.01	1,241	308	1.03	1,212
Pennsylvania	4,894	961	1.06	4,618	1,445	1.06	4,623
Puerto Rico	670	244	1.00	670	436	1.02	654
Rhode Island	605	62	1.00	605	103	1.02	596
South Carolina	1,970	377	1.01	1,953	549	1.02	1,925
South Dakota	1,071	88	1.24	861	112	1.25	855
Tennessee	2,810	496	1.03	2,717	734	1.05	2,677
Texas	6,730	1,355	1.19	5,640	2,037	1.26	5,345
Utah	878	150	1.02	857	196	1.03	850
Vermont	537	66	1.00	537	85	1.01	533
Virgin Islands	337	5	1.00	337	11	1.02	330
Virginia	3,352	575	1.19	2,828	850	1.21	2,767
Washington	2,212	400	1.07	2,063	588	1.10	2,018
West Virginia	1,881	208	1.45	1,296	294	1.47	1,276
Wisconsin	3,219	520	1.02	3,151	699	1.03	3,122
Wyoming	542	48	1.00	542	62	1.01	538

NOTE: Post-stratification was used to force the weight sums of MFFS respondents to agree with totals obtained from the October 2003 version of the EDB (i.e., the post-stratified weight sum). The unequal weighting effect (UWE) measures the amount of variance inflation above an equally weighted sample. The effective sample size is the number of respondents divided by the UWE.

The overall unequal weighting effect after post-stratification was 1.31, compared with 1.22 for the initial sampling weights. This increase can be interpreted as the "price" paid (in terms of variance inflation) for ensuring that the weighted distribution of respondents reflects the distribution of the October 2003 version of the EDB. Although the unequal weighting effect varied among the post-strata, the increase attributable to post-stratification did not seriously affect the effective sample sizes.

3.1.7 Effects of Weighting

Weighted analysis of the CAHPS® MFFS survey data offers the following advantages:

- **Bias reduction**. Weights can eliminate selection bias caused by sampling beneficiaries at different rates. For example, sample members in the county with the highest sampling rate were selected at 380 times the rate of sample members in the county with the lowest sampling rate. In addition, weights can reduce nonresponse bias that results from differential response rates. It should be noted that case-mix adjustment (CMA) also may be used to reduce response bias.

- **Improved coverage**. Changes in the composition of the MFFS population during the 5 months between sample selection (in August) and the end of data collection (in February) made complete coverage of the target population impossible. Post-stratification of the design weights of respondents ensures that the weighted distribution of respondents reflects the true (as of October 2003) fee-for-service Medicare population distribution with respect to age, gender, race, dual eligibility, and geounit.

The primary disadvantage of weighted analysis is the increased variances that often accompany weighted estimates. This is not the case for the CAHPS® MFFS survey, however, since an overall UWE of 1.31 suggests that our design-consistent weights have only a slight effect on the efficiency of the estimates in comparison to a randomly selected sample of the same size. We found no discernible increase in the variances when weights were used to generate state-level estimates for two CAHPS® ratings (Rate Medicare and Rate Health Care) and three CAHPS® composites (Needed Care, Care Quickly, and Good Communication). Another disadvantage of weighted analysis often cited by researchers is that special software often is needed to analyze weighted data. However, recent changes to the CAHPS® macro enable weighted estimates of CAHPS® outcomes to be generated easily.

We evaluated the effect of the MFFS analysis weights on the accuracy of the survey estimates by comparing the mean square errors (MSEs) of weighted estimates to the corresponding MSEs of unweighted estimates. The MSE, defined as the sum of the bias squared and the variance, is used to measure the combined effect of bias and variance on the survey estimates. We assumed that the weighted estimates represent relatively unbiased estimates because of the bias reduction and improved coverage that the weights offer. We estimated the bias associated with the unweighted estimates as the deviation from the corresponding weighted estimate.

We used the CAHPS® macro with the CMA factors to generate both the weighted and unweighted estimates for state estimates of the two CAHPS® ratings and three composites. The estimates are the percentage of Medicare beneficiaries reporting the most positive CAHPS® outcome. We included CMA in the evaluation because most of our survey estimates are case-mix adjusted. In addition, comparing weighted and unweighted CMA estimates enabled us to account for any reduction in response bias that is attributable to the CMA factors. If weights do not reduce the MSEs of key CAHPS® estimates (compared to unweighted), then there is little motivation for their use in the analysis. In this situation, the weights are said to be "ignorable." Examples of situations where weights are ignorable include the following:

- There is little or no difference between weighted and unweighted estimates.
- The variance inflation attributable to weighting exceeds the squared reduction in bias (i.e., the weighted MSE is greater than the unweighted MSE).

In ***Figure 3.1***, we plotted the weighted and unweighted root MSEs (i.e., the square root of the MSEs) to display the differences in the same scale as the percentage estimates. For example, in the plot labeled "Rate Medicare," the root MSE for the unweighted CMA estimate for North Dakota (ND) is 2.4 percent, compared with 1.2 percent for the weighted CMA estimate. Because the variances of the weighted and unweighted estimates are virtually identical, the 1.2 percent difference in the root MSEs can be attributed to the bias of the unweighted CMA estimate. Overall, the results indicate that the weights are ignorable for many state estimates, especially those for the Good Communication composite. However, the weights are nonignorable (in terms of reduced MSEs) for a number of state estimates of the overall ratings of Medicare and health care and the Needed Care and Care Quickly composites. Because all of the root MSEs appear on or below the diagonal, we conclude that the weighted analysis of the CAHPS® MFFS survey data can improve the accuracy of state-specific estimates of CAHPS® outcomes without adversely affecting the associated statistical power. Bias tends to be a more important component of MSE for larger sample sizes.

3.2 Data Presentation

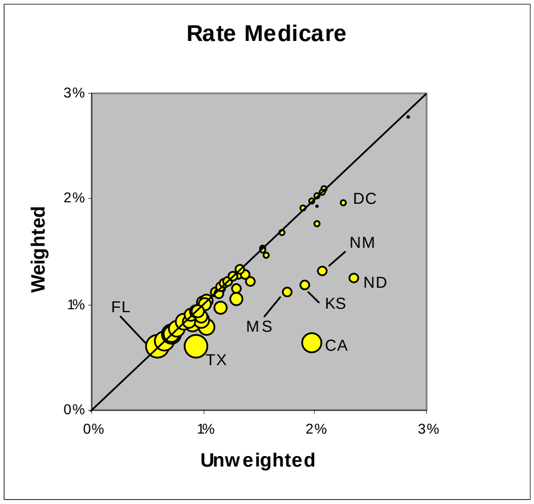
In this section, we present new aspects of the MFFS survey data and suggest general methods for using the data for analytical purposes.

3.2.1 New and Revised Analysis Variables

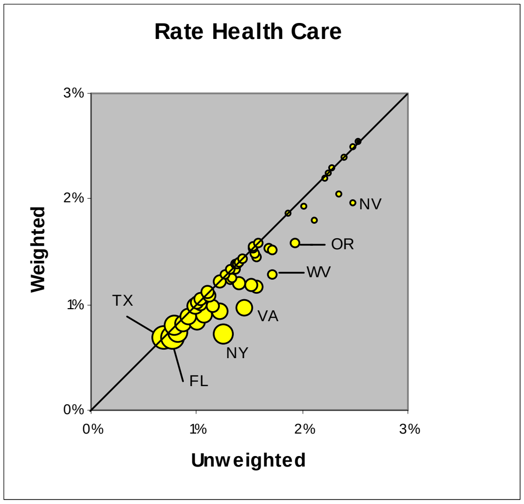
Race/Ethnicity—There was interest in exploring new methods of analyzing race and ethnicity in an attempt to estimate more exclusive groupings of the response categories across two separate questions on the survey: Q84, "Are you of Hispanic or Latino origin or descent?" and Q85, "What is your race?" For these analyses, we constructed several new variables and added a category to the race variable that incorporated responses to the survey question on Hispanic origin. The new race variable, NEWRACE7, has no imputed values. The categories of NEWRACE7 are as follows:

- White
- African American

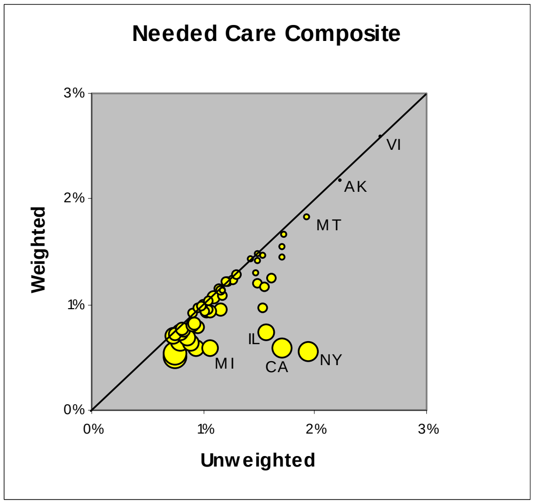
Figure 3.1
Comparison of root MSEs of weighted and unweighted state estimates of CAHPS® outcomes



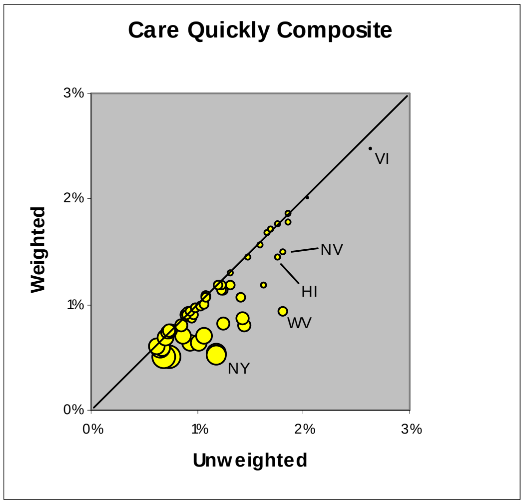
Rate Medicare



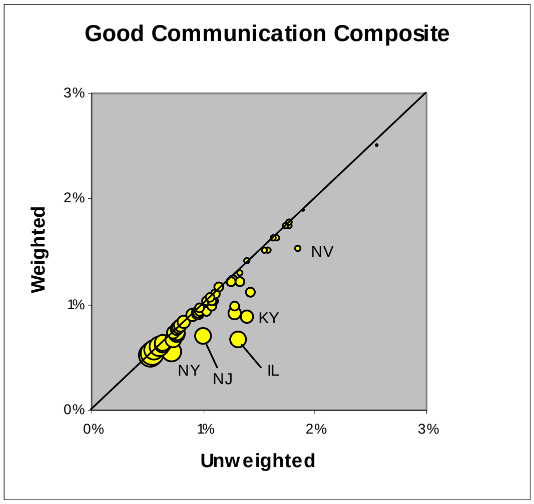
Rate Health Care



Needed Care Composite



Care Quickly Composite



Good Communication Composite

Discussion

Mean Square Error (MSE) formalizes the tradeoff between low bias and low variance, in that MSE = (Bias)^2 + Variance. Plotting SQRT(MSE) allows for the display of variation in the original units of measure (i.e., proportion of Medicare beneficiaries rating the selected indicator most positively). The states with greater numbers of respondents (i.e., larger bubbles) tend to congregate near the origins of these plots. This is to be expected, since variance is reduced as sample size increases. Observations in the lower right half of each plot indicate that a state's weighted estimated proportion has less MSE than its unweighted counterpart. Since most of the observations in these plots fall in this region, we can conclude that the use of weights is beneficial, because they yield more efficient estimates of the true population proportions and preserve statistical power.

- Asian
- Native Hawaiian/Pacific Islander
- American Indian/Alaska Native
- Other/Indeterminate
- Hispanic

We also constructed simplified versions of this new race variable by collapsing some of the categories. NEWRACE6 combines Asians and Native Hawaiian/Pacific Islander:

- White
- African American
- Asian/Native Hawaiian/Pacific Islander
- American Indian/Alaska Native
- Other/Indeterminate
- Hispanic

NEWRACE3 is a further simplification of NEWRACE7 and contains only three categories:

- White
- African American
- Other

The new analytic race variables were created retrospectively for the 2000–2002 MFFS surveys. These variables were constructed using the survey responses, the information available on the CMS EDB, and the variable NEWHISP. NEWHISP is a flag identifying Hispanics from the response to the ethnicity question on the survey. If the response to this question was missing, we then looked at the following to identify Hispanics: whether a Spanish survey was requested, if the individual resided in Puerto Rico, or if the individual had a Spanish surname.

3.2.2 Sample Size and Statistical Significance

Large sample sizes, like the sample sizes for the MFFS surveys, often produce statistically significant results regardless of the sizes of the minimum detectable differences.

Statistical significance should not be confused with *practical significance*. A small real effect, of no practical importance, may be statistically significant in a very large sample. A nonsignificant result does not demonstrate that there is no effect. It means that the data are consistent with there being no effect, and in small samples, this can happen even when the real effect is large. For analyses with large sample sizes, both statistical and practical significance should be considered.

Care should also be taken when performing statistical analyses of cross-classifications of several domains. Some cells in the cross-classifications might turn out to have small sample sizes (for instance, race by education cross-classifications might yield small sample sizes for Asians who have completed less than a high school education). For such cases, we recommend that a result (e.g., estimates, p-values) based on a small sample size be marked with an asterisk denoting it as potentially unreliable (in a statistical sense) if either the sample size is less than a fixed number of individuals or if its coefficient of variation is greater than some designated value. For such cases, we might also recommend the use of relevant small area estimation techniques to produce more reliable estimates.

3.2.3 Trend File

The MFFS trend data file was expanded to include the survey results from all 4 years of the MFFS survey (2000 through 2003). Likewise, the crosswalk that tracks differences in the survey instruments over the years was expanded to include the fourth year. (The crosswalk is contained in the complete report for this task [RTI International, 2004c].) As before, the crosswalk includes information on whether differences in questions across years are differences in question wording and/or differences in skip patterns. Some of the differences in the questions used to calculate the composites and ratings are highlighted below.

In the 2000 survey, the following questions are different from those in the 2001 through 2003 surveys:

1. The first question comprising the Care composite (Question 4 in the 2000 survey), “How much of a problem, if any, was it to get a personal doctor or nurse you are happy with?”, does not get asked if beneficiaries do not answer “yes” to Question 2 asking if they got a new personal doctor. Note that Question 2 does not exist in the 2001 through 2003 surveys, and Questions 3–5 are skipped if Question 2 is “no.”

2. The “rate your personal doctor” question (Question 7) is also different in 2000 because of Question 2. Question 5 is the gateway to Question 7; however, if Question 2 is answered “no,” respondents never get to Question 5 but do get to Question 7, regardless of how they would have answered Question 5. Again, this is different in 2001, 2002, and 2003.

In the 2002 and 2003 surveys, the following questions are different from 2000 and 2001:

1. The “rate your specialist” question in 2002 and 2003 is NOT skipped by the question “Did you or your doctor think you needed to see a specialist?”, but in 2001 and 2000 “rate your specialist” is skipped by that question.

2. The fourth question in the Care composite ("In the last 6 months, how much of a problem, if any, were delays in health care while you waited for approval from Medicare?") can be skipped by two different questions. In 2001 and 2000, one question did not exist.

Users of the trend file should be aware of the differences in the survey instruments through the years. These differences might lead to statistically significant trends that may not be real because of the question differences. The trend file assumes that the samples from each of the survey years are independent of each other and, hence, can be combined as one sample.

3.2.4 Minimum Data Set Institutionalized

We acquired an extract of the November 1, 2003, Long-Term Care Minimum Data Set (MDS) for sample members from the 2000 through 2002 CAHPS® MFFS surveys. The MDS is a standardized, primary screening and assessment tool of health status that forms the foundation for the comprehensive assessment of all residents of long-term care facilities certified to participate in Medicare or Medicaid. The information provided in the MDS allows us to identify sample members who had a nursing home stay before, during, and after data collection. We are not able to collect this information from the MFFS survey sampling frame (i.e., the CMS EDB) alone. However, the MA survey is able to identify and prospectively exclude the institutionalized from its samples, because the information is available on its sampling frame, the GHP.

For each of the samples from 2000, 2001, and 2002, we have created datasets that contain a subset of the 594 variables available on the MDS. From these, we have identified a number of MFFS beneficiaries who have been institutionalized, as well as the timing and duration of their institutionalization(s). We have begun to examine the characteristics of this group and are currently assessing the effects on response rate within the MFFS population. We plan to report our findings in the 2004 Final Report.

SECTION 4
DATA COLLECTION

by Linda L. Dimitropoulos, Ph.D., RTI, and Carol Prindle, Ph.D., RTI

4.1 Introduction

In this section, we describe the data collection protocol and preliminary results for the 2003 Medicare CAHPS® Fee-for-Service (MFFS) Survey. One of the benefits of conducting an annual survey is that it provides an opportunity to learn from each year of the study and implement changes to the methodology. Although we noted a slight decline in response rate in the 2003 survey year, each of the previous 2 years of data collection since the 2000 MFFS survey realized incremental increases in the response rate that can be attributed to modifications in the methodology from year to year. For example, the decision to use a special delivery courier for the third-wave mailing was based on the results of an experiment conducted during the 2001[1] survey year that showed a significant increase in response when the survey was delivered via special delivery than by either U.S. Priority or First-Class mail. We also attribute some of the increase in response to better access to resources for contacting and locating beneficiaries. One of those resources provided telephone numbers for a large proportion of the sample through the Social Security Administration (SSA). This change increased the number of nonrespondents that RTI could contact by telephone during the follow-up phase of the survey and reduced the number of cases that RTI would have to direct to the third-wave mail follow-up.

4.2 Data Collection Schedule

The MFFS survey is a self-administered mail survey that offers sample members the option of calling a toll-free number and completing the survey over the telephone. There is strong evidence that providing multiple modes of responding improves overall response (Dillman, 2000). The follow-up data collection effort for nonrespondents to the mail survey includes a telephone follow-up of nonrespondents for whom RTI had a telephone number and a third-wave survey package mailed to nonrespondents for whom RTI did not have a telephone number.

The data collection period for the 2003 MFFS began with the mailout of the prenotification letter on September 10, 2003, and ended with the close of the telephone follow-up on February 21, 2004. The overall data collection schedule is shown in ***Table 4.1***.

---

[1] The experiment compared response to a third-wave survey sent by U.S. First-Class mail, U.S. Priority mail, or special delivery courier. The results showed that respondents were 2.7 times more likely to return the survey when it was mailed via overnight courier than when it was mailed by either of the other two methods.

Table 4.1
2003 MFFS data collection schedule

Activity	Date
Prenotification letters sent	09/10/2003
Toll-free line opened	09/10/2003
Inbound telephone center opened	09/11/2003
First survey mailed	09/15/2003
First thank you/reminder letter sent	09/22/2003
Cutoff date for Wave 1	10/20/2003
Second survey mailed	11/05/2003
Second thank you/reminder letter sent	11/10/2003
Cutoff date for Wave 2 for telephone follow-up	11/26/2003
Cutoff date for Wave 2 for Wave 3 mailing	12/12/2003
Telephone follow-up began	12/03/2003
Third survey sent (Federal Express)	01/05/2004
Third thank you/reminder letter sent	01/29/2004
Cutoff date for returned mail surveys	02/21/2004
Toll-free line closed down	02/21/2004
Telephone follow-up ended	02/21/2004

4.3 Mail Survey

The data collection plan for the mail survey followed the traditional method of making five contacts: (1) advance letter, (2) first survey package, (3) thank you/reminder postcard, (4) second “replacement” package, and (5) second thank you/reminder postcard. In addition, we mailed a third survey package that, for the first time, was followed by a third thank you/reminder postcard sent to all remaining sample members.

RTI selected a stratified random sample of 178,650 noninstitutionalized MFFS beneficiaries from the CMS Enrollment Database (EDB) (see Section 3 for details). The sample was drawn from 277 geographically distinct areas in the United States, Puerto Rico, and Virgin Islands. Sample sizes in geounits ranged from 506 to 2,000. After the sample was selected, the address file was run through the U.S. Postal Service National Change of Address (NCOA) database, which updates the addresses for those respondents who have moved since the sample was drawn. Any sample file address that was not updated through NCOA was mailed using the original CMS-provided address, which was modified to conform to the standard postal format with three lines of address: name; street address; and city, state, zip code.

NCS Pearson printed all materials and processed and scanned the returned questionnaires. RTI provided oversight to NCS Pearson to ensure that the correct survey procedures were implemented and that specified quality assurance and quality control procedures were followed. NCS Pearson submitted proofs of all materials prior to printing, and all survey materials were reviewed and approved by CMS and RTI before being sent to the respondents.

Quality control is very important to the MFFS, and a number of quality control procedures have been put into place to minimize errors during data collection. The flow of the cases through this complex process is monitored by the use of two sets of disposition codes that are assigned to cases as they progress through the data collection activities. The codes allow staff to track and monitor production and to create status reports. The first set of codes is used to track the mail survey, and the second set of codes is used to track cases that were handled by telephone. The telephone disposition codes provide a means to monitor and report case status and production. The codes are also used to determine next steps in the contacting process.

4.4 Inbound Respondent Calls

All survey materials included the study's toll-free number for respondents to call if they had questions about the study or to request a telephone interview. The toll-free line was in operation 6 days a week, from 8:00 a.m. EST until 8:00 p.m. EST Monday through Friday and from 8:00 a.m. to 5:00 p.m. EST on Saturday. A voice mail system was used to take information from callers after hours and to prompt callers to leave their name, phone number, and ID for later contact by telephone center staff. All voice mail messages were returned the same day (for early morning messages) or the following day (for overnight messages). If a sample member indicated a specific time to call back, this preference was accommodated. During the peak calling period following the first-wave mailing, if the call-in line was busy due to high call volume, a message was placed on the answering machine notifying callers that the system was experiencing a high volume of calls and telling them that they could stay on the line or leave their number for a callback. A maximum of 12 callback attempts was made to each respondent.

4.5 Nonresponse Follow-up

The nonresponse follow-up included both a telephone and mail component and started approximately 3 weeks after the second survey was mailed. The telephone follow-up was conducted with nonrespondents for whom RTI found a telephone number. All other nonrespondents were followed up by sending a third survey package by mail.

4.5.1 Telephone Follow-up

The nonresponse follow-up by telephone began on December 3, 2003, with 68,591 cases that had not responded to either of the first two mailings by November 26, 2003. Telephone interviewing was conducted 7 days a week (with the exception of December 23, 24, and 31, 2003, and January 1, 2004). RTI trained 215 interviewers, including 19 bilingual Spanish-English interviewers.

RTI obtains telephone numbers for sample members through two sources: the SSA database and commercial telephone matching vendors. Frequently, we obtain two different numbers for a sample member from each of these sources, so we include both numbers in the telephone number sample file. The telephone number file for this survey year contained two or more unique phone numbers for 25,960 of the follow-up cases (14.53 percent). If a case had two telephone numbers associated with it, the interviewers dialed the telephone number received from the SSA first. If after six attempts at the first number no contact was made, the second number was attempted. This procedure represents a change from 2002, where the RTI-provided number was dialed first because we did not know how the SSA numbers would perform. This change was made in 2003 based on the experience of 2002 where the highest percentage of completed cases was obtained from the SSA number. When contact was made with Spanish-speaking households, the calls were routed to Spanish-speaking interviewers. Initial soft refusals were reinserted in the calling queue and directed to interviewers who specialized in refusal conversion techniques. In 2003 (as in 2002), refusals were called back 48 hours after the original attempt, not 5 days later as they were in 2000 and 2001. Interviewers were instructed to contact directory assistance if they called numbers that were disconnected or incorrect. If no new number was provided by directory assistance, the call was recorded as “Disconnected—No Directory Assistance Listing.” These cases were then sent back to RTI for tracing on a flow basis. Cases for which RTI’s Tracing Operations Unit (TOPS) found a correct phone number were returned to the telephone center, where further attempts were made to conduct a telephone interview with the sample member.

4.5.2 Third-Wave Mail Follow-up

A final or third-wave survey package was sent to nonrespondents for whom RTI did not have a telephone number. The third-wave mailing was sent approximately 5 weeks after the second-wave mailing to allow for returned surveys from the second wave to be processed. This year, the third-wave questionnaire mailing was followed by a third thank you/reminder postcard. This postcard was different from the Wave 1 and 2 thank you/reminder postcards, which were mailed only to those sample members who received the first- and second-wave questionnaire mailings. The new Wave 3 thank you/reminder postcard was mailed to all remaining sample members (35,014 members), not just to the 2,666 sample members who received the third-wave Federal Express questionnaire packet. The postcard was mailed on January 29 and 30, 2004. The purpose of the extra postcard was to boost response rates.

The acquisition of the telephone numbers from SSA helped to reduce the number of cases requiring a third-wave mailing by more than 50 percent between 2001 and 2002 and even further reduced it in 2003. In 2002, 8,342 cases were eligible for the Wave 3 mailing because they had not responded and we did not have a phone number for them, compared with 17,961 cases in 2001. In 2003, only 2,666 cases were eligible for the third wave, a reduction of more than two-thirds. The third-wave mailing was sent using special 2-day delivery, which differed from 2002 only insofar as 2002 used overnight delivery. Use of special delivery was based on the findings of the experiment conducted during 2001, which showed that sample members were 2.7 times more likely to return the completed survey if it was delivered by a special delivery overnight courier than by U.S. Priority or First-Class mail. In 2003, the third-wave mailing realized a rate of return of completed surveys of 12.7 percent, less than in 2002 (18.6 percent) but more than in 2001 (10.5 percent).

4.5.3 Tracing and Locating

RTI used a number of resources to find new address information for beneficiaries who moved and to locate telephone numbers for the sample members. First, the sample file was matched to the U.S. Postal Service NCOA database. This process updated 6,089 sample addresses, or 3.41 percent of the 2003 sample. This number was somewhat higher than the 3,835 updated addresses found in 2002, representing 2.16 percent of that year's sample.

Only 414 surveys were returned as undeliverable after completion of the 2003 survey. In 2000, there were 2,597 packages returned as undeliverable, and 1,316 packages were returned as undeliverable in 2001. The smaller number of undeliverable packages can be attributed for the most part to the extended schedule.

Since the sample file drawn from the EDB does not include telephone numbers, and because telephone numbers are critical to success in the telephone follow-up, it is necessary to begin the search for telephone numbers as early as possible in the process. In the 2000 and 2001 MFFS, RTI used two outside vendors (Telematch and First Data) and the RTI TOPS unit as resources for telephone numbers. For 2002 and 2003, the project team was able to access additional telephone numbers from the SSA.

The tracing for MFFS is done in an iterative process. In past years, the sample file was first processed through Telematch, a vendor with access to national databases that can match names and addresses with corresponding phone numbers. Cases that were returned without a telephone number were then processed through First Data. The CMS office in Puerto Rico has been instrumental in obtaining telephone numbers for sample members who reside in Puerto Rico and for whom our vendors cannot supply numbers. All remaining cases were sent to TOPS for more extensive case-by-case tracing.

Obtaining the telephone numbers from SSA was the primary reason for the increase in the number of sample member telephone numbers found in both 2002 and 2003. Overall, the 2003 tracing effort produced telephone numbers for 95.91 percent of the sample, which represents an increase of 1.62 percent over 2002.

4.6 Data Collection Results

The response rate among eligible sample members for the 2003 MFFS survey was 69.3 percent, which is 1.3 percent lower than the 70.6 percent response rate achieved during 2002. The response rate varied somewhat by geographic area, but response in each area was sufficient (at least 300 completed cases per geounit) to provide measures of CAHPS® composites and ratings for all 277 geounits and for each of the 50 U.S. states, the District of Columbia, Puerto Rico, and the Virgin Islands.

As noted earlier, the English version of the survey was sent to all sample members except the 1,200 sample members in Puerto Rico, who each received a Spanish survey package. RTI received 2,404 completed Spanish mail surveys, and another 700 Spanish surveys were completed by telephone (inbound or outbound). Of those completing a Spanish survey, 605 (19.49 percent) were known to live in Puerto Rico. The number of completed Spanish surveys was significantly higher in 2003 than in 2002. In 2002, we received only 1,059 Spanish completes, 554 by mail and 505 by phone. The reason for the increase is unknown.

The only sample members considered to be ineligible for the MFFS survey are those found to be deceased during the data collection period of performance. For 2003, the 4,002 deceased accounted for only 2.2 percent of the sample.

Table 4.2 provides a summary of the results of data collection from 2003 by a number of demographic variables. It also provides a comparison to the response rates among eligibles from 2000 through 2003. In 2003, the response rate among eligibles was lower than in 2002 by 1 to 2 percent in almost all categories of respondents.

The biggest drop was an almost 25 percent decrease in the percentage of respondents who categorized themselves as “other or unknown” race. Until this year, the number of respondents categorizing themselves as “other or unknown” race had risen. In 2003, the sample for this category was 7,882 (only 4.4 percent of the entire sample), and a total of 3,849 respondents selected this race category. The response rate among sample members in this category jumped from 49.2 percent in 2000 to 66.5 percent in 2001, rose to 74.2 percent in 2002, then fell to just 50.09 percent in 2003. Another decrease was in the percentage of sample members in the age category of 44 or younger, a decrease from 50.3 percent in 2002 to 46.5 percent in 2003. Also, in past years there was a steady increase in the number of people responding among the Puerto Rican beneficiaries. In 2000 the response rate was 45.6 percent for Puerto Rico, in 2001 the response rate increased to 57.5 percent, and in 2002 the rate again increased to 59.2 percent.

In 2003, however, the number of beneficiaries who responded from Puerto Rico dropped to 55.8 percent. This year also was the first time that Virgin Island beneficiaries were included in the sample, and only 48.1 percent of the 700 sampled beneficiaries responded.

4.7 Recommendations for the 2004 MFFS Survey

We saw a steady increase in response rates over the first 3 years of the MFFS survey, largely because of improvements made to the methodology as a result of our rigorous study of ways to improve efficiency and response rates. The overall response rate in 2000 was 63.9 percent, which increased to 68.0 percent in 2001 and further to 70.6 percent in 2002. The slight decline in 2003 to 69.3 percent does not seem to indicate a problem with the methodology but rather a natural fluctuation in response. Additional adjustments made to the data collection plan included adding 4 weeks to the schedule in 2001 to allow time for processing receipts from the Wave 1 mailing before preparing the Wave 2 mailing. This change reduced overlap mailing considerably and allowed us to take more time to track undeliverable addresses and telephone numbers. In 2003, we added 2 weeks to the end of the schedule to allow time to process all telephone numbers completely through 12 attempts. We also designed a study to explore the costs and benefits of increasing the level of effort (defined as number of call attempts) during the telephone follow-up (Campbell et al., 2004). This study is described in Section 5.

Table 4.2
2003 MFFS survey response rates

	2003 sample		2003 respondents		Response rates among eligibles (%)			
	Count	Percent	Count	Percent	2003	2002	2001	2000
Overall	178,650	100.00	120,974	100.00	69.3	70.6	68.0	63.9
Telephone status								
No number found	7,311	4.09	4,349	3.59	62.8	63.2	64.7	56.4
Number found	171,339	95.91	116,625	96.41	69.5	71.1	69.1	67.5
Gender (EDB)								
Male	76,888	43.04	51,916	42.92	69.2	70.8	68.5	65.0
Female	101,762	56.96	69,058	57.08	69.3	70.5	67.7	63.0
Age group (EDB)								
44 or younger	5,046	2.82	2,318	1.92	46.5	50.3	56.8[a]	51.3[a]
45-64	17,513	9.80	10,655	8.81	61.8	63.9		
65-69	38,850	21.75	27,380	22.63	71.2	73.1	70.0	67.7
70-74	38,200	21.38	27,391	22.64	72.7	74.7	71.8	69.1
75-79	33,670	18.85	24,070	19.90	72.9	73.8	71.2	67.8
80-84	24,526	13.73	16,932	14.00	71.1	71.1	69.9	59.5[b]
85 or older	20,845	11.67	12,228	10.11	62.8	62.9	60.7	
Race (EDB)								
White	154,548	86.51	107,828	89.13	71.30	71.5	69.4	66.0
Black	16,220	9.08	9,297	7.69	59.13	59.1	55.6	51.0
Other	7,882	4.41	3,849	3.18	50.09	74.2	66.5	49.2
Dually eligible (EDB)								
Yes	24,447	13.68	14,243	11.77	60.49	61.3	56.9	50.5
No	154,203	86.32	106,731	88.23	70.64	72.0	69.6	66.1
Beale code[c]								
Central metro counties 1M+ pop	54,547	30.86	34,080	28.41	64.00	64.7	61.8	58.4
Fringe metro counties 1M+ pop	6,809	3.85	4,688	3.91	70.30	72.6	70.1	65.9
Metro counties of 250K to 1M+ pop	42,058	23.80	28,489	23.75	69.28	70.5	68.7	63.5
Metro counties of <250K pop	19,525	11.05	13,801	11.50	72.25	74.3	71.3	67.4
Urban pop of 20K +, adjacent to metro area	9,130	5.17	6,436	5.36	71.83	74.2	70.9	68.1
Urban pop of 20K +, not adjacent to metro area	6,470	3.66	4,633	3.86	73.16	75.5	72.6	67.0
Urban pop of 2,500 to 19,999, adjacent to metro area	16,294	9.22	11,710	9.76	73.60	74.7	72.0	68.4

(continued)

Table 4.2
(continued)

	2003 sample		2003 respondents		Response rates among eligibles (%)			
	Count	Percent	Count	Percent	2003	2002	2001	2000
Urban pop of 2,500 to 19,999, not adjacent to metro area	14,368	8.13	10,564	8.81	75.14	76.1	73.8	68.9
Completely rural, adjacent to metro area	2,612	1.48	1,905	1.59	74.44	76.2	72.6	69.2
Completely rural, not adjacent to metro area	4,827	2.73	3,574	2.98	75.75	77.6	74.9	70.1
Census division								
New England	12,532	7.01	8,402	6.95	68.48	69.8	67.5	62.7
Middle Atlantic	23,204	12.99	14,433	11.93	63.81	64.8	62.8	59.7
East South Central	13,866	7.76	9,620	7.95	71.09	72.1	70.8	64.9
West South Central	17,720	9.92	12,016	9.93	69.44	70.3	66.0	63.8
South Atlantic	36,901	20.66	24,738	20.45	68.61	70.6	67.7	63.9
East North Central	27,713	15.51	19,492	16.11	71.84	72.9	70.7	65.9
West North Central	17,357	9.72	12,710	10.51	74.86	76.4	74.7	70.5
Pacific	16,647	9.32	10,814	8.94	66.39	68.1	64.4	61.9
Mountain	10,810	6.05	7,742	6.40	72.97	73.8	72.5	67.4
Puerto Rico and Virgin Islands[d]	1,900	1.06	1,007	0.83	54.23	59.2	57.5	45.6

[a] In the 2000 and 2001 MFFS surveys, beneficiaries aged 64 or younger comprised the youngest age category available as a response to the question, “What is your age now?” Beginning in 2002, this age category was further broken out into responses for beneficiaries aged 44 or younger and beneficiaries aged 45 to 64.

[b] In the 2000 survey, beneficiaries aged 80 or older comprised the oldest age category available as a response to the question, “What is your age now?” Beginning in 2001, this age category was further broken out into responses for beneficiaries aged 80 to 84 and beneficiaries aged 85 or older.

[c] Puerto Rico and Virgin Islands have been excluded from percentage calculations because Beale codes are not assigned in these regions.

[d] Virgin Islands included in the 2003 sample only.

NOTE: EDB = CMS Enrollment Database.

SECTION 5
COSTS AND BENEFITS OF IMPROVING RESPONSE RATES IN THE MFFS SURVEY

by Larry N. Campbell, M.S., RTI; Linda L. Dimitropoulos, Ph.D., RTI;
G. Gordon Brown, Ph.D., RTI; and Lisa A. Carpenter, B.S., RTI

5.1 Introduction

This section presents the results of a study conducted to explore the costs and benefits of increasing the level of effort (defined as number of call attempts) during the telephone follow-up phase of the Medicare CAHPS® Fee-for-Service (MFFS) Survey, an annual mail survey of beneficiaries enrolled in fee-for-service Medicare. It is critical that the survey reach an appropriate level of response to conduct the necessary subgroup analyses. The telephone follow-up is particularly important to the MFFS survey because we know that beneficiaries who respond by telephone tend to respond more favorably on the CAHPS® ratings of Medicare (Pugh et al., 2002). The project team has conducted a number of studies to identify methods to increase response rates by using resources efficiently.

The 2003 MFFS survey was conducted primarily as a mail survey with an inbound computer-assisted telephone interview (CATI) option. Beneficiaries who did not respond to either of the first two mailings were placed into a telephone follow-up group if a telephone number could be obtained for them. Beneficiaries for whom a telephone number could not be obtained were followed up by mail. The nonresponse follow-up, by telephone or special delivery mail, is the final step in the standard mail survey methodology and has been shown to reliably boost response among nonrespondents to earlier attempts (Dillman, 2000).

An important question with respect to the telephone follow-up is, “What is the maximum number of calls that should be made to sufficiently reduce nonresponse bias?” This question assumes that the added effort will allow us to penetrate a segment of the population that was not adequately represented and, further, that the answer is important even in the presence of mode response bias. This study is an attempt to answer the following research questions using the 2003 CAHPS® MFFS data:

1. Are we penetrating an underrepresented population segment by increasing the level of effort with additional calls?

2. At what point do we experience diminishing returns with the number of call attempts sufficient to cap the level of effort?

3. What are the cost implications of increasing the level of effort?

4. Are the CAHPS® measures affected by the additional responses?

5.2 Methods

5.2.1 Data Issues

The telephone center supplied a file containing a detailed call history for each beneficiary participating in the telephone follow-up. Each telephone call attempt was assigned a status code. The type of code indicated whether another attempt would be made. Complete and final codes, such as refusals, beneficiaries found to be institutionalized, and beneficiaries not available for the duration of the survey, resulted in no further contact. In contrast, a pending code, such as no answer, busy, or not available at the moment, generated additional attempts.

To determine the number of attempts for each sample member, we used the counter from the CATI software, which follows industry standards. At the beginning of the telephone follow-up, we set a maximum of 10 attempts. However, midway through the process, we decided to extend data collection for 2 weeks and increase the maximum number of call attempts to 12. The analysis was conducted using 67,413 valid cases. Demographic variables used in the modeling, including race, age, gender, dual eligibility, and disability status, came from the CMS Enrollment Database (EDB) and are available for all sample members.

5.2.2 Why Survival Analysis?

We used survival analysis—specifically, the discrete time proportional hazards model—to address whether we were penetrating an underrepresented population segment and to what degree we experienced diminishing returns with the number of call attempts. Survival analysis is appropriate for “time to event” data when there are censored observations in the data. The event in this instance is an individual who is contacted and completes the survey. Final codes, those that result in no further contact but are not a complete, were considered censored.

5.2.3 Empirical, Predicted, Conditional, and Cumulative Probabilities

We used two distinct concepts in calculating response rates or response probabilities. The first is conditional probabilities, which are the response rates given that the individual is still in the population at a given attempt. The cumulative probability is the probability of a response at a given attempt given that the person is in the population at the beginning of the study. The conditional response rate answers the question, “Given that a person is still in the population at the k^{th} attempt, what is the probability that they respond at the $k+1^{th}$ attempt?” The cumulative response rates answer the question, “Given that a person is in the population at the beginning of the study, what is the probability that they respond on the k^{th} attempt?”

Two methods are used to estimate conditional or cumulative response rates. The first method calculates the response rates without controlling for any covariate or making any adjustments for the sample design. These response rates are referred to as the empirical response rates or empirical response probabilities.

The second method calculates modeled response rates by controlling for all covariates of interest and using the sample weights. These response rates are referred to as the predicted response rates or predicted probabilities. The predicted response rates that were based on the models tended to be similar, if not identical, to the empirical response rates. As a result, we used both the empirical and predicted response rates where appropriate.

5.2.4 Modeling

We used SUDAAN's SURVIVAL and LOGISTIC procedures to fit all models and produce all statistics of interest. For a description of the discrete proportional hazards model, see the SUDAAN Language Manual (RTI International, 2004d). We elected to test the proportionality assumption inherent to survival models. The proportionality assumption states that the regression coefficients based on the covariates are independent of the number of attempts. We found no evidence to indicate that the proportionality assumption was being violated.

5.2.5 Attempts and Covariates

A way of assessing the impact of multiple calls on the CAHPS® measure was to examine interactions between demographic variables of interest and the response variables that comprise the CAHPS® composites. To conduct this analysis, we used only individuals that completed the follow-up survey and re-coded the responses to dichotomous variables. We then used these dichotomous variables in logistic regression as the response variables and used "attempts" as a covariate. If the "attempts" variable was statistically significant in the model, then it would provide evidence that the response was a function of the number of attempts.

5.3 Results

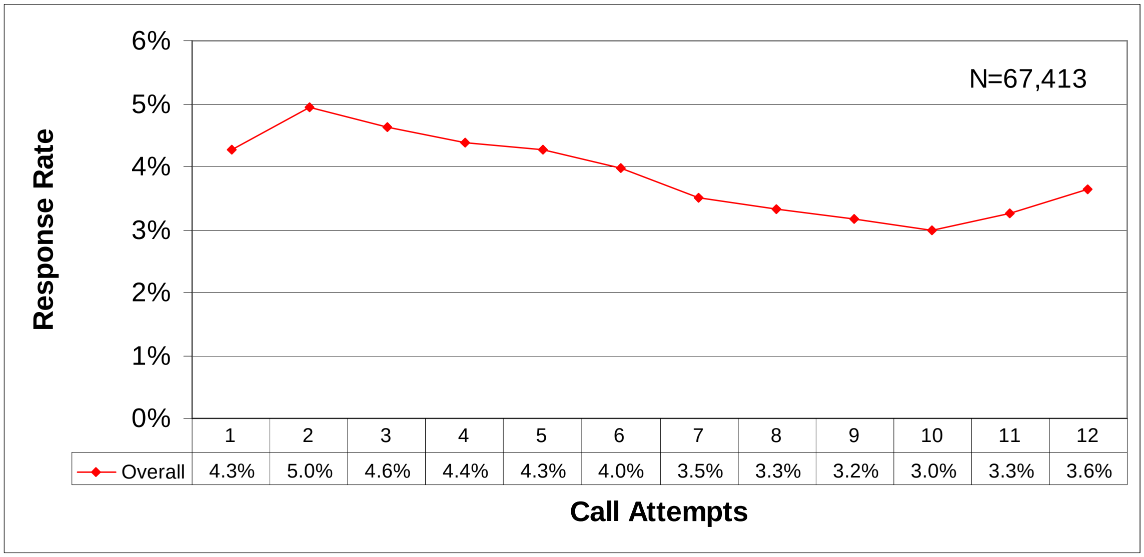
5.3.1 Diminishing Returns

The overall response rate tends to diminish, as illustrated by the predicted response probabilities shown in ***Figure 5.1***. The mean response probability for the first six call attempts is 4.41 percent, whereas the mean for the last six is 3.31 percent, or roughly a 1.1 percent drop. Conducting a simple T-test of the mean of the first six compared with the mean of the last six gives a p-value of 0.00013. One note of interest is a slight increase in the conditional response probabilities for the 11th and 12th attempts. The response probabilities are distinctly declining from the 2nd to the 10th attempts. The 11th and 12th attempts have values of 3.3 percent and 3.6 percent, respectively. Potential reasons for this upturn are hypothesized in Section 5.4.

5.3.2 Effects on CAHPS® Measures

The logistic regression models indicate that most of the CAHPS® measures are unaffected by the number of call attempts. For all of the models that we fit, the linear and quadratic terms for "attempts" were not significant factors in predicting CAHPS® measures. With one exception, the tests of the interactions of the covariates with attempts were not significant. The only significant interaction was the "Age*attempts" interaction. However, this interaction was only significant as a quadratic effect for Question 20. We surmise that this significant effect was an artifact of the data and will not be present in future years.

Figure 5.1
Predicted conditional response rate for each attempt



5.3.3 Underrepresented Population

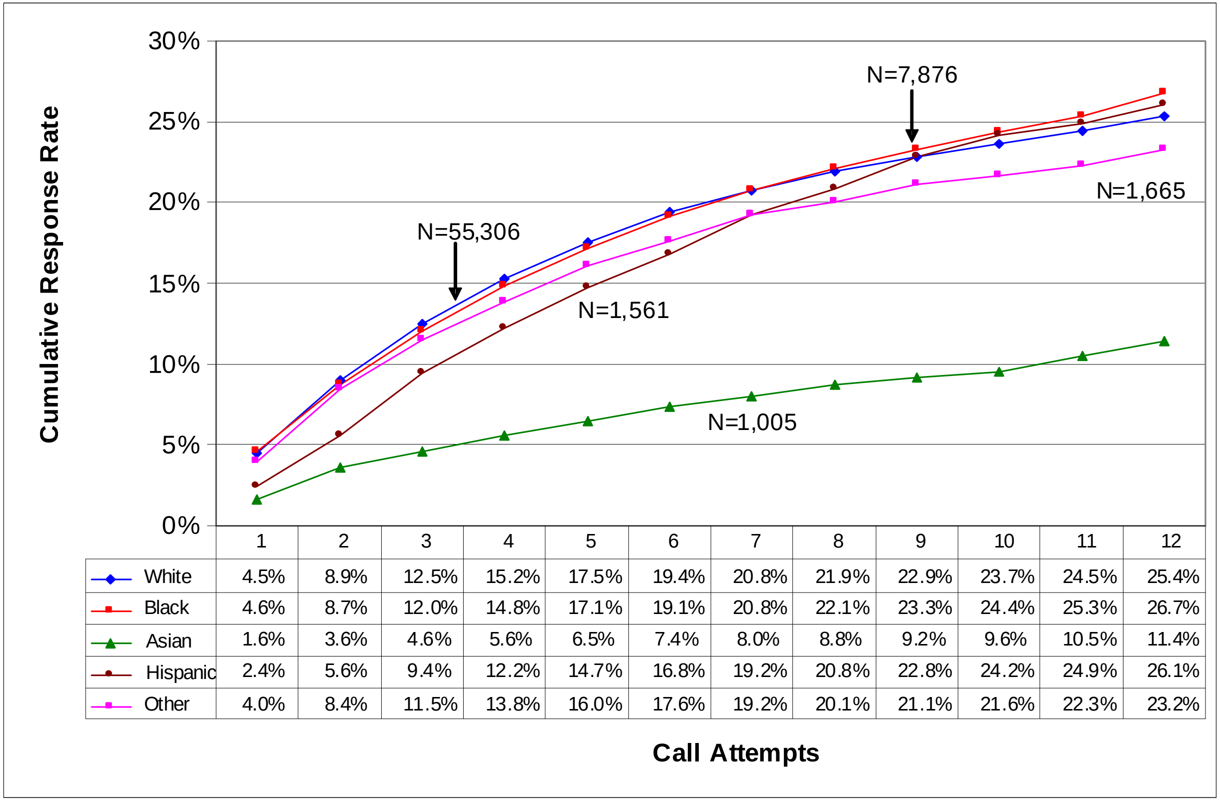
We found little evidence indicating that segments of the population are being sampled at different rates for the later call attempts versus the earlier attempts. ***Figure 5.2*** depicts the empirical cumulative response for race. As demonstrated by the graph, all of the races tend to follow the same pattern. Response rates tend to be higher in the earlier attempts and tend to decline as attempts increase. While Asians tend to respond at a lower rate, none of the paths show a significant difference from the others. We also found this type of trend to be true for age, gender, urbanicity, dual eligibility, and disabled status.

5.4 Discussion

It is possible that additional call attempts, particularly when the number of attempts approaches the maximum, will yield responses from a group of beneficiaries whose demographics are different from those who responded to earlier attempts. Because it is desirable to have the respondents fully represent the population sampled, it makes sense to consider additional call attempts to reach an underrepresented population segment.

With respect to the demographic variables we examined—age, race, gender, dual eligibility, and disability—we did not find that additional call attempts resulted in penetrating a population segment with different demographic characteristics. We observed the same general result for age, gender, dual eligibility, and disability, further supporting the assertion that response distributions resulting from a maximum number of call attempts do not differ from those distributions that result from fewer attempts. Thus, there is insufficient justification for the use of increased call attempts as a strategy for reaching an underrepresented segment of this population.

Figure 5.2
Empirical overall cumulative response by race



	1	2	3	4	5	6	7	8	9	10	11	12
◆ White	4.5%	8.9%	12.5%	15.2%	17.5%	19.4%	20.8%	21.9%	22.9%	23.7%	24.5%	25.4%
■ Black	4.6%	8.7%	12.0%	14.8%	17.1%	19.1%	20.8%	22.1%	23.3%	24.4%	25.3%	26.7%
▲ Asian	1.6%	3.6%	4.6%	5.6%	6.5%	7.4%	8.0%	8.8%	9.2%	9.6%	10.5%	11.4%
● Hispanic	2.4%	5.6%	9.4%	12.2%	14.7%	16.8%	19.2%	20.8%	22.8%	24.2%	24.9%	26.1%
■ Other	4.0%	8.4%	11.5%	13.8%	16.0%	17.6%	19.2%	20.1%	21.1%	21.6%	22.3%	23.2%

Call Attempts

The economic principle of diminishing returns is relevant to the interpretation of our findings. We can expect that, at some point, additional call attempts will produce a negligible improvement in response rate and will therefore cease to be a cost-effective strategy.

To produce Figures 5.1 and 5.2, we used the CAHPS$^{®}$ MFFS survival modeling of the predicted conditional response rate for each call attempt. The down sloping curve in Figure 5.1 indicates a slight decline in response rate with additional attempts. However, on the basis of response rate alone, cost aside, there is no point on this curve beyond which one would not choose to continue with further attempts, since each further attempt is producing a significant response.

We believe the slight tailing up of the curve for attempts 10 through 12 is an artifact of our having increased the maximum number of calls from 10 to 12 attempts midway through the telephone follow-up. We also used refusal conversion specialists for many of the calls for attempts 10 through 12.

A second critical factor in determining the number of call attempts to make is cost. ***Table 5.1*** presents a cost analysis that allows us to evaluate the benefit of an additional telephone attempt using the marginal costs of a completed survey.

Table 5.1
Cost analysis

Attempt	Marginal cost per complete (units)
1	35
2	34
3	36
4	38
5	38
6	41
7	44
8	46
9	48
10	50
11	46
12	38

NOTE: Total number of calls: 500,000; average call attempt cost: 1 unit; completed survey cost: 13.7 units.

Using the total number of call attempts, 500,000, and assigning a relative cost of 1 unit for the average call attempt and 13.7 units for a completed telephone survey, we calculated the marginal costs per completed survey for each number of attempts. For instance, in Table 5.1, the marginal cost per completed survey is 34 units for attempt number 2. This is the incremental cost, above all attempt number 1 costs, to complete each survey with a second attempt.

As with the response rate, if there were significantly diminishing returns for additional attempts with respect to cost, we would observe that the marginal cost per completed survey increased substantially with additional attempts. Although we do observe some increase in marginal cost per complete, it is only a moderate increase. This finding suggests that, if the budget is sufficient to support additional attempts, there will not be a significant increase in the cost per completed survey.

Our final question is whether we obtain different answers to the core CAHPS® questions from beneficiaries who answer after many call attempts than from those who answer after fewer attempts. Without such a difference in the CAHPS® measures, we could remove the concern of response bias, and our desire for precision would remain as the driving statistical factor for the number of attempts to make. Our modeling of the CAHPS® measures, taking into account the same demographic variables of age, race, gender, dual eligibility, and disability status, indicated that the number of attempts had no statistically significant effect on the CAHPS® outcomes.

5.5 Conclusion

When choosing the number of attempts to make in the telephone follow-up phase of the CAHPS® MFFS annual survey, whether response bias is at issue (underpenetrated population segments) or a concern for diminishing returns and cost, it makes sense to set the maximum number of attempts at 12. Without either a significantly diminishing response rate or increasing marginal costs for a completed survey, the threshold for the maximum number of attempts, where the costs exceed the benefits, is beyond 12. As long as the budget is sufficient, the maximum number of attempts on this survey should be set no lower than 12, satisfying both scientific and cost concerns.

SECTION 6
CASE-MIX ADJUSTMENT

by Marc N. Elliott, Ph.D., RAND

The Medicare CAHPS® Fee-for-Service (MFFS) Survey project is centered around two types of comparisons: beneficiary comparisons of MFFS and Medicare Advantage (MA) within local areas and administrative comparisons of MFFS across local areas. Case-mix adjustment (CMA) is a central element in these comparisons. CMA attempts to remove from ratings and reports of care response patterns that are systematically associated with such patient-level characteristics as demographics, socioeconomic status, and general health status, which may vary considerably across reporting units. These systematic patterns of association may reflect "response bias," response patterns that do not correspond to actual differences in quality of care. In any event, these are patient characteristics that are generally agreed to be beyond the control of providers or plans once they have been selected by beneficiaries. The goal of CMA can therefore be thought of as follows: to estimate the ratings and reports that a plan or collection of MFFS providers would have received if all providers and plans treated the same standardized population of patients (Medicare beneficiaries). This adjustment should make attributions of ratings and reports to MFFS providers and MA plans more appropriate, supporting better decision making by beneficiaries and quality improvement by quality improvement organizations (QIOs) and CMS.

The two goals of MFFS CMA (within-MFFS comparison and MFFS-vs.-MA comparison) suggest two different, but similar, CMA models. ***Table 6.1*** describes the independent variables recommended for case-mix adjustment. This set of variables is the same as that used in the previous year.

The present study found that the case-mix adjusters employed in 2001 and 2002 MFFS-vs.-MA CMA (age, education, self-rated health status, self-rated mental health status, and proxy respondent status[1]) constitute an effective case-mix model for both comparison purposes. Self-rated health, self-rated mental health, and education were the three most important CMA variables. An indicator of dual eligibility further enriches the within-MFFS model. These findings are consistent with CMA results for 2000, 2001, and 2002.

Within-MFFS CMA employs the above independent variables plus dummy variables corresponding to the geographic units being compared (county-based sampling stratum, state, or CMS region) in a linear regression. In these regressions, CAHPS® ratings in reports serve as dependent variables, sometimes in their original forms and sometimes dichotomized to correspond to displays of data to consumers. Although age is important for adjusting the rating of Medicare, the most important CMA variables for within-MFFS CMA in 2002 and 2003 were education and self-rated mental health.

---

[1] Although proxy respondent status has only a small empirical effect on CMA, it has been included because many stakeholders feel it is important for the face validity of CMA.

Table 6.1
Independent indicator variables used in MFFS case-mix adjustment, 2003

SAS variable name	Description	Response options
AGE (AGE44, AGE4564, AGE6569, AGE7579, AGE8085, AGE85)	Age	<44, 45-64, 65-69, 70-74, 75-79, 80-85, >85
EDUC (LESS8GRD, SOMEHIGH, SOMECOLL, COLLGRAD, COLLMORE)	Education	<8th grade, some high school, high school graduate or GED, some college (but less than 4-year degree), 4-year college graduate, >college graduate (some graduate school beyond the 4-year degree)
GHP (EXCEL, VERYGOOD, FAIR, POOR)	General health perception	Excellent, very good, good, fair, poor
MHP (MHEXCEL, MHGOOD, MHFAIR, MHPOOR)	Mental health perception	Excellent, very good, good, fair, poor
(PROXY, ANSPROXY)	Proxy respondent status	No assistance on survey, someone helped but did not answer for you, someone answered for you
DUALELIG[a,b]	Dual-eligibility indicator (eligible for Medicaid program)	Yes, no

[a] CMS data contain the indicator of state buy-in, which is a proxy for dual-eligibility status; state buy-in can exist for an individual who is not actually on Medicaid.

[b] Recommended for within-MFFS use only.

In MFFS-vs.-MA CMA, these same variables from Table 6.1 (minus the dual-eligibility indicator) also serve as independent variables in a linear regression, but dummies correspond to MA plans, with MFFS treated as an additional “plan.” Although the direction of CMA coefficients is similar for MFFS and MA, the magnitudes of the effects sometimes differ. In 2000 and 2001, the well-established tendency of healthier beneficiaries[1] to rate their care more positively or to report better health care experiences was considerably stronger in MA than in MFFS, with MA slopes generally 50 percent to 100 percent larger than MFFS slopes for the general self-rated health item for most subjective global ratings and many

---

[1] As measured by general self-rated health.

objective report items. In other words, ratings and reports of one's health care were considerably more sensitive to one's (general) health status in MA than in MFFS. In 2002 and 2003, this pattern was largely restricted to the global ratings. If this is a reliable trend, and if one considers the report items to be more objective, one possible interpretation of these findings would be that health status–based differences in MFFS and MA experiences may be diminishing, though not the perceptions of those differences. Interestingly, the self-rated mental health item did not follow this pattern: mentally healthier beneficiaries reported more positively than less mentally healthy beneficiaries to the same extent in MFFS and MA from 2000 through 2003.

A major implication of the difference in general health status coefficients is that the difference between the case-mix adjusted mean of an MA plan and an MFFS reporting entity depends on the reference population. Case-mix adjustment to a healthy reference population would be relatively more favorable to MA, and case-mix adjustment to an unhealthy reference population would be relatively more favorable to MFFS. In Medicare Compare consumer materials, from 2000 through 2003, MFFS-vs.-MA CMA used the midpoint of MFFS beneficiary and MA beneficiary characteristics as the reference population. Because of the generally poorer health status of MFFS beneficiaries (even excluding the dually eligible), the general health perception (GHP) component of CMA tends to adjust in favor of MFFS relative to MA.

In comparing MFFS and MA, there was concern that underlying geographic factors not captured in a case-mix model might inappropriately influence MFFS-vs.-MA comparisons. To ensure geographic equivalence of state-level comparisons, county-based geographic equivalence weights (GEW) were created in the 43 states[1] where MA exists. These weights were then combined with MFFS nonresponse weights.

Comparison weights have gone from moderate adjustments in favor of MA in 2001 to very small in adjustments in 2002 and 2003. One interpretation is that the MFFS sample was initially scarce in the geographic regions that had the least positive Medicare experiences among those regions with MA penetration. The shrinking effect of the comparison weights may be attributable to the reallocation of MFFS sample into the counties with high MA penetration but low population that were initially underrepresented, in the efforts to reduce the comparison weights design effect. In other words, the geographic distribution of the MFFS sample is much better matched to MA in 2003 than it was in 2001.

The impact of case-mix adjustment on within-MFFS comparisons has remained moderate. The adjustments for the most affected states are quite substantial for both between-state comparisons of MFFS and within-state comparisons of MFFS with MA. Nationally, case-mix adjustment has gone from moderate adjustments in favor of MA in 2001 to small adjustments in favor of MA in 2002 to moderate adjustments in favor of MFFS in 2003. A similar pattern exists for case-mix adjustment of state-level comparisons of MA and MFFS, except that the amount of adjustment of these estimates by CMA is notably larger in 2003 than in 2001 and 2002.

---

[1] Including the District of Columbia.

Adjustments favoring MA probably correspond to MA having a higher proportion of certain types of negative responders: the young and the better educated. Adjustments favoring MFFS probably correspond to MFFS having a higher proportion of a different class of negative responders: the unhealthy. The shift from adjustments favoring MA to adjustments favoring MFFS could mean that age and education selection into MA is becoming weaker, but health selection is becoming stronger. Future research should investigate trends in MFFS-vs.-MA case-mix demographics.

SECTION 7
ESTIMATING THE EFFECTS OF PROXY RESPONDENTS IN MEDICARE CAHPS®

by Marc N. Elliott, Ph.D., RAND, and Kelly Chong, UCLA
School of Public Health

Many health surveys, such as the Medicare CAHPS® Fee-for-Service (MFFS) Survey, allow the use of proxy respondents or assistance to ensure that the experiences of the most vulnerable seniors are not omitted. On these surveys, assistance may range from help with reading or writing to having a proxy, such as a family member, respond in the place of the senior. The question that naturally arises is whether proxies give responses comparable to what would have been self-reported.

Proxy respondents can be the exception or the rule. General surveys of seniors, such as the MFFS survey, may need proxies only for a vulnerable minority of the population. Other applications, such as a family member survey of nursing home experiences, may deliberately seek proxy responses for a majority of respondents or even all respondents.

This section reports the results of an analysis we conducted of the effects of proxy respondents in Medicare CAHPS®.

7.1 Methods: Data, Case-Mix Adjustment, and Propensity Score Weighting

Our analysis was based on the 2001 MFFS survey. Of the 110,215 respondents, 5 percent had someone respond for them (proxy respondent), 14 percent received help from a proxy (proxy assistance), and 81 percent received no help (no proxy). Proxy assistance includes someone reading the survey to the respondent, writing responses for the respondent, or translating languages for the respondent. Proxy assistance might still allow for proxy influence in that the proxy is discussing the items with the respondent. In our study, spouses and life partners comprised 43 percent of all proxies, other family members comprised 46 percent (two-thirds of these being children), and the remaining 11 percent were not family members. We examined 27 outcomes for proxy effects: the four global ratings and the 23 report items that constitute the composites reported on the Medicare Personal Plan Finder web page (www.medicare.gov/MPPF/DefaultVersion/home.asp). Global ratings are overall subjective assessments, whereas report items are more specific and thought to be more objective. We disaggregated composites into report items to examine proxy effects at the item level.

Traditionally, case-mix adjustment (CMA), a regression-based approach, has been used to adjust for proxy effects. Medicare CAHPS® uses CMA, controlling for age, health status (overall and mental), and education and distinguishing proxy respondent and proxy assistance cases from no proxy cases. This approach assumes that all observations are equally relevant to the comparison and that the effects of age, health status, and education are constant across the whole range. A potential weakness of this approach in examining proxy effects is that those needing proxy assistance are fairly different from those not needing such assistance in terms of health status and other important covariates.

Propensity score weighting (PSW) models the probability of needing proxy assistance from exogenous predictors and then gives greater weight to the respondents most relevant to the comparison (those who resemble the other proxy statuses in their characteristics). In a way, PSW restricts the comparison to respondents who were "on the border" of needing proxy help. The standard CMA models are run with these weights. We included this technique in case the standard technique was misestimating proxy effects. In the cases that show little difference, we can conclude that the standard CMA was unbiased.

7.2 Building a Propensity Score Model

The trinomial outcome (proxy respondent/proxy assistance/no proxy) was fit with multinomial logistic regression. Candidate predictors were survey-reported age, education, race/ethnicity, general health perception, mental health perception, SF-12 physical score, and SF-12 mental score. Because there could have been proxy effects on predictors, we eliminated mental health perception, the one predictor for which proxy respondent and proxy assistance answers differed. This approach may have been overly conservative, since true differences might have existed between those groups, but we wanted to err on the side of excluding predictors from the PSW model that themselves might be subject to strong proxy effects. Bivariately significant terms were retained for a multivariate model. Ordinal variables were treated as categories if they departed significantly ($p<0.05$) from linearity. Continuous predictors were tested for quadratic effects, which were retained if significant. All significant two-way interactions were retained. Multinomial logistic regression generated three predicted probabilities for each observation summing to one. Propensity score weights were the inverse of the predicted probability for the group to which a respondent belonged. For example, a person with no proxy help might have predicted probabilities of 0.1, 0.4, and 0.5 for proxy respondent, proxy assistance, and no proxy, respectively. In that case, the weight would have been 1/0.5, since the 0.5 corresponds to the category to which the person actually belonged.

Secondary analyses used CMA to compare the three proxy statuses among those with proxy respondents, with and without PSW.

7.3 Results

Table 7.1 summarizes the effects of proxy respondents and proxy assistance for global ratings and report items, using simple (unweighted) CMA and PSW CMA.

As can be seen, proxies are associated with consistently less favorable global ratings and generally less favorable reports. Using PSW CMA scores as a reference, simple CMA appears to overestimate proxy effects by 22 percent to 43 percent for ratings and 60 percent to 100 percent for reports. Proxy effects are less than half as large for report items as they are for global ratings and are about half as large for proxy assistance as for proxy response.

Table 7.1
Summary of proxy effects

	Global ratings		Report items	
	Simple CMA	PSW CMA	Simple CMA	PSW CMA
Proxy respondent	4/4 lower	4/4 lower	13/23 lower 4/23 higher	9/23 lower 4/23 higher
	0.22 std. mean effect size	0.18 std. mean effect size	0.08 std. mean effect size	0.05 std. mean effect size
	p<0.0001	p<0.002	p<0.05	p<0.05
Proxy assistance	4/4 lower	3/4 lower	15/23 lower 1/23 higher	6/23 lower 3/23 higher
	0.10 std. mean effect size	0.07 std. mean effect size	0.06 std. mean effect size	0.03 std. mean effect size
	p<0.01	p<0.01	p<0.05	p<0.05

NOTE: CMA = case-mix adjustment; PSW = propensity score weighting.

Spouse proxy respondents were more positive than other family members and nonfamily members; other family members and nonfamily members were virtually identical in their proxy respondent effects. Furthermore, PSW had essentially no effect on these estimates. Spouses were significantly more positive than other proxy respondents for all four global ratings (p<0.001, mean effect size 0.34 std. dev., largest effect size 0.50 std. dev.) and for 16 of 23 report items (p<0.05, mean effect size 0.16 std. dev., largest effect size 0.45 std. dev.). Synthesizing the primary and secondary analyses reveals that, after PSW, spouse proxy respondents have very little deviation from expected self-report (a mean effect size of less than 0.03 std. dev. for both ratings and reports). Almost all deviations of proxy respondents from expected self-report are attributable to nonspouse proxies, who differ by an average effect size of 0.37 std. dev. for ratings and 0.14 std. dev. for reports. In other words, spouses answer very much how the beneficiary would have been expected to answer, but other proxy respondents rate the experiences substantially less positively than the beneficiary likely would have. These deviations average together to produce small to moderate proxy respondent effects.

7.4 Summary

Proxy respondent effects are generally very small for objective report items but can be somewhat larger for global ratings, especially when proxies are not spouses. Proxy assistance effects are very small. Standard CMA can overestimate proxy effects by failing to compare beneficiaries with comparable health status. Spouses are very accurate proxy respondents.

Proxy effects are likely to have substantial CMA effects only on surveys in which proxy respondents account for a large proportion of response. These effects are likely to be small on the Medicare CAHPS® surveys but might be larger for subreports on the disabled or those in fair to poor health, or for surveys of nursing home residents. Objective reports are preferable to subjective ratings when using proxy respondents. When available, spouses should serve as proxies. Adjustment should be made for nonspouse proxy respondents if these are a substantial proportion of all responses. This adjustment will not be possible in the 2004 Medicare CAHPS® surveys, which lack information on the relationship of the proxy to the beneficiary. There may be some value in reintroducing a simplified version of the proxy relationship question that distinguishes between spouse and nonspouse proxies. Propensity score weighting or matching can improve the estimation of these adjustments somewhat.

SECTION 8
SUBGROUP ANALYSES

by Shulamit L. Bernard, Ph.D., RTI; Erica R. Brody, M.P.H., RTI;
and Nathan D. West, M.P.A., RTI

One of the key objectives of the Medicare CAHPS® Fee-for-Service (MFFS) national implementation project is to provide information to help beneficiaries choose among health plan options. Currently, through surveys of the Medicare Advantage, or MA (formerly Medicare Managed Care, or MMC), enrolled population and beneficiaries in Original (or fee-for-service) Medicare, beneficiaries residing in areas in which there is a choice of plans are able to access data comparing CAHPS® indicators of patient satisfaction with MA and MFFS plans (http://www.medicare.gov). In this section, we provide the results of our analyses of data from the 2003 CAHPS® MFFS survey. (The complete survey instrument is provided in Appendix A.)

We performed analyses of Medicare subgroups to gain a better understanding of the differences in health services experience and satisfaction among Medicare beneficiaries by geographic levels (national, regional, and state level), sociodemographic characteristics, health plan options, and health status. The MFFS population is quite heterogeneous in terms of demographic characteristics, region of residence, supplemental insurance (whether with or without prescription drug coverage or Medicaid), and health-related characteristics. These subgroups of the MFFS beneficiaries have vastly different experiences with and expectations of the health care system and, thus, may perceive the quality of and access to services differently.

The analyses presented in this section examine differences across selected subgroups for the most-positive CAHPS® ratings and responses (i.e., “10,” “Always,” “Not a Problem,” or “Yes”). A total of nine performance indicators (five composite indicators and four rating indicators) were used from the 2003 CAHPS® MFFS survey:

- Needed Care Composite
- Good Communication Composite
- Care Quickly Composite
- Respectful Treatment Composite
- Medicare Customer Service Composite
- Rate Personal Doctor
- Rate Specialist
- Rate Health Care
- Rate Medicare

For each indicator, the results are stratified by the following MFFS subgroups:

- Sociodemographics
  - Age
  - Education
  - Ethnicity
  - Race
  - Gender
- Health status
  - Self-reported general health
  - Self-reported mental health
  - Chronic illness
  - Hospitalization
  - Disability (added in 2002 and 2003)
- Access to health care
  - Insurance status
  - Personal doctor

8.1 Methods

8.1.1 CAHPS® Measures

We began our individual-level analysis by calculating unweighted and weighted frequencies for all categorical variables in the data set. We present and summarize the weighted frequencies and cross-tabulations for key variables on the subgroup analysis report for 2003 (RTI International, 2004b). The analyses presented in this report examine differences across selected data aggregation options for the most-positive CAHPS® ratings and responses (i.e., “10,” “Always,” “Not a Problem,” or “Yes”) that have been case-mix adjusted. ***Table 8.1*** presents the survey questions used to construct each of the CAHPS® indicators and ratings.

To obtain ratings and composites, we used the CAHPS® 3.4 Survey and Reporting Kit macro, which enables us to report on both means and proportions. Furthermore, the ratings and composites at different aggregation levels were case-mix adjusted using models

Table 8.1
CAHPS® performance indicators and ratings

Indicators and ratings	2003 survey questions
Needed Care Composite[a] (with numerical response categories of 1 = A big problem, 2 = A small problem, 3 = Not a problem)	▪ Since you joined Medicare, how much of a problem, if any, was it to get a personal doctor or nurse you are happy with? (Q11) ▪ In the last 6 months, how much of a problem, if any, was it to see a specialist that you needed to see? (Q13) ▪ In the last 6 months, how much of a problem, if any, was it to get the care, tests, or treatment you or a doctor believed necessary? (Q26) ▪ In the last 6 months, how much of a problem, if any, were delays in health care while you waited for approval from Medicare? (Q28)
Good Communication Composite[a] (with numerical response categories of 1 = Never, 2 = Sometimes, 3 = Usually, 4 = Always)	▪ In the last 6 months, how often did doctors or other health providers listen carefully to you? (Q32) ▪ In the last 6 months, how often did doctors or other health providers explain things in a way you could understand? (Q33) ▪ In the last 6 months, how often did doctors or other health providers show respect for what you had to say? (Q34) ▪ In the last 6 months, how often did doctors or other health providers spend enough time with you? (Q35)
Care Quickly Composite[a] (with numerical response categories of 1 = Never, 2 = Sometimes, 3 = Usually, 4 = Always)	▪ In the last 6 months, when you called during regular office hours, how often did you get the help or advice you needed? (Q18) ▪ In the last 6 months, not counting the times you needed healthcare right away, how often did you get an appointment for health care as soon as you wanted? (Q22) ▪ In the last 6 months, when you needed care right away for an illness, injury, or condition, how often did you get care as soon as you wanted? (Q20) ▪ In the last 6 months, how often were you taken to the exam room within 15 minutes of your appointment? (Q29)
Respectful Treatment Composite (with numerical response categories of 1 = Never, 2 = Sometimes, 3 = Usually, 4 = Always)	▪ In the last 6 months, how often did office staff at a doctor's office or clinic treat you with courtesy and respect? (Q30) ▪ In the last 6 months, how often were office staff at a doctor's office or clinic as helpful as you thought they should be? (Q31)
Medicare Customer Service Composite (with numerical response categories of 1 = A big problem, 2 = A small problem, 3 = Not a problem)	▪ In the last 6 months, how much of a problem, if any, did you have with paperwork for Medicare? (Q54) ▪ In the last 6 months, how much of a problem, if any, was it to find or understand this information? (Medicare information in written materials or on the Internet) (Q50) ▪ In the last 6 months, how much of a problem, if any, was it to get the help you needed when you called Medicare customer service? (Q52)

(continued)

Table 8.1
(continued)

Indicators and ratings	2003 survey questions
Rate Medicare (with option to rate 0 [worst health plan possible] to 10 [best personal health plan possible])	Using any number from 0 to 10 where 0 is the worst health plan possible and 10 is the best health plan possible, what number would you use to rate Medicare? (Q55)
Rate Health Care (with option to rate 0 [worst health care possible] to 10 [best personal health care possible])	Using any number from 0 to 10 where 0 is the worst health care possible and 10 is the best health care possible, what number would you use to rate your health care in the last 6 months? (Q36)
Rate Personal Doctor (with option to rate 0 [worst doctor or nurse possible] to 10 [best personal doctor or nurse possible])	Using any number from 0 to 10 where 0 is the worst personal doctor or nurse possible and 10 is the best personal doctor or nurse possible, what number would you use to rate your personal doctor or nurse? (Q6)
Rate Specialist (with option to rate 0 [worst specialist possible] to 10 [best personal specialist possible])	Using any number from 0 to 10 where 0 is the worst personal specialist possible and 10 is the best specialist possible, what number would you use to rate your specialist? (Q15)

[a] Composites featured on the Medicare Personal Plan Finder web page available to Medicare beneficiaries on the Medicare web site (www.medicare.gov/MPPF/DefaultVersion/home.asp).

developed for the within-MFFS comparisons. (The development of the specific models is discussed in the 2003 case-mix report (RTI International, 2004a). The CAHPS® 3.4 software allows specification of the level of aggregation, and we aggregated to the geounit, state, CMS region, and the nation as a whole. We then specified stratification variables to compare variation among subsets of Medicare beneficiaries. The stratification variables included age, gender, education, race, ethnicity, self-perceived general health status, self-perceived mental health status, chronic illness, overnight hospitalization, personal doctor or nurse, disability, and insurance.

8.1.2 MFFS Enrollee Characteristics

The subgroups that were used for stratification across the selected CAHPS® indicators are shown in ***Table 8.2***, with data sources indicated.

When we stratified by a variable that is in the case-mix model, we removed that particular variable from the case-mix model before running the macro. For example, when we stratified by the variable “age,” we removed age from the case-mix model. We followed this same procedure for each stratification variable that was in the case-mix model.

The analyses described in this section do not focus on statistical significance because with such a large sample, even fractional percentage point differences may prove to be statistically significant but not substantive. Therefore, our focus is on substantive percentage point differences for the various indicators.

Table 8.2
Subgroups used for stratification across selected CAHPS® indicators

Variable/construct	Categories	Data source
Age	18–45 46–64 65–69 70–74 75–79 80+	Self-reported. If missing/not reported, then determination was made from the Enrollment Database (EDB) file.
Education	8th grade or less Some high school, but did not graduate High school graduate or GED Some college or 2-year degree 4-year college degree More than 4-year college degree	Self-reported.
Ethnicity	Hispanic or Latino Medicare beneficiaries Not Hispanic or Latino Medicare beneficiaries	Self-reported. If missing/not reported, then determination was made via an algorithm that used additional information from the survey data, from the EDB file, and the Spanish Surname List. A detailed description of the algorithm employed can be found in the full 2002 subgroup analysis report (Brody et al., 2003).
Race	Medicare beneficiaries of White race Medicare beneficiaries of Black race Medicare beneficiaries of other race	Self-reported. If missing/not reported, then determination was made via an algorithm that used information from the EDB file. A detailed description of the algorithm employed can be found in the full 2002 subgroup analysis report (Brody et al., 2003).
Gender	Male Female	Self-reported. If missing/not reported, then determination was made from the EDB file.
Health status	Medicare beneficiaries who reported their physical health status as excellent, very good, good, fair, or poor Medicare beneficiaries who reported their mental health status as excellent, very good, good, fair, or poor Medicare beneficiaries who reported chronic illness Medicare beneficiaries who reported being hospitalized overnight or longer in the last 12 months	Self-reported.

(continued)

Table 8.2
(continued)

Variable/construct	Categories	Data source
Insurance status	Medicare beneficiaries who are also enrolled in Medicaid (i.e., the dually eligible)	EDB file. If EDB file indicates not on Medicaid, then code remaining insurance categories as below.
	Medicare beneficiaries who have additional health care insurance coverage without a prescription drug benefit	Self-reported. If response to Question 1[a] is "Yes" and response to Question 2[b] is not "Medicaid" and response to Question 47[c] is either "No" or missing, then assign to this category.
	Medicare beneficiaries who have additional health care insurance coverage including a prescription drug benefit	Self-reported. If response to Question 47 is "Yes," then assign to this category.
	Medicare beneficiaries who have no additional insurance	Self-reported. If response to Question 1 is "No" and response to Question 47 is "No" or missing; OR if response to Question 1 is "Yes" and only response to Question 2 is "Medicaid"; OR if response to Question 1 is missing and response to Question 2 is "I don't have health insurance other than Medicare" and response to Question 47 is "No" or missing; OR if did not respond to Question 1 and Question 2 and response to Question 47 is "No," then assign to this category.
	Medicare beneficiaries who did not respond to this question (missing)	Self-reported. If no response to Question 1, 2, or 47, then assign to this category.
Personal doctor	Medicare beneficiaries who reported having a personal doctor or nurse	Self-reported.
Disability		EDB file. This indicator was created from the beneficiary's history of entitlement reason codes (BENE_ENTLMT_RSN_CD). If any of the last 10 entitlement codes indicated disability (values of 1 or 3) then the beneficiary was assigned as disabled.

[a] Question 1 text: "Do you have any other insurance that pays at least some of the cost of your health care?"

[b] Question 2 text: "Please mark the box below for each type of health insurance that you have."

[c] Question 47 text: "Not including Medicare, do you have any other health insurance that pays at least some of the costs of medicines prescribed by doctors and other health providers?"

NOTE: For discussion of the case-mix model, see Section 6.

8.1.3 Case-Mix Adjustment

CMS is required by the 1997 Balanced Budget Act (BBA) to provide beneficiaries with information that will enable them to choose between Medicare plan options. This requirement, also reiterated in the Medicare Modernization Act of 2003, necessitates the construction of CAHPS® ratings and composites that can be compared across managed care plans and between MA and MFFS options. The implication for the construction of the composites from the MFFS survey is that they be created in as like a manner as possible to those from the MA survey.

Because CMS intends to provide quality information to support Medicare beneficiaries' choice of Medicare health plan options, it is essential that differences between the composition of Medicare beneficiaries in MFFS and in MA be adequately adjusted for when data are reported. For MFFS, this adjustment must be made on the reporting-unit level and, in order to allow like comparisons, must be comparable in rigor and scope to the adjustment made on the MA sample. Case-mix adjusted consumer ratings can provide more valid health plan comparisons than can unadjusted ratings by controlling for factors related to systematic response biases. Adjusted data are therefore potentially more appropriate for comparing the quality of care delivered.

Case-mix adjustment for systematic bias is useful when comparing assessments of different plans or regions if members of a particular demographic group that is more or less inclined than others to assign poor ratings to bad care are disproportionately enrolled in a particular plan or, as in the case of within-MFFS comparisons, these members reside in a particular geographic area. In many markets, MFFS beneficiaries tend to be older and frailer than MA beneficiaries. To present fair comparisons, the influence of plan composition must be accounted for in the reporting statistic. A similar argument can be made for comparison of ratings and composites for different geographic units within the MFFS population. For these reasons, all ratings and composites used to compare MFFS and MA, or regions within the MFFS population, are case-mix adjusted.

8.2 Characteristics of 2003 CAHPS® MFFS Population

We briefly summarize sociodemographic characteristics, health status, and insurance status of all survey respondents weighted to the U.S. population of MFFS beneficiaries below (see ***Table 8.3***).

Table 8.3
Demographic, health status, and insurance: Frequencies weighted to the CAHPS® MFFS population (n = 120,974)

Descriptive variable	Percent	Descriptive variable	Percent
Sociodemographic characteristics		**Health status (self-reported)**	
Age		***Self-perceived general health status***	
18-44 years	2	Excellent	6
45-64	10	Very good	20
65-69	22	Good	33
70-74	21	Fair	29
75-79	19	Poor	12
80 years or older	25	***Self-perceived mental health status***	
Gender		Excellent	23
Male	43	Very good	29
Female	57	Good	30
Education		Fair	14
8th grade or less	14	Poor	4
Some high school, but did not graduate	14	***Had a physical/medical condition that lasted at least 3 months***	72
High school graduate or GED	35	***Hospitalized overnight or longer in the last 12 months***	23
Some college or 2-year degree	20	***Disabled***	19
4-year college degree	8	**Access to care (self-reported)**	
More than 4-year college degree	9	***Had a personal doctor or nurse***	88
Race		***Insurance—plans in addition to Medicare***	
White	89	Additional insurance with prescription drug coverage	47
Black	8	Additional insurance without prescription drug coverage	26
Other	3	No additional insurance	13
Ethnicity		Dual eligible/Medicaid	14
Hispanic or Latino	5	Missing	< 1
Not Hispanic or Latino	95		

NOTE: Due to rounding, percentages may sum to slightly less or more than 100 percent.

8.2.1 Sociodemographics

The majority of beneficiaries (62 percent) were 65 to 79 years of age, and one out of four beneficiaries was aged 80 or older. Only 12 percent of the population was under 65 years of age. More than half (57 percent) were female. More than one-quarter (28 percent) of respondents had less than a high school education, approximately one-third were high school graduates, and the remaining 37 percent had more than a high school education. Most beneficiaries (89 percent) were White, 8 percent were Black, 5 percent were Hispanic or Latino,[1] and 3 percent were of other/unknown race.

8.2.2 Health Status

Approximately one-quarter of respondents reported excellent or very good health, 33 percent reported good health, and about 40 percent were in fair or poor health. More than half (52 percent) reported excellent or very good mental health, more than one-quarter (30 percent) were in good mental health, and around 18 percent reported being in fair or poor mental health. Almost three of four beneficiaries (72 percent) reported having a physical or medical condition that lasted at least 3 months, and almost one-quarter (23 percent) reported having been hospitalized overnight in the past year. Approximately 19 percent of the beneficiaries were disabled, as indicated by information provided in the CMS Enrollment Database (EDB). Please note that the category “disabled” includes only beneficiaries whose initial eligibility for Medicare was based on disability.

8.2.3 Access to Care

An overwhelming majority (88 percent) of beneficiaries reported having a personal doctor. Of the total respondents, 47 percent reported having additional insurance with prescription drug coverage, and 26 percent reported having health insurance in addition to Medicare but no supplemental coverage for prescription drugs. Further, 14 percent were dually eligible for both Medicare and Medicaid (as indicated in the EDB file), and 13 percent reported having no insurance in addition to Medicare. Detailed information about supplemental insurance coverage can be found in ***Table 8.4***.

8.3 Key Findings

8.3.1 Findings From 2000 Through 2003

Overall, on a national level, little change has occurred in the CAHPS® composites over the past 4 years. During the 4-year period (2000 through 2003), between 84 percent and 89 percent of MFFS beneficiaries reported that they always received needed care (see ***Figure 8.1***). The percentage of most-positive responses was slightly lower in 2003 (85 percent) and 2002 (84 percent) than in 2001 (89 percent) and 2000 (87 percent), but remains quite high.

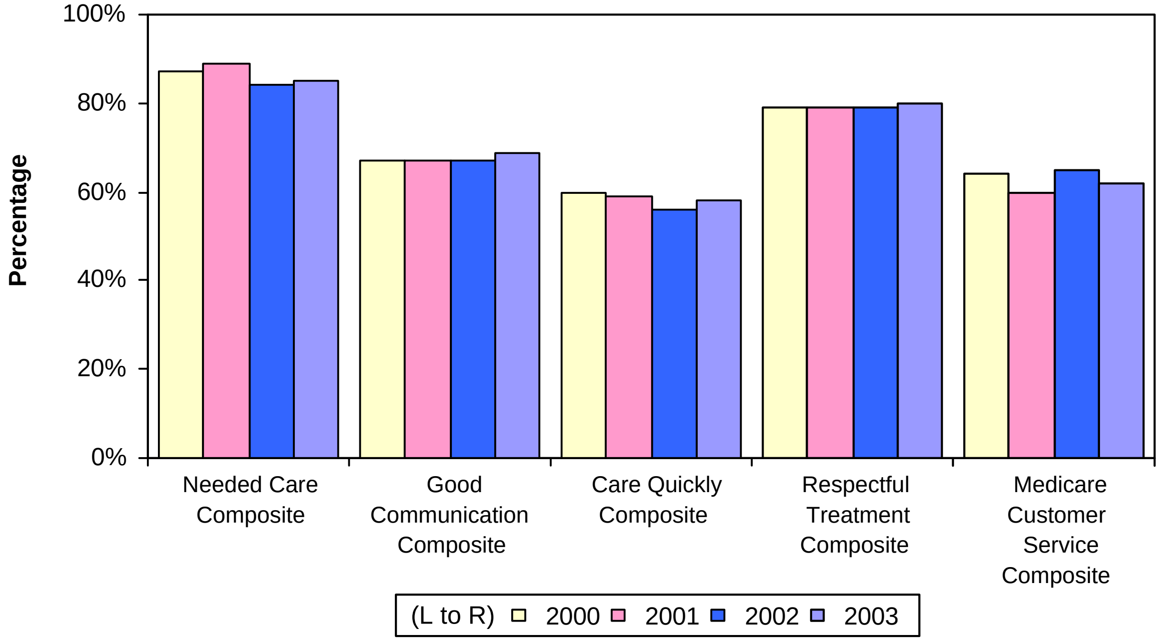
---

[1] Indicators of Hispanic ethnicity and race are from two separate questions (Questions 84 and 85, respectively, from the 2003 CAHPS® Medicare Satisfaction Survey) and are reported on each group regardless of answers to the other question.

Table 8.4
Insurance in addition to Medicare

Categories	Percent
Medigap	9.2
Employer, union, or retiree health coverage	17.7
Veteran's benefits (VA benefits)	1.1
Military retiree benefits (TRICARE)	3.1
Other	25.9
Dually eligible/Medicaid	14.0
Reporting more than one supplemental plan	15.6
No additional insurance	13.0
Missing	0.4

Figure 8.1
CAHPS® composites: Percent of beneficiaries providing most-positive responses (2000–2003)



The proportion of beneficiaries who reported always getting care quickly during the 4-year period was lower; however, more than half of beneficiaries (58 percent) reported no problems getting care quickly in 2003. The small differences in results of the Care Quickly composite between the 2003, 2002, and 2001 surveys may be the result of changes in the wording of questions that make up this composite indicator (see ***Table 8.5***).

Most-positive responses for both the Respectful Treatment composite and the Good Communication composite increased slightly in 2003, after being consistent at 79 percent and 67 percent, respectively, during the first 3 years of the CAHPS® MFFS survey. Throughout the 4-year period, approximately four out of five beneficiaries reported optimal experiences with being treated respectfully by providers and their staff, whereas only two-thirds of beneficiaries provided most-positive responses to the Good Communication composite.

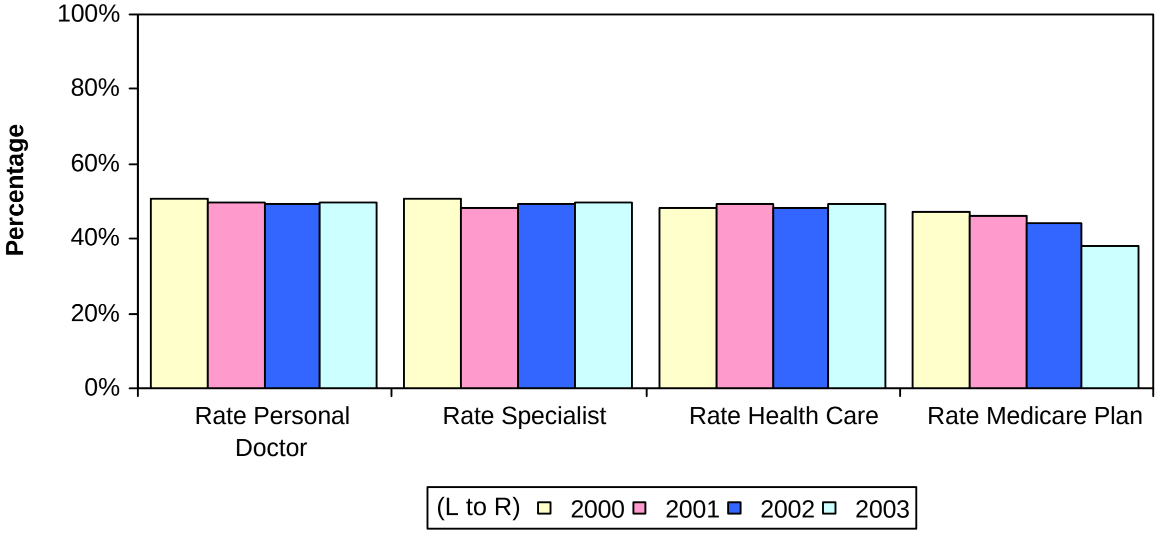
Of the composites, the Medicare Customer Service composite fluctuated the most during the 4-year survey period. For example, 64 percent of beneficiaries reported most-positive experiences in 2000, down to 60 percent in 2001, up to 65 percent in 2002, and back down again to 62 percent in 2003.

Table 8.5
Question wording changes for Care Quickly Composite items (2000–2003)

2000–2001	2002	2003
In the last 6 months, when you called during regular office hours, how often did you get the help or advice you needed?	In the last 6 months, when you called during regular office hours, how often did you get the help or advice you needed?	In the last 6 months, when you called during regular office hours, how often did you get the help or advice you needed?
In the last 6 months, how often did you get an appointment for regular or routine care as soon as you wanted?	In the last 6 months, how often did you get an appointment for health care as soon as you wanted?	In the last 6 months, not counting the times you needed health care right away, how often did you get an appointment for health care as soon as you wanted?
In the last 6 months, when you needed care right away for an illness or injury, how often did you get care as soon as you wanted?	In the last 6 months, when you needed care right away for an illness, injury or condition, how often did you get care as soon as you wanted?	In the last 6 months, when you needed care right away for an illness, injury, or condition, how often did you get care as soon as you wanted?
In the last 6 months, how often did you wait in the doctor’s office or clinic more than 15 minutes past your appointment time to see the person you went to see?	In the last 6 months, how often did you see the person you came to see within 15 minutes of your appointment time?	In the last 6 months, how often were you taken to the exam room within 15 minutes of your appointment?

Approximately half of beneficiaries rated health care, their specialist, and their personal doctor a “10” during the 4-year survey period (see ***Figure 8.2***). Ratings of the Medicare health plan were slightly lower overall and decreased over the 4-year period. The most noteworthy change in the ratings for 2003 was a 6 percentage point decrease in beneficiaries’ rating the Medicare health plan a “10.” Specifically, only 38 percent of beneficiaries rated Medicare a “10” in 2003, a large deviation from the gradual downward trend in the previous 3 years—47 percent in 2000, 46 percent in 2001, and 44 percent in 2002. The cause of the observed downward trend is not clear; however, increased news coverage of Medicare, especially related to Medicare reforms, may have made beneficiaries scrutinize Medicare more closely and thus contributed to the substantial decrease observed between the 2002 and 2003 surveys. In addition, there have been slight changes in the wording of the survey question corresponding to beneficiary plan ratings (see ***Table 8.6***) that may have also contributed to this trend. Variation in the CAHPS® composites can best be observed among states or markets within states. Data for CAHPS® results by CMS region, state, and substate areas are available from the 2003 subgroup analysis report (RTI, 2004b).

Figure 8.2
CAHPS® ratings: Percent of beneficiaries providing ratings of “10” (2000–2003)

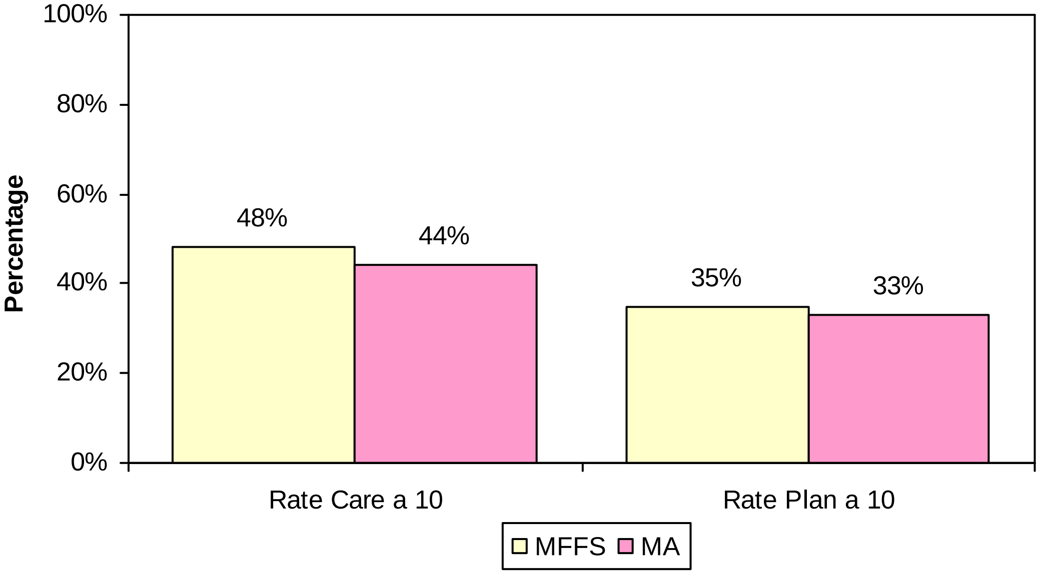


CAHPS® indicator results among the MFFS beneficiary population were compared with beneficiaries participating in the MA program. When making these comparisons, we included only MFFS results from areas where beneficiaries have a choice of MFFS and MA health plans, rather than results from all beneficiaries enrolled in the MFFS program, resulting in slight differences in observed rates. For example, in 2003, 38 percent of overall beneficiaries rated the Medicare health plan a “10,” whereas this value is 35 percent for the MFFS population calculated only for areas where beneficiaries have a choice of plans (see ***Figure 8.3***).

Table 8.6
CAHPS® survey question to assess beneficiary satisfaction with Medicare plan (2000–2003)

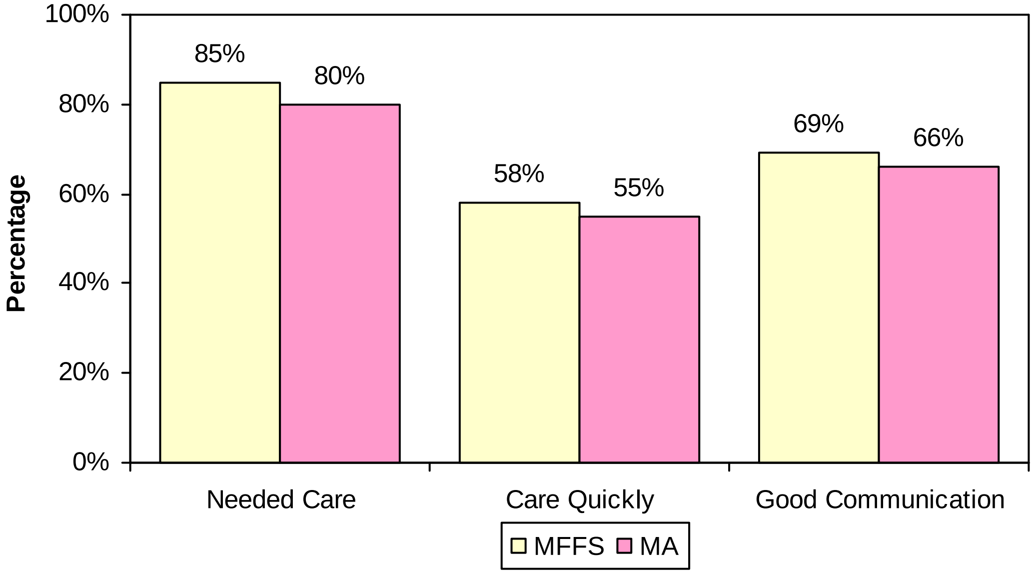
- 2000 & 2001—How would you rate all your experience with Medicare? Use any number from 0 to 10 where 0 is the worst health plan possible, and 10 is the best health plan possible.
- 2002—How would you rate all your experience with Medicare? Using any number from 0 to 10 where 0 is the worst health plan possible, and 10 is the best health plan possible, what number would you use to rate your health plan?
- 2003—Using any number from 0 to 10 where 0 is the worst health plan possible, and 10 is the best health plan possible, what number would you use to rate Medicare?

Figure 8.3
Comparison between MA and MFFS for CAHPS® ratings: Percent of beneficiaries providing ratings of "10" (2003)



In this case-mix adjusted comparison, patient experiences with getting care quickly and good communication were very similar for the MFFS and MA beneficiary populations. In contrast, a higher percentage of MFFS beneficiaries provided most-positive responses to the Needed Care composite during the 4 years compared with their peers enrolled in MA. For example, in 2003, 85 percent of MFFS beneficiaries reported a most-positive response for getting needed care, compared with only 80 percent of MA beneficiaries (see ***Figure 8.4***). MFFS beneficiaries were also slightly more likely than MA beneficiaries to provide the best ratings for their health care and the Medicare health plan. For example, in 2003, 48 percent of MFFS beneficiaries rated their health care a "10" in contrast to only 44 percent of MA beneficiaries (see ***Figure 8.3***).

Figure 8.4
Comparison between MA and MFFS for CAHPS® composites: Percent of beneficiaries providing most-positive responses (2003)



8.3.2 Sociodemographic Characteristics

8.3.2.1 2003 Highlights

For all CAHPS® composites and the rating of the Medicare plan, beneficiaries between 18 and 45 years of age were less likely to provide most-positive responses than all other age groups. For the Needed Care and Respectful Treatment composites and ratings of the Medicare plan and overall health care in 2003, there was at least a 14 percentage point difference between the proportion of most-positive responses reported by beneficiaries 80 years of age or older and those in the youngest age group, 18 to 45 years. For example, 87 percent of beneficiaries aged 80 or older provided the most-positive response for the Needed Care composite, compared with only 70 percent of beneficiaries aged 18 to 45.

Black beneficiaries were generally more likely than White beneficiaries or beneficiaries of other races to provide most-positive responses to CAHPS® indicators in 2003. The only exception to this finding was for the Needed Care composite, where 84 percent of White beneficiaries provided the most-positive response compared with 81 percent of Black beneficiaries. Hispanic beneficiaries were more likely to provide most-positive responses to the CAHPS® ratings during 2003. We found considerable differences in CAHPS® ratings between Hispanic and non-Hispanic beneficiaries, with 6 to 17 percentage points separating the CAHPS® scores reported by Hispanic compared with non-Hispanic beneficiaries. For example, in 2003, 54 percent of Hispanic beneficiaries rated Medicare a “10” compared with 38 percent of non-Hispanic beneficiaries.

With the exception of the Needed Care composite, there was a consistent tendency for lower ratings and composites to be associated with increasing beneficiary education. There was a greater than 10 percentage point difference between the proportion of most educated and least educated beneficiaries providing most-positive responses. For example, 33 percent of beneficiaries with more than a college degree rated their health care a “10” in 2003, whereas almost twice as many beneficiaries (60 percent) with an eighth grade education or less rated their health care a “10.” Similar patterns of results by education were observed during prior survey years, 2000 through 2002.

8.3.2.2 Findings From 2000 Through 2003

A similar proportion of male and female MFFS beneficiaries provided most-positive responses to the CAHPS® composite indicators, whereas 3 percent to 5 percent more women provided best possible responses to the four CAHPS® ratings during the 4-year period. There was a consistent age effect, with younger beneficiaries less likely to report most-positive experiences than older beneficiaries during all 4 years. CAHPS® indicator scores similarly varied by education during the 4-year period; less educated beneficiaries were consistently more likely to report most-positive experiences than more educated beneficiaries. Blacks were generally more likely to report most-positive ratings of satisfaction with care than were Whites and beneficiaries of other races. For example, slightly over half of Blacks rated their overall health care a “10” across all 4 years compared with only 46 percent to 47 percent of Whites during 2000 through 2003. Hispanics rated Medicare and their overall health care slightly higher than non-Hispanics, but there was generally little to no difference in their composite scores during the 4-year period.

8.3.3 Health Status

8.3.3.1 2003 Highlights

For most indicators, at least 10 percent more beneficiaries in excellent health provided most-positive responses than those in poor health. In 2003, there was a 15 percentage point difference between the proportion of the healthiest beneficiaries and sickest beneficiaries providing most-positive responses for the Needed Care composite (80 percent vs. 65 percent for general health perception; 76 percent vs. 61 percent for mental health perception).

The self-reported presence of a chronic illness had no impact on several composites but a minor impact on the Needed Care and Medicare Customer Service composites, both of which had fewer most-positive responses from beneficiaries with a chronic illness than from beneficiaries with no chronic illness. A smaller proportion of beneficiaries who had a chronic illness provided most-positive ratings of “10” for all four indicators reflecting ratings of beneficiary satisfaction, compared with beneficiaries who did not report having a chronic illness.

One notable difference in 2003 was rating Medicare; only 36 percent of chronically ill beneficiaries rated Medicare a “10” compared with 43 percent of beneficiaries with no chronic illness. Beneficiaries who had been hospitalized overnight at least once during the past year were more likely to rate their specialists a “10”; however, hospitalization during the past year had little effect on other CAHPS® scores during the 4 years. In 2003, disability seemed to have no effect on the five CAHPS® composites; however, disabled beneficiaries gave a slightly higher percentage of "10" ratings for their personal doctor, specialist, health care, and the Medicare plan.

8.3.3.2 Findings From 2000 Through 2003

Across all indicators, we found a strong association between health status and CAHPS® scores; beneficiaries reporting better general and mental health status were more likely to provide most-positive responses for each CAHPS® indicator during the 4 years.

8.3.4 Access to Care

8.3.4.1 2003 Highlights

Beneficiaries who were dually eligible for Medicare and Medicaid reported greater satisfaction and better experiences than those who reported having additional insurance and prescription drug coverage; this was the case for the Good Communication and Medicare Customer Service composites and all ratings in 2003.[1] In particular, over half of dually eligible beneficiaries rated Medicare a "10" compared with 35 percent of beneficiaries who had additional insurance and prescription drug coverage.

Beneficiaries who reported having a personal doctor were more likely to provide most-positive responses, generally by at least 10 percentage points, for the five CAHPS® composites and their ratings of health care and specialists in 2003. For example, 49 percent of beneficiaries who had a personal doctor rated their health care a "10" compared with 38 percent of beneficiaries who did not have a personal doctor. Ratings of the Medicare plan were not affected by whether beneficiaries reported having a personal doctor or not.

8.3.4.2 Findings From 2000 Through 2003

Beneficiaries who reported having additional insurance but no prescription drug benefits were less satisfied and reported worse experiences than those with additional insurance and prescription drug benefits during each of the 4 years.

8.3.5 Geographic Variation

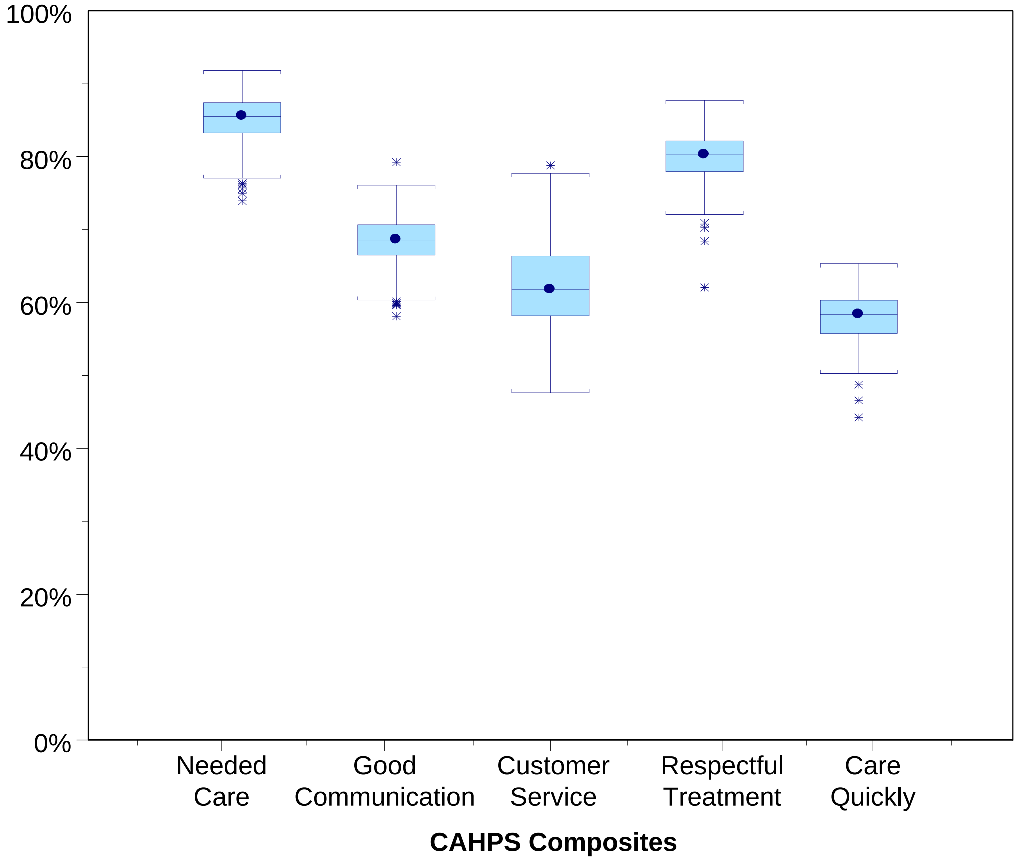
8.3.5.1 2003 Highlights

Figures 8.5 and 8.6 are box and whisker plots that display the variation in percentages of most-positive responses and best possible ratings across CAHPS® indicators by geounit. The shaded boxes represent the interquartile range, and the dot within the box represents the median of the distribution. The pair of lines closest to the shaded box approximates the range of observed values, while the asterisks further away from the shaded area represent outliers, values that are more than 1.5 times above or below the interquartile range.

---

[1] Dual eligibility for Medicaid was determined using the EDB file, whereas information about additional insurance and prescription drug coverage was obtained via beneficiary self-report collected using the CAHPS® survey.

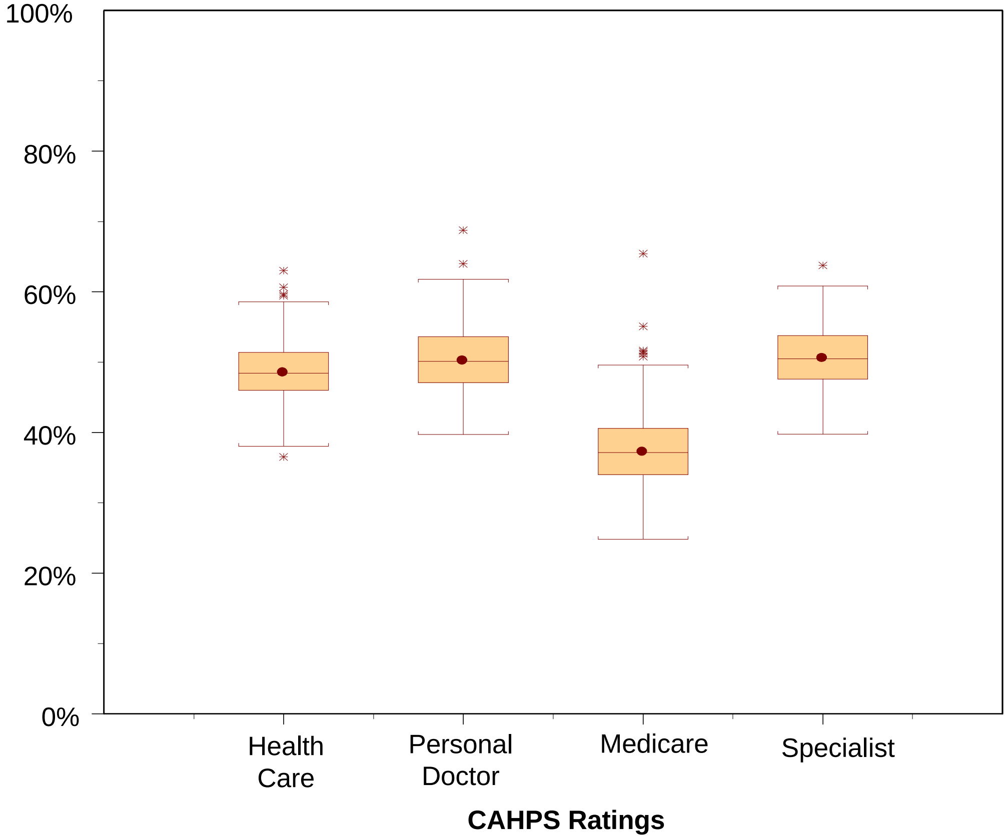
Figure 8.5
Variation in percentage of most-positive responses across CAHPS® composites by geounit



Median percentages of most-positive responses varied across the different composites, ranging from a high of approximately 85 percent for the Needed Care composite to a low of approximately 58 percent for the Care Quickly composite. In contrast, the median percentages of best possible ratings among geounits for the ratings were similar. The median percentages of most-positive ratings for three of the ratings were approximately 50 percent, while this figure was approximately 38 percent for rating Medicare.

We noted some small, regional differences in CAHPS® scores. Enrollees residing in Maine and to a lesser extent other New England states, such as Vermont, consistently rated their Medicare experience and satisfaction higher than the national average. On the other hand, beneficiaries residing in the West—particularly Arizona, New Mexico, and Nevada—rated their Medicare experience and satisfaction lower than the national average. Overall, a higher than average proportion of MFFS beneficiaries living in the South provided most-positive responses across all composites compared with people in the West, who provided lower than average ratings on most CAHPS® composites. The one exception to the South providing higher most-positive responses was Florida, which generally reported lower percentages of most-positive responses than the national average.

Figure 8.6
Variation in percentage of best possible CAHPS® ratings by geounit



8.3.5.2 Findings From 2000 Through 2003

Beneficiaries residing in the Midwest generally rated their Medicare experience and satisfaction consistent with the national average. Beneficiaries residing in Puerto Rico provided considerably higher ratings than average for all four CAHPS® ratings. This trend was also observed for the Needed Care and Customer Service composites during 2001 and 2002.

8.3.6 Conclusion

Overall, CAHPS® scores seem to be relatively stable during this 4-year period. It was surprising to see the decrease in the rating of Medicare, which (as was discussed) may be explained by the prominence of the Medicare program and benefits in the news. However, the meaning of that decrease is not clear because it is not associated with other aspects of the Medicare program as measured by other ratings and composites where we do not see similar decreases.

8.4 Supplemental Analysis Using Claims Data: Relationship Between CAHPS® Scores and Hospital Encounters for Ambulatory Care Sensitive Conditions

8.4.1 Objective

During the past year, we conducted a study to examine whether the incidence of hospital encounters for ambulatory care sensitive conditions (ACSCs) is associated with self-reports of poor access to care or dissatisfaction with health care services among MFFS beneficiaries. Specifically, we examined whether beneficiaries had a hospitalization, observation stay, or emergency room visit for at least one of the following conditions during the 12 months preceding administration of the CAHPS® survey:

- Cellulitis
- Congestive heart failure
- Dehydration
- Bacterial pneumonia
- Septicemia
- Ischemic stroke
- Urinary tract infection
- Asthma
- Chronic obstructive pulmonary disease
- Acute diabetic events among Medicare beneficiaries with diabetes
- Lower limb peripheral vascular disease among Medicare beneficiaries with diabetes

8.4.2 Methods

We estimated weighted and case-mix adjusted means for the nine CAHPS® indicators using the CAHPS® 3.4 Survey and Reporting Kit macro. Case-mix adjustment variables included:

- Education
- Proxy assistance
- Proxy respondent
- Self-reported health

- Self-reported mental health
- Claims-based measure of health status—hierarchical condition categories (HCC) score[1]

Further, we used two-sample t-tests for differences in means for all pairwise comparisons ($p<0.05$ with Bonferroni corrections).

8.4.3 Results and Implications

Based on 2000 data, our results show that 6.1 percent of MFFS beneficiaries experienced at least one inpatient admission, observation stay, or emergency room visit for an ACSC during the 12 months preceding administration of the CAHPS® survey. Although beneficiaries with an ACSC were modestly more likely to report problems getting needed care, and this difference was statistically significant, the mean difference in this CAHPS® composite was very small (***Table 8.7***). Overall, we found virtually no differences in mean CAHPS® scores between beneficiaries who did or did not experience an ACSC encounter. Mean satisfaction scores as measured by the CAHPS® in the MFFS population are very high and do not appear to vary by ACSC hospitalization. This finding suggests that CAHPS® scores reflect a domain of health care quality distinct from that represented by hospitalizations for ACSCs.

Table 8.7
Comparison of mean CAHPS® scores for MFFS beneficiaries by ACSC indicator

Indicator	ACSC	No ACSC	p-value
Experience with care			
Needed care	2.80	2.81	p<.01
Care quickly	3.41	3.40	ns
Good communication	3.57	3.57	ns
Satisfaction with care			
Rate Medicare	8.68	8.70	ns
Rate care	8.79	8.81	ns
Rate specialist	8.83	8.85	ns
Rate personal doctor	8.85	8.86	ns
Customer service			
Respectful treatment	3.73	3.73	ns
Medicare customer service	2.52	2.51	ns

NOTE: ACSC = ambulatory care sensitive condition; ns = difference is not significant.

---

[1] The hierarchical condition categories (HCC) are constructed as a result of aggregating International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM), diagnostic codes into 184 clinically meaningful diagnostic categories organized into multiple body system or disease type hierarchies that distinguish disease severity of related diagnoses. Higher scores indicate greater disease severity.

Because this analysis used a composite measure of different ACSCs, we cannot know if a relationship exists between CAHPS® scores and the incidence of hospitalizations for specific conditions. Therefore, we recommend that additional analyses be conducted to examine whether there is a relationship between CAHPS® measures and ACSCs for specific conditions.

8.5 Supplemental Analysis Using Claims Data: Relationship Between CAHPS® Scores and Claims-Based Quality Measure for Beneficiaries With Diabetes

8.5.1 Objectives

During this survey year, using the individual as the unit of analysis, we examined the relationship between measures of patient satisfaction and claims-based clinical measures of (a) the quality of preventive care delivered to MFFS beneficiaries with diabetes and (b) complications from diabetes. We hypothesized that CAHPS® measures of patient experience (i.e., composite measures of getting needed care, getting care quickly, and good communication with providers) were more likely to be associated with clinical measures of quality than attitudes represented by CAHPS® ratings and patient experiences with Medicare and provider customer service.

We focused this analysis on MFFS beneficiaries with diabetes. Diabetes is a common and widespread disease in the United States, and it is approaching epidemic proportions. Diabetes is not curable; however, a substantial body of research has demonstrated that receipt of appropriate medical care and patient self-care practices can delay or prevent crippling complications from diabetes, such as blindness, lower extremity disease, and cardiovascular complications. Because care guidelines for management of diabetes are clearly defined, we chose this chronic disease to evaluate the relationship between patient satisfaction and clinical measures of quality among the MFFS population.

8.5.2 Methods

Study Sample—To conduct this study, we used data from the 2000 CAHPS® MFFS survey. The sample of beneficiaries was drawn from a sampling frame constructed from the CMS Enrollment Database (EDB). The frame comprised over 30 million beneficiaries who were enrolled in MFFS for at least the prior 6 months and resided in any state in the United States, the District of Columbia, or Puerto Rico.[1] The number of respondents for the 2000 survey was 103,551, yielding a response rate of 64 percent. These data were merged with 1999 through 2001 Medicare claims data derived from Medicare Parts A & B claims files using a unique identifier assigned to each beneficiary.

For this analysis, we included only those 2000 MFFS respondents with a diabetes diagnosis as evidenced by ICD-9 codes on the data file. To construct a 12-month interval of

---

[1] Additional information about the sampling strategy can be found in the final report for the 2000 MFFS survey (RTI and RAND, 2001).

claims for each respondent prior to the survey date,[1] we drew from Medicare claims data representing health care utilization from 1999 through 2001. Persons were classified as having diabetes if they met the following criteria based on research conducted by Hebert et al. (1999):

- A claim from inpatient, outpatient, skilled nursing facility, or home health agency with a diagnosis of diabetes (250.xx, 357.2, 362.01, 362.02, or 366.41), excluding women diagnosed during pregnancy.

  AND one of the following additional criteria:

  - One face-to-face acute encounter for diabetes in a hospital inpatient or emergency room setting

    OR

  - Two face-to-face nonacute encounters at least 7 days apart

This algorithm yielded 16,532 diabetic beneficiaries, approximately 16 percent of the MFFS survey respondents in 2000.

Variable Construction—

Claims-Based Health Status Measure. We used HCCs to create a claims-based health status measure. The HCCs aggregate ICD-9-CM diagnostic codes into 184 clinically meaningful diagnostic categories and are organized into multiple body system or disease type hierarchies that distinguish disease severity of related diagnoses (Pope et al., 2000). For this analysis, the health status measure was equal to the beneficiary's HCC risk score divided by the national mean of HCC of all Medicare beneficiaries in 2000, resulting in a range of 0.13 to 12.29 and a mean of 1.53, with higher scores indicating greater disease severity.

Clinical Quality of Care Indicators. Using ICD-9 and current procedural terminology (CPT) codes from the Medicare claims data representing the year preceding and the year following the survey date of each respondent, we created three measures related to receipt of recommended services to manage and prevent the development of complications from diabetes and one measure representing the presence of complications from diabetes. If the survey date for a respondent was not available, the midpoint of the survey period was assigned as the reference date for that individual. Each of these measures is described below.[2]

Prevention Measures. Diabetes management that effectively prevents or delays complications from the disease involves both healthy lifestyle choices related to physical

---

[1] The 2000 implementation of the CAHPS® survey among the MFFS population was fielded during September 2000 through January 2001.

[2] Additional information about the specification of each indicator can be found in Chen et al. (2003).

activity and diet, as well as receiving appropriate preventive care services from health care providers (American Association of Clinical Endocrinologists and American College of Endocrinology, 2002; National Institute of Diabetes and Digestive and Kidney Disease, 2003). Using claims data, we created three measures reflecting whether beneficiaries received a subset of recommended care to prevent complications from diabetes:

- "Physiological test" is an indicator of whether beneficiaries received at least one of the following physiological tests during the year preceding the survey: hemoglobin A1c test, lipid profile, or test for microalbuminuria.
- Self-blood glucose monitoring and participation in diabetes education were grouped into a single indicator, "Self-care."
- "Eye exam" is an indicator of whether beneficiaries received eye exams.

Disease Progression. In addition to measuring whether beneficiaries received appropriate preventive care, we measured the prevalence of complications from diabetes as an indicator of health care quality. We hypothesized that beneficiaries with complications from diabetes may not have received optimal clinical care. Initially, we created variables representing five different disease progression measures: cardiovascular disease, lower extremity disease, neurological complications or stroke, diabetes complication, and diabetic eye disease. However, the prevalence of each of these conditions was relatively low, ranging between 2 percent and 20 percent of the diabetics, so that we combined all of these indicators into one disease progression variable; 34 percent of diabetics had been diagnosed with at least one of the complications of diabetes under study.

Statistical Analysis—We calculated CAHPS® ratings and composites for the population of MFFS beneficiaries with diabetes using the CAHPS® 3.4 Survey and Reporting Kit macro. All analyses were performed in SAS using ordinary least squares (OLS) regression, according to CAHPS® macro guidelines. Mental health perception, general health perception, age, education, proxy status, and HCC risk score divided by the 2000 national mean functioned as covariates. For composite outcomes, all component items within a composite were weighted equally, with mean imputation for missing items within a composite. Standard errors were appropriately adjusted for this imputation.

The CAHPS® macro enabled us to report on both means and proportions, as well as to determine the level of aggregation. In the Results section that follows, we present the weighted and case-mix adjusted proportions among beneficiaries who did and did not receive specific preventive care services and among beneficiaries who acquired complications of diabetes at the national level for the most-positive CAHPS® ratings and responses (i.e., "10," "Always," "Not a Problem," or "Yes"). Variables making up the case-mix model included age, education, self-reported health, self-reported mental health, proxy, and a claims-based health status measure based on the HCC.

We used the two-sample t-test for differences in proportions to make pairwise comparisons. Our sample sizes for all groups of interest exceed the usual rules of thumb for large sample tests; thus, by the central limit theorem, the normal approximation to the binomial distribution was used to make the pairwise comparisons. We held the family-wise error rate at alpha = 0.05 and accounted for the multiple comparisons using the Bonferroni adjustment.

8.5.3 Results

Characteristics of Respondents—In general, persons with diabetes who had already experienced complications of the disease were older and in poorer health than the overall population of persons with diabetes. For example, 24.7 percent of diabetics with complications, in contrast to only 20.2 percent of all persons with diabetes, were older than 80. More than one-fourth of persons with complications reported much worse health than 1 year ago, whereas only 17.6 percent of all persons with diabetes classified their general health in this way. Further, the average HCC index of health status was 2.25 for persons who had complications compared with 1.5 for the general population of beneficiaries with diabetes. Diabetics who received preventive services under study were not substantially different from the overall population of persons with diabetes.

Medicare claims data showed that more than three of four persons with diabetes received at least one of the following physiological tests during the past year: hemoglobin A1c, lipid profile, or microalbumin test. Fifty-eight percent of MFFS beneficiaries with diabetes received a diabetic eye exam during the previous year, and almost two of five beneficiaries either took a diabetes education course or engaged in self-blood glucose monitoring, as evidenced by Medicare claims data.

Relationship Between Clinical and Patient Satisfaction Measures—Overall, there was a relationship between average scores of CAHPS® indicators of patient experience and satisfaction with care and two of the measures of clinical quality: (a) receiving a dilated eye exam during the past year and (b) receiving at least one of three physiological tests recommended for detecting early complications of diabetes, as shown in ***Table 8.8***.

Physiological Tests—More than three of four MFFS beneficiaries (78 percent) received at least one of the following three recommended physiological tests during the previous year: hemoglobin A1c test, microalbumin test, or lipid profile. There was a statistically significant association between this measure of health care quality and all of the CAHPS® indicators with the exception of one indicator, Medicare Customer Service. On average, beneficiaries who had received at least one of the recommended tests provided higher ratings of their experience obtaining needed care, receiving care without delay, and communication with their provider. In addition, these beneficiaries also rated their personal doctor, specialist, the Medicare health plan, and health care in general higher than those who had received none of the recommended tests in the previous year.

Table 8.8
Association of mean CAHPS® scores for MFFS with claims-based indicators of clinical quality

	Eye exam			Physiological tests			Self-care			Disease progression		
Measure	Yes	No		Yes	No		Yes	No		Yes	No	
Experience with care												
Getting needed care	2.83	2.81	*	2.82	2.80	**	2.82	2.82	ns	2.82	2.82	ns
Care quickly	3.42	3.43	ns	3.42	3.39	**	3.43	3.42	ns	3.42	3.41	ns
Good communication	3.60	3.61	ns	3.60	3.58	*	3.61	3.60	ns	3.59	3.60	ns
Satisfaction with care												
Rate personal doctor	8.99	8.95	ns	8.98	8.91	*	8.96	8.98	ns	9.00	8.95	*
Rate specialist	8.97	8.90	*	8.95	8.85	*	8.95	8.93	ns	8.95	8.93	ns
Rate care	8.93	8.88	*	8.92	8.78	**	8.93	8.90	ns	8.91	8.89	ns
Rate plan	8.84	8.70	**	8.81	8.66	**	8.83	8.75	**	8.81	8.77	ns
Customer service												
Medicare customer service	2.53	2.52	ns	2.53	2.50	ns	2.53	2.53	ns	2.56	2.51	**
Respectful treatment	3.73	3.75	ns	3.74	3.72	*	3.75	3.74	ns	3.74	3.73	ns

*p<0.05; **p<0.01.

NOTE: ns = difference is not significant.

Eye Exam—Fifty-eight percent of MFFS beneficiaries with diabetes (n=9,569) received an eye exam in the year prior to responding to the CAHPS® survey. The average patient satisfaction ratings of the Medicare plan, specialist care, and health care in general were higher among those receiving eye exams than persons with diabetes who had not received an eye exam in the previous year. Similarly, the average rating of the Care Quickly composite was higher among those receiving an eye exam in the past year.

Self-Care and Diabetes Education—Approximately 40 percent of beneficiaries had engaged in self-blood glucose monitoring (as evidenced by Medicare claims for glucometers, test strips, or other monitoring devices) or had attended diabetes self-management training classes. This indicator of care quality was significantly associated with only one of the CAHPS® indicators: beneficiaries who had engaged in self-blood glucose monitoring or diabetes education provided higher ratings of the Medicare plan, on average, than other beneficiaries.

Disease Progression—One of three MFFS beneficiaries with diabetes (n=5,656) had received health care services to treat a complication of diabetes during the year prior to responding to the CAHPS® survey. This indicator was associated with patient ratings of their main doctor or nurse and satisfaction with Medicare customer service; beneficiaries with complications from diabetes provided more positive responses to these indicators, on average, than beneficiaries without complications.

Discussion and Implications—Overall, we found that patient satisfaction measures were related to a subset of the clinical measures of health care quality under study. Specifically, beneficiaries with diabetes who received recommended eye exams and at least one of the recommended physiological measures were more satisfied with their health care than their peers who did not receive these preventive care services. In contrast, the clinical measures of self-care and disease progression were not associated with the CAHPS® indicators.

However, the significant differences found among persons who received recommended eye exams and physiological tests represent very small differences in average CAHPS® scores. Therefore, one should consider whether the significant differences reported here reflect meaningful differences in patient satisfaction. Additional research could investigate the relationship between patient satisfaction and alternative clinical measures of health care quality.

8.5.4 Depression Among 2000 and 2001 MFFS Beneficiaries

A third project that we have begun using MFFS claims data involves identifying beneficiaries diagnosed with depression using ICD-9 codes and examining the relationship between depression diagnosis and mental health component scores derived from the SF-12 questions included on the CAHPS® MFFS survey. To date, we have begun examining the following different algorithms for identifying beneficiaries:

- **Depression #1 (Strict definition).** Persons without bipolar disorder or schizophrenia who meet one of the following criteria:
    - 296.2 or 296.3 as the primary diagnosis for an inpatient admission; or
    - 296.2 or 296.3 as any diagnosis for at least two outpatient visits; or
    - 296.2 or 296.3 for any diagnosis for at least one outpatient visit and at least one inpatient admission.
- **Depression #2.** Persons without bipolar disorder or schizophrenia who meet one of the following criteria:
    - a diagnosis of 298, 311, 300.4, 301.12, or 309.1 as the primary diagnosis for an inpatient admission; or
    - an occurrence of 298, 311, 300.4, 301.12, or 309.1 as any diagnosis for at least two outpatient visits; or

  - an occurrence of 298, 311, 300.4, 301.12, or 309.1 as any diagnosis for at least one outpatient visit and at least one inpatient admission.

- **Depression #3.** Persons without bipolar disorder or schizophrenia who meet the following criterion:

  - 298, 311, 300.4, 301.12, or 309.1 in any field.

We anticipate continuing work on this analysis during the next year.

SECTION 9
REFERENCES

American Association of Clinical Endocrinologists and the American College of Endocrinology: The American Association of Clinical Endocrinologists medical guidelines for the management of diabetes mellitus: The AACE system of intensive diabetes self-management—2002 update. Endocr Pract 8(Suppl 1):40-82, 2002.

Brody, E., West, N., Carpenter, L., et al.: Medicare Fee-for-Service National Implementation Subgroup Analysis: Chartbook of Results. Report prepared for the Centers for Medicare & Medicaid Services. Research Triangle Park, NC. RTI International, 2003.

Campbell, L., Dimitropoulos, L.L., Carpenter, L., et al.: The costs and benefits of improving response rates of the CAHPS® Medicare Fee for Service Survey. Proceedings of the Joint Statistical Meetings of the American Statistical Association, 2004.

Chen, H., Khavjou, O., and Finkelstein, E.: Medicare Quality Monitoring System Version 3 Specifications—Processes of Care and Progression of Disease Measures, 1992–2001. Report prepared for the Centers for Medicare & Medicaid Services under Contract No. 500-95-0058. Research Triangle Park, NC. RTI International, 2003.

Cleary, P.D., Zaborski, L., and Zasklavsky, A.M.: Evaluation of Effects of Prescription Medicine Questions in MMC CAHPS®. Report prepared for the Centers for Medicare & Medicaid Services, 2002.

Deville, J.C., and Sarndal, C.E.: Calibration estimation in survey sampling. J Am Stat Assoc 87:376-382, 1992.

Dillman, D.A.: Mail and Internet surveys: The Tailored Design Method (2nd ed., pp. 149-193). New York. Wiley, 2000.

Dimitropoulos, L.L., Campbell, L.N., and Iannacchione, V.G.: The Effect of Method of Delivery and Response Propensity on Response to a Nonresponse Follow-up. Paper presented at the annual meeting of the American Association for Public Opinion Researchers. Nashville, TN, May 2003.

Elliott, M.N., Solomon, M., Suttorp, M., et al.: Geographic Unit Recommendations for National CAHPS® Medicare Fee-for-Service Surveys. Report prepared for the Centers for Medicare & Medicaid Services, 2000.

Hebert, P.L., Geiss, L.S., Tierney, E.F., et al.: Identifying persons with diabetes using Medicare claims data. Am J Med Qual 14(6):270-277, 1999.

National Institute of Diabetes and Digestive and Kidney Disease: Prevent Diabetes Problems: Keep Your Diabetes Under Control. NIH Pub. No. 03-4349. 2003. Available at: http://diabetes.niddk.nih.gov/dm/pubs/complications_control/index.htm#8. Accessed April 15, 2004.

Pope, G., Ellis, R.P., Ash, A.S., et al.: Diagnostic Cost Group Hierarchical Condition Category Models for Medicare Risk Adjustment. Final report submitted to the Health Care Financing Administration under Contract No. 500-95-048. Waltham, MA. Health Economics Research, Inc., December 2000.

Pugh, N., Iannacchione, V.G., Lance, T.X., et al.: Evaluating mode effects in the Medicare CAHPS® Fee-for-Service Survey. Proceedings of the American Statistical Association, Section on Survey Research Methods, 2002.

RTI and RAND: Implementation of Medicare CAHPS® Fee-for-Service Survey: Final Report for Year 1. Prepared for the Centers for Medicare & Medicaid Services under Contract No. 500-95-0061. 2001.

RTI International: Analysis of Case-Mix Strategies and Recommendations for Medicare Fee-for-Service CAHPS®: Case-Mix Adjustment Report: 2003. Final report prepared for the Centers for Medicare & Medicaid Services under Contract No. 500-95-0061. Research Triangle Park, NC. RTI International, 2004a.

RTI International: Medicare CAHPS® Fee-for-Service Survey, Subgroup Analysis Results (2000–2003). Report prepared for the Centers for Medicare & Medicaid Services. Research Triangle Park, NC. RTI International, 2004b.

RTI International: Sample Selection, Weighting, and Data Presentation Report: 2003. Final report prepared for the Centers for Medicare & Medicaid Services under Contract No. 500-95-0061. Research Triangle Park, NC. RTI International, 2004c.

RTI International: SUDAAN Language Manual, Release 9.0. Research Triangle Park, NC. RTI International, 2004d.

Singh, A.C., and Folsom, R.E.: A generalized exponential model of sampling weight calibration for extreme values, nonresponse, and poststratification. Proceedings of the Section on Survey Research Methods of the American Statistical Association, 2000.

Appendix A:
2003 MEDICARE CAHPS$^{®}$ FEE-FOR-SERVICE SURVEY