# Adjusting Performance Measures to Ensure Equitable Plan Comparisons

Alan M. Zaslavsky Ph.D., Lawrence B. Zaborski, M.S., M.A., Lin Ding, Ph.D., James A. Shaul, M.H.A., Matthew J. Cioffi, B.S., and Paul D. Cleary, Ph.D.

*When comparing health plans on scores from the Medicare Managed Care Consumer Assessment of Health Plans (MMC-CAHPS®) survey, the results should be adjusted for patient characteristics, not under the control of health plans, that might affect survey results. We developed an adjustment model that uses self-reported measures of health status, age, education, and whether someone helped the respondent with the questionnaire. The associations of health and education with survey responses differed by HCFA administrative region. Consequently, we recommend that the case-mix model include regional interactions. Analyses of the impact of adjustment show that the adjustments were usually small but not negligible.*

## INTRODUCTION

In 1995, the Agency for Healthcare Research and Quality (then called the Agency for Health Care Policy and Research) initiated a cooperative agreement with RAND, Harvard, and Research Triangle Institute to conduct the CAHPS® study. The goals of the CAHPS® project included developing a standardized survey that could be used to assess the experience of consumers in different types of insurance arrangements and health care plans (Weinberger, 1999). In 1997, HCFA funded the CAHPS® consortium to develop a version of CAHPS® suitable for assessment of the experiences of Medicare beneficiaries in managed care. HCFA now uses that survey to assess Medicare managed care plans annually.

Several methodological problems complicate the measurement and reporting of health care data, particularly when reports draw comparisons among health plans, as is the case in the MMC-CAHPS® project. Among the challenges is the need to adjust appropriately for patient characteristics such as patients' health and sociodemographic characteristics, which are not under the control of plans and which may affect CAHPS® scores.

There are at least two reasons why it might be desirable to adjust plan CAHPS® scores. First, there are certain processes that one would expect to vary according to the characteristics of patients. For example, one CAHPS® question is "…how much of a problem did you have finding or understanding the information…from your health plan?" Although it is desirable to communicate clearly with all patients, it probably is harder to do so with patients who have less education than with other patients. Second, certain personal patient characteristics might influence the response to questions, even if the process of care is the same for all patients.

To develop a case-mix adjustment model for the MMC-CAHPS® data, we first reviewed published studies. Next, we analyzed five data sets from surveys of health

maintenance organization (HMO) populations in different parts of the United States and identified potentially important case-mix variables (Cioffi et al., 1998). Finally, we analyzed MMC-CAHPS® data to evaluate alternative models. We analyzed data from the first MMC-CAHPS® survey, based on CAHPS® 1.0, and from the second and third surveys, based on CAHPS® 2.0 (Cleary et al., 2000).

## PREVIOUS RESEARCH

Given that there are few published studies of factors affecting health plan ratings, we reviewed studies of hospital care, ambulatory medical services, and health plans. Patient characteristics that have been identified as correlates of patient reports about their health care include (1) patient sociodemographic characteristics, (2) overall perceived health status, (3) functional status, (4) diagnoses or conditions, (5) length of relationship with provider or health plan and prior use of services, (6) whether the survey was completed by a proxy, and (7) institutional status (Cleary and McNeil, 1988; Aharony and Strasser, 1993; Weiss, 1988; Hall et al., 1990; Cleary et al., 1992; Zapka et al., 1995; Kane, Maciejewski, and Finch, 1997; Kippen, Strasser, and Joshi, 1997; Lee and Kasper, 1998).

Better self-reported health is consistently associated with higher ratings of health care services by consumers and patients (Cleary and McNeil, 1988; Aharony and Strasser, 1993; Marshall, Hays, and Mazel, 1996; Hall, Milburn, and Epstein, 1993; Roberts, Pascoe, and Attkisson, 1983; Hall et al., 1998). Current general health status tends to be the strongest predictor of patient or consumer satisfaction with health care services (Hall et al., 1990; Cleary et al., 1992; Zapka et al., 1995; Kane, Maciejewski, and Finch, 1997; Kippen, Strasser, and

Joshi, 1997; Lee and Kasper, 1998). Perceived improvement in health also has a strong positive association with health care ratings (Kane, Maciejewski, and Finch, 1997; Kippen, Strasser, and Joshi, 1997).

The few studies that have investigated the relationship between satisfaction ratings and the presence of specific medical conditions have yielded inconsistent results (Hall et al., 1990; Zapka et al., 1995; Kippen, Strasser, and Joshi, 1997). Emotional distress and social-activity limitations have been found to be negatively associated with satisfaction ratings, although work limitations or other limitations due to emotional health status have not (Marshall, Hays, and Mazel, 1996; Greenley, Young, and Schoenherr, 1982).

Early investigations revealed that older patients are generally more satisfied than younger patients with their medical care (Cleary and McNeil, 1988; Aharony and Strasser, 1993; Zapka et al., 1995), although findings are not consistent (Weiss, 1988; Kane, Maciejewski, and Finch, 1997), and some studies have found that this association is not present in the oldest groups of patients (Hall et al., 1990; Lee and Kasper, 1998).

Although some studies have found that females were less satisfied than males (Cleary et al., 1992), the preponderance of studies found that gender is not a significant predictor of satisfaction (Cleary and McNeil, 1988; Aharony and Strasser, 1993; Weiss, 1988; Hall et al., 1990; Zapka et al., 1995; Kane, Maciejewski, and Finch, 1997; Lee and Kasper, 1998). Studies of the association between respondents' race and ratings of their medical care and health insurance plans have had inconsistent results (Weiss, 1988; Zapka et al., 1995; Kane, Maciejewski, and Finch, 1997; Kippen, Strasser, and Joshi, 1997). Evidence about the association between education levels and consumer ratings of medical care and health plans is also

inconsistent (Weiss, 1988; Zapka et al., 1995; Kane, Maciejewski, and Finch, 1997; Kippen, Strasser, and Joshi, 1997).

Most investigations of the relationship between income and consumer reports about health care find that those with higher incomes provide modestly higher or similar ratings, compared with those with lower incomes (Weiss, 1988; Hall et al., 1990; Cleary et al., 1992; Zapka et al., 1995; Kane, Maciejewski, and Finch, 1997). A study of Medicare beneficiaries indicated that more income is associated with higher levels of health care satisfaction (Lee and Kasper, 1998).

Possession of additional health insurance coverage such as Medicaid or private supplemental policies may be an important predictor of health plan satisfaction. One study found that Medicare fee-for-service beneficiaries with secondary insurance were more likely than others to be highly satisfied with their medical care (Lee and Kasper, 1998). Groups with both Medicaid and Medicare coverage were more likely to be highly satisfied with their health care than those with Medicare only.

Older individuals are often cognitively or physically impaired and may be unable to complete a survey or unavailable to respond (Corder, Woodbury, and Manton, 1996). As a result, surveys are often completed by a close relative or caregiver. Proxies tend to rate the patient's health status lower than the patients rate themselves (Epstein et al., 1989; Rothman et al., 1991; Magaziner et al., 1988; Sprangers and Aaronson, 1992; Magaziner et al., 1996; Rubenstein et al., 1984; Magaziner et al., 1997). Subjective health-status dimensions tend to be more greatly underrated by proxies. Reports about observable physical or functional characteristics, such as the ability to walk upstairs or dress, are much more consistent between patient and proxy than less concrete or more private health dimensions, such as emotional well-being (Epstein

et al., 1989; Rothman et al., 1991; Magaziner et al., 1988; Sprangers and Aaronson, 1992; Magaziner et al., 1996; Rubenstein et al., 1984; Magaziner et al., 1997). Reports about the patient's general health status tend to have the lowest levels of concordance (Magaziner et al., 1988). Some studies have found that the association between proxy and patient reports increased with higher education levels for either respondent (Sprangers and Aaronson, 1992; Hays et al., 1995) and with increased contact between patient and proxy respondent (Epstein et al., 1989; Sprangers and Aaronson, 1992; Rubenstein et al., 1984). An increased burden of caring for the patient is associated with more negative ratings by proxies relative to those given by patients (Epstein et al., 1989; Magaziner et al., 1988). Reports about more subjective aspects of health appear to be influenced much more by the proxy's own level of psychological distress, age, and health status than by characteristics of the patient (Rothman et al., 1991).

Few studies have investigated the effect of proxy ratings on reports about medical care and health plans (Epstein et al., 1989; Lavizzo-Mourey, Zinn, and Taylor, 1992). One study comparing responses of 60 elderly patients with their proxies' satisfaction ratings found the association to be modest (Epstein et al., 1989). Proxies consistently rated the patient's care more negatively than did patients themselves. The majority (62 percent) of proxies in this study were spouses. Another study found modest associations between proxy and patient ratings. However, in that study, surrogates generally rated the patient's care more positively than did patients themselves (Lavizzo-Mourey, Zinn, and Taylor, 1992).

## PRELIMINARY DATA AND ANALYSES

The data sets we used for preliminary analyses were chosen, in part, to complement the existing literature. Only one pub-

lished study investigated the relationship between consumers' ratings and sociodemographic characteristics specifically for Medicare beneficiaries. We selected data sets that included recent consumer evaluations of their health insurance plans, in addition to ratings of their medical care (Table 1). Three of the five data sets included a substantial number of Medicare beneficiaries from diverse geographic regions of the United States. Two of the data sets (Washington and private employer) included many questions that matched items on the 1997 MMC-CAHPS®. Each data set included information on age, sex, education, general health status, and race. Information about chronic or disabling conditions, physical functioning, emotional well-being, income, and proxy responses was included in many of the data sets.

## Minnesota Data

In 1995, the Minnesota Health Data Institute (MHDI) conducted a study of all health plans in Minnesota for commercially insured groups, Medicare, and medical assistance programs. The study included traditional indemnity and managed care plans. More than 17,000 surveys were collected for 46 different health plans. The MHDI used a survey that was similar to the National Committee on Quality Assurance's (NCQA's) Annual Member Health Care Survey (AMHCS) (NCQA 1.0).

## NCQA Survey

In 1996, NCQA required that health plans submitting results for the Health Plan Employer Data and Information Set (HEDIS®) report data for the AMHCS. A total of 43 health plans submitted results for more than 18,000 completed surveys. The majority of data compiled by the NCQA was for commercially insured groups. We limited our analyses to data from Medicare beneficiaries (862 responses).

The NCQA data presented us with two unique opportunities to study the relationship between self-reported health-status measures and health care ratings for Medicare beneficiaries. First, the data included the Medical Outcomes Study (MOS) Short Form 12 (SF-12) health survey (Ware, Kosinski, and Keller, 1996), which collects data about physical functioning and emotional well-being. Additionally, the NCQA survey included unusually complete data (26 questions) about chronic or disabling conditions that respondents had.

## Medicare Current Beneficiary Survey (MCBS)

The MCBS is administered each year to measure the economic and quality of life effects of the Medicare program on enrollees. Each year approximately 16,000 Medicare beneficiaries or their proxies are interviewed, of whom about 4,000 are new to the panel. Although the MCBS focuses on economic aspects of the enrollees' experience with Medicare, it also includes a number of questions about their experiences with their doctors and medical care. The MCBS is unique in that it collects satisfaction information for a nationally representative sample of Medicare beneficiaries. We used the 1996 MCBS, which includes a supplemental sample of beneficiaries in managed care organizations.

## Private Employer Survey

In fall 1998, a large private employer sponsored an evaluation of the plans offered to employees and retirees, using the CAHPS® adult core survey. Survey responses were

**Table 1**

**Data Set Components for Preliminary Analyses**

| Project | Minnesota | NCQA | Washington | Private Employer | MCBS |
|---|---|---|---|---|---|
| Survey Used | NCQA Annual Member Health Care Survey Hybrid | NCQA Annual Member Health Care Survey | CAHPS® Adult Core Version 1.0 | CAHPS® Adult Core Version 1.0 with SF-12 Attached | 1996 MCBS |
| Method of Data Collection | Mail | Not Available | Mail | Mail | Personal Interview |
| Percent Responding | Not Available | 72 | 52 | 61 (Salary) 45 (Hourly) | Not Available |
| Dates of data collection | April-June 1995 | 1996 | Summer 1997 | Fall 1997 | 1996 |
| Number of Health Plans | 46 | 114 | 20 | 11 | Not Available |
| Number of Respondents | 17,591 | 862 | 8,310 | 4,678 | 16,411 |
| Medicare[2] | Yes | Yes | No | Yes, not able to identify. | Yes |
| Sample Description | Survey of commercial, Medicare, and Medicaid plan in the State to provide information to consumers. | A mix of enrollees ranging from children to Medicare clients. No plans were exclusively Medicare. | State employees. | Large manufacturing company employees and their dependents. Stratified by hourly and salary positions. | Rotating panel design to provide information about Medicare beneficiaries. 1996 survey included a supplemental sample of enrollees in managed care. |
| **Demographics**[3] Percent in Managed Care | Not Available | Not Available | 92.9 | 82.2 | 22.7 |
| Age Percent 65 or Over Percent 18-64 | 97.9 2.1 | 94.8 5.2 | 1.6 98.4 | 21.0 79.0 | 85.5 14.5 |
| Percent Male | 34.0 | 27.0 | 39.8 | 72.6 | 44.1 |
| Education Percent with Some High School or Less Percent High School Graduate Percent with Some College or More | 27.7 35.4 36.9 | 26.7 31.7 41.6 | 1.2 16.4 82.2 | 9.0 29.0 62.0 | 39.8 33.0 27.2 |
| Percent White | Not Available | 92.0 | 86.5 | 85.0 | 79.7 |

[1] Limited to Medicare enrollees.

[2] At least part of the data.

[3] Shown for Medicare sample.

NOTES: NCQA is National Committee on Quality Assurance. MCBS is Medicare Current Beneficiary Survey. CAHPS® is Consumer Assessment of Health Plans.

SOURCE: Data from the CAHPS®, Agency for Healthcare Research and Quality; data analysis by the authors.

collected from 4,678 current employees and retirees in 11 different health plans. Both managed care and traditional indemnity plans were included in the evaluation.

These data were unique because they included SF-12 health-status measures as well as CAHPS® survey data. These data allowed us to explore the relationships between self-reported health measures and consumer ratings that were exactly the same as those collected with the 1997 MMC-CAHPS®.

### Washington State

An evaluation of 20 health plans offered to State employees by the Washington State Health Care Authority obtained 8,310 responses. The study was conducted in summer 1997 using the CAHPS® adult core survey, which includes a number of items that matched those in the 1997 MMC-CAHPS®.

### ANALYSES

We estimated linear models in which the dependent variable is the response on a survey item or set of items (composite) and the independent variables are case-mix adjusters. In the data sets in which sample size was adequate and the data distinguished among multiple plans, we included dummy variables for each of the plans. When we control for plan effects, the case-mix coefficients represent within-plan effects of the adjuster variables. We tested the predictive power of variables individually and in combination.

### Results of Preliminary Analyses

Because of the large number of analyses involved, we do not present details of the empirical results from the preliminary

analyses in this article. A more detailed description of the analyses and results is available from the authors (Cioffi et al., 1998).

### Self-Reported Health Status

In each of the preliminary data sets, current general health status was the strongest predictor of health care and health plan satisfaction for both commercially insured and Medicare enrollees. Individuals who rated their general health levels higher also gave higher ratings of their health care services.

A general health-status variable was analyzed as a continuous or categorical variable in the MCBS, Washington, private employer, and NCQA data sets. The continuous variable accounted for the same amount of variation as the categorical variable, because ratings of medical care and of health plans improved by about the same amount for each step on the general health-status response scale.

### Physical Functioning, Comorbidities, and Chronic Conditions

Our preliminary analyses generally indicated that measures of physical-functioning limitations were not significant independent predictors of care ratings. Analyses of data from Medicare beneficiaries in the NCQA data set also indicated that physical functioning was not a significant predictor, after controlling for emotional status, general health status, and age. Analysis of the private employer data, which consists of retirees and current employees, also revealed that work or activity limitations, physical functioning limitations, and limitations due to pain were infrequently related to health care ratings after controlling for age, general

health status, and emotional well-being. In the Washington State data, a variable indicating whether the survey respondent needed help with personal care or meeting routine needs, or had a condition that interfered with his or her independence, with work, or with school activities, was not a significant predictor for the respondent's ratings of overall medical care, specialists, or personal doctors. Having a limitation was significantly associated only with the respondent's overall rating of health plan.

We studied the association between medical care and health plan ratings, and having a chronic or disabling condition, using the MCBS, NCQA, and private employer data sets. A variable constructed from the NCQA data set indicated the number of conditions (from a list of 26) an individual reported having. For the private employer data set, we analyzed a variable that indicated whether respondents had a medical condition that had lasted for 3 or more months. Neither chronic nor disabling conditions were significant predictors of satisfaction outcomes in these data sets. In the MCBS, however, individuals who reported having any of four physical conditions tended to provide higher health care satisfaction ratings.

Emotional Well-Being

In the NCQA analyses, a general measure of emotional well-being was a significant predictor of health care ratings. Feeling calm and peaceful was the most important emotional-status predictor of higher levels of satisfaction in the private employer data analysis. Respondents with fewer work and social limitations due to emotional distress gave higher satisfaction ratings, but these results were inconsistent. Feeling energized was not an inde

pendent predictor of satisfaction levels, because it was strongly correlated with feeling calm and peaceful.

Age

Older adults are generally more satisfied with their medical care and health plan services. However, ratings do not increase monotonically with age for those over age 65. Among Medicare beneficiaries in Minnesota, older individuals are less satisfied than younger ones with their health plan, medical care, and access to care. There were no significant differences among age groups for appointment access and physician choice.

Results from Medicare enrollees in the NCQA data file indicated that satisfaction increases with age until the 80-84 or 85-89 age groupings, at which point it levels off or declines. Evidence from the MCBS suggests that satisfaction decreases with age, with most of this effect resulting from lower satisfaction among the oldest respondents (those over age 85).

There is little evidence that age affects the relationship between ratings and the other sociodemographic characteristics. No significant interaction effect was discovered between age and health status. The effects of education, income, and Hispanic background were related to age for those 85 years and over, but these interactions were inconsistent and only marginally significant.

Sex

Sex was not significantly related to health plan or medical care ratings. Analyses of employed adults and Medicare beneficiaries indicated that females were more satisfied than males, but the effects were only marginally significant.

Race

The relationships between race and ratings of health plan or medical care were inconsistent in the preliminary data sets. Among retirees and current employees in the private employer data, race was a significant predictor only for ratings of specialists but not for health plans, personal doctors, or overall health care. Black persons, Asians, and Pacific Islanders tended to be more satisfied than white persons in this study. (Throughout this discussion, the term "white persons" refers to white people who are not Hispanic.) In the Washington State data, race had a significant relationship only with the specialist ratings: Hispanic persons and Asians were less satisfied than white people.

Among Medicare beneficiaries in the NCQA data set, Hispanic people tended to give lower ratings than white persons of their health plans and of getting approvals or referrals for care. No other differences were significant. Analyses of the MCBS data set revealed somewhat different results. Black people tended to be less satisfied with both their doctors and medical care than white people. Hispanic persons were more satisfied than white people with their doctors, but less satisfied than white people with their care.

Education

More educated Medicare beneficiaries in the Minnesota and MCBS data sets rated their care higher than those who were less educated. There was no relationship, however, among Medicare enrollees in the NCQA data set. In the private employer and Washington data, those with more education tended to be less satisfied with their medical care and health insurance plan. An ordinal education variable predicted ratings as well as a set of categorical variables, because satisfaction levels changed roughly linearly.

Income

Only the 1996 MCBS data set included an income variable that allowed us to study the relationship between the income levels of Medicare beneficiaries and ratings of medical care. Our results were consistent with findings from a study showing that elderly respondents with higher income levels tended to rate their medical care better than other respondents (Lee and Kasper, 1998). Increases in satisfaction at higher income levels were generally modest. Effects of secondary sources of health insurance coverage were not assessed in any of the preliminary data sets.

**Proxy Respondents**

We were able to assess the impact of response by a proxy only for the MCBS data set. Proxy ratings of medical care are only marginally higher than those provided by the intended Medicare respondents.

**Context Variables**

We studied several variables describing the social context or community in which the respondent lives, using ZIP Code level 1990 U.S. census data. We considered seven variables, each of which is measured as percentage of residents in the respondent's ZIP Code who belong to the respective group: Ethnicity (Black, Asian, Hispanic), College-Educated, High-Status Occupation, Urban Resident, Public Assistance Recipients (overall and among those over the age of 65).

The NCQA analysis revealed that individuals who live in areas with a high percentage of Asians were more likely to report satisfaction with their plan and with ability to get referrals. Respondents from areas where there is a high concentration of black residents and from densely

populated urban areas report fewer problems with their health care. These respondents report greater satisfaction with their plan overall and also with the quality of care received. In the MCBS data, residents of areas with large concentrations of college-educated individuals and residents of areas with many persons with high-status occupations tended to be more satisfied with their doctors and rated their overall care higher.

In the Washington State analysis, residents of more urbanized areas were more satisfied with doctors, specialists, and health plans. Respondents from areas with high percentages of college graduates or persons with high-status occupations tended to provide higher ratings for both their doctors and their overall health care. However, when we include both of these highly correlated variables in our model, only education had a significant positive effect on the CAHPS® scores (an effect that is opposite to that of individual level education in this particular analysis), while occupation has a positive effect on satisfaction with doctor.

## SUMMARY OF PRELIMINARY ANALYSIS RESULTS

The effect of a few patient characteristics, particularly health status and age, are consistent across multiple studies, while others have effects that are either weak or inconsistent. Some of the inconsistencies might be attributable to the diverse settings and populations studied. In particular, population-based studies might confound case-mix effects with selection of some groups or patients into more or less favorable situations. In the next section, we report analyses of Medicare CAHPS® data that compare patients' reports within a single reimbursement system and a large number of health plans.

## MMC-CAHPS® Data

Instrument

The MMC-CAHPS® survey, fielded in 1997, included all items of the CAHPS® 1.0 adult core instrument (Hays et al., 1999) and 28 additional MMC-specific items (Cleary, Zaslavsky, and Cioffi, 2000). Of 85 items, 4 elicit overall ratings and 34 elicit reports of respondent experiences. Other questions are used to determine the applicability of particular report questions or ask about sociodemographic characteristics, health status, and health care utilization. The MMC-CAHPS® survey fielded in 1998 and 1999 included all items of the CAHPS® 2.0 adult core instrument and 41 additional MMC-specific items. The potential case-mix variables in the MMC-CAHPS® questionnaire are available from the author, including 10 variables from the survey and 6 variables based on respondents' ZIP Codes.

## Sample

For each MMC-CAHPS® survey, HCFA drew a stratified sample of Medicare beneficiaries who had been enrolled in an eligible plan. Eligible plans included all health plans with Medicare contracts in effect on or before January 1 in the year preceding the survey and in business for 2 years. Contracts that covered large areas were divided into geographically defined reporting units. A simple random sample of up to 600 members was drawn from each plan or reporting unit.

For each survey, we deleted cases sampled from contracts that had ceased activity, had only one beneficiary (two plans in the second year) or had been terminated, and beneficiaries that left their plan before the survey was administered, as well as deceased and institutionalized beneficiaries.

In the first survey, there were 89,802 valid surveys from 119,267 eligible beneficiaries, representing a response rate of 75 percent. In the second survey, there were 123,000 valid surveys from 152,144 eligible beneficiaries, representing a response rate of 81 percent. In the third survey, there were 166,072 valid surveys from 202,775 eligible beneficiaries, representing a response rate of 82 percent.

## Survey Procedures

Survey data collection took place from February to May 1998 for the first survey and from September to December for the second (1998) and third (1999) surveys. Although there were slight modifications in survey protocols, the basic approach was comparable each year. The survey firm mailed a preliminary notification letter, followed by the survey. Non-respondents were sent a reminder postcard, and if no survey was received, a duplicate survey was sent. Interviewers contacted respondents by telephone to complete missing items and to followup for non-response, if a telephone number could be obtained.

## MMC Analyses

The statistical criteria for usefulness of a variable for case-mix adjustment include both its predictive power in the pooled within-plan regression model and the degree of between-plan variability in the variable, relative to its within-plan variability. In the analyses presented here, we combined information about predictive power and between-plan variability to obtain an overall summary of the impact of the variable on adjustment: the ratio of the variance of the adjustments for the new variable to the between-plan variance of the

unadjusted means (Zaslavsky, 1998). We first evaluated explanatory power (EP) using a linear specification and then tested the improvement in fit by replacing it with variables for each response category.

We also investigated the possibility that the effects of case-mix variables would vary by region. We assigned all responses from each plan to the single region in which the plan had the largest enrollment, as determined by the Medicare Managed Care Market Penetration for All Medicare Plan Contractors Quarterly State/County/ Plan data file. This allowed us to adjust each plan using a single model and facilitated comparisons among plans operating in the same area. For most plans, 70 percent or more of their enrolled population was within a single region. Because of the small number of plans and managed care enrollees in several HCFA regions in the first year of CAHPS® data examined, we combined regions 5, 7, and 8 for case-mix modeling. For consistency across years, we retained that grouping for each year.

We tested interactions between region and a linear effect of age, education, and reported health status, in models predicting the four CAHPS® general ratings. To evaluate whether it was necessary to create an interaction term for each age category, we assessed alternative models, comparing a model in which a regional interaction was estimated for each category with a model containing a regional interaction with the linear age effect. Similar analyses were performed for the education and health-status interactions.

We analyzed all 3 years of MMC-CAHPS® data. To simplify the presentation, we show only year three results in the tables. Tables containing all 3 years are available from the authors.

## MMC Results

We examined the associations between adjuster variables and CAHPS® ratings and the explanatory power of potential adjuster variables to select a final adjustment model. In Table 2, we present two sets of analyses. In Model 1, we present the explanatory power of variables controlling only for age and health. In Model 2, we present the explanatory variables after controlling for a set of core variables. Later, we discuss the rationale for using these two approaches and the results. We discuss both the predictive strength and explanatory power of variables but present data only on explanatory power for all variables. The coefficients for the core models are available from the authors.

Age and general health status were chosen for inclusion in the core model because the literature and preliminary analyses indicated that they are consistently the strongest predictors of satisfaction and because they were the case-mix adjusters in the standard model for the core CAHPS® project. In the MMC data, there was a positive relationship between age and the CAHPS® ratings, even when other demographic, health status, and contextual variables were entered into the equation. The younger group (under 65, essentially all disabled) and the group age 65 to 69 years tended to give the lowest ratings. The fractions in the extreme age groups, under age 65 and over age 80, varied greatly between plans. This suggests that age would have some impact on case-mix adjustment.

Health status was consistently the strongest positive predictor of consumers' ratings for all measures tested. On the other hand, there was less between-plan variance for health status than for some other variables. Nonetheless, because of its predictive power, health status is an important variable in the case-mix model (Table 2).

In Model 1 in Table 2, we first controlled for age and health and then entered each of the other potential adjuster variables separately to determine their EP. Controlling for health status and age, individuals with more education rated their health plans, medical care, personal doctors, and specialists lower than those with less education. This relationship was consistent for all models that were tested, even when all possible predictors were included in the model. The between-plan variance for education is large compared with that for age and health-status variables. It has the largest or second largest EP on all of the four global ratings for all 3 years. For all the composites, education had the largest EP in at least 1 of the years and for Plan Paperwork, it was the most important in all 3 years (data not shown). Based on these results, we decided to include education as part of our base case-mix model.

The MMC-CAHPS® survey asked whether beneficiaries received help filling out the survey and what type of help they received. The two proxy variables were PROXY (helped with the survey in any way) and ANYPROXY (somebody answered the survey for the subject). Both variables were significant predictors of most of the ratings and composites, but both had small EPs because their contribution to the predictive power of the model was relatively small, and the proportion of individuals receiving help did not vary much across health plans. Nevertheless, adjusting for proxy responses may be important because of common concerns that the inability of some beneficiaries to complete a survey by themselves will compromise the validity of the survey results. Thus, we included this variable in the case-mix model despite its limited impact on adjustments of scores.

Model 2 in Table 2 controls for age, health, education, and proxy responses, and tests the explanatory power of all other

**Table 2**

**Explanatory Power of Each Potential Adjuster**

|  | Ratings of | | | | | | | |
|---|---|---|---|---|---|---|---|---|
|  | Health Plan | | Medical Care | | Personal Doctor | | Specialists | |
| Adjusters | Model 1 | Model 2 | Model 1 | Model 2 | Model 1 | Model 2 | Model 1 | Model 2 |
| Age | 0.431 | — | 0.241 | — | 0.144 | — | 0.091 | — |
| Health | 0.274 | — | 0.295 | — | 0.140 | — | 0.104 | — |
| Education | 0.745 | — | 0.453 | — | 0.339 | — | 0.194 | — |
| Proxy | 0.035 | — | 0.024 | — | 0.024 | — | 0.014 | — |
| Proxy Answer | 0.003 | — | 0.003 | — | 0.004 | — | 0.002 | — |
| Male | 0.005 | 0.002 | 0.004 | 0.001 | 0.005 | 0.003 | 0.002 | 0.001 |
| Race |  |  |  |  |  |  |  |  |
| White | 0.030 | 0.007 | 0.070 | 0.045 | 0.207 | 0.163 | 0.000 | 0.000 |
| Black | 0.030 | 0.009 | 0.076 | 0.047 | 0.148 | 0.107 | 0.005 | 0.001 |
| Hispanic | 0.014 | 0.005 | 0.007 | 0.004 | 0.017 | 0.013 | 0.002 | 0.001 |
| Asian | 0.012 | 0.004 | 0.009 | 0.006 | 0.005 | 0.002 | 0.016 | 0.011 |
| Native American | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Other | 0.004 | 0.003 | 0.002 | 0.001 | 0.001 | 0.002 | 0.004 | 0.003 |
| Medical Condition | 0.001 | 0.003 | 0.011 | 0.015 | 0.012 | 0.015 | 0.024 | 0.026 |
| ADLs | 0.012 | 0.005 | 0.005 | 0.002 | 0.002 | 0.000 | 0.003 | 0.001 |
| IADLs | 0.022 | 0.010 | 0.011 | 0.005 | 0.002 | 0.000 | 0.004 | 0.002 |
| Independent | 0.026 | 0.014 | 0.029 | 0.020 | 0.009 | 0.004 | 0.011 | 0.008 |
| ZIP Code Variables |  |  |  |  |  |  |  |  |
| Percent College Degree | 0.468 | 0.075 | 0.279 | 0.054 | 0.187 | 0.025 | 0.074 | 0.003 |
| Percent Urban | 0.000 | 0.014 | 0.042 | 0.011 | 0.006 | 0.029 | 0.038 | 0.069 |
| Percent Black | 0.051 | 0.013 | 0.122 | 0.072 | 0.171 | 0.110 | 0.005 | 0.000 |
| Percent Hispanic | 0.306 | 0.120 | 0.209 | 0.104 | 0.353 | 0.195 | 0.148 | 0.116 |
| Percent Asian | 0.047 | 0.007 | 0.040 | 0.012 | 0.003 | 0.023 | 0.010 | 0.021 |
| Percent Over Age 65 Receiving Public Assistance | 0.162 | 0.024 | 0.308 | 0.158 | 0.287 | 0.129 | 0.050 | 0.016 |

NOTES: Explanatory power=(variance of adjustments for a variable)/(variance of plan means). All values are multiplied by 1,000 for legibility. In Model 1, the base model for calculation of age and health explanatory power is a null model (with only plan effects). For all other variables, Model 1 includes age and health, with plan differences absorbed as well. In Model 2, the base model for all variables includes age, health, education, and proxy. ADLs is activities of daily living. IADLs is instrumental activities of daily living.

SOURCE: Data from the Consumer Assessment of Health Plans®, Agency for Healthcare Research and Quality; data analysis by the authors.

potential adjusters. These analyses indicated that ZIP Code Hispanic, public assistance (senior), and/or self-reported Asian race had some marginal explanatory power. ZIP Code education, which appeared potentially important in tests with the first base model, was not important after controlling for individual education. The influence of each is much less than that of age, health, and beneficiary education, however. Also, the influence of these variables was not consistent across all dependent variables or the three surveys. Of the variables tested, ZIP Code Hispanic appeared to be the most important. The inconsistency in these results and the age of the census data on which these ZIP Code variables are based would argue against including these variables in the MMC-CAHPS® case-mix adjustment model. However, we examined regional interactions and a model that includes ZIP Code Hispanic as a potential model option in subsequent analyses.

Respondents reporting more medical conditions provided higher ratings of their health plan, medical care overall, specialists, and personal doctors. This counterintuitive finding may indicate that it is not the mere presence of disease that leads to lower satisfaction ratings, but the level of severity and disabling effect that accompanies the disease. In addition, individuals who use health care services more frequently might be more knowledgeable about their condition and be more likely to report a condition on the survey. Higher use of services may also indicate increased satisfaction with the services received. However, the mean number of chronic conditions did not vary much across Medicare plans. Therefore, adjustment for the prevalence of medical conditions would not have much impact on health plan ratings.

MMC-CAHPS® respondents were asked three questions about having a health problem that (1) caused them to need help with personal care needs, such as eating, dressing, or getting around the house, (2) caused them to need help with routine needs, such as everyday household chores, doing necessary business, shopping, or getting around for other purposes, and (3) seriously interfered with their independence, participation in the community, or quality of life. Two of the physical-functioning indicators were related to ratings, even after controlling for general health status. Respondents with a physical limitation that interfered with independence, participation in the community, or quality of life rated their health plans, medical care, specialists, and personal doctors lower. Respondents that needed help with personal care were more likely to give lower ratings of the health plans and medical care overall and marginally lower ratings of specialists. Needing help with routine needs such as household chores or shopping was not a significant predictor of ratings due to its high correlation with needing help with personal care needs.

Although physical-functioning indicators were significant predictors for the Medicare population, their predictive power was modest compared with self-reported general health status, and they varied little across plans. Therefore, including these variables in the model would have little effect on the outcomes.

Males reported lower ratings than females of their health plans and personal doctors in year one and lower scores on all ratings in years two and three. However, even when sex was a significant indicator, its predictive power was small and it had the smallest variation between health plans. Therefore, it had very little impact in the case-mix adjustment.

The relationships between race and health plan or medical care ratings were not consistent. Asian Medicare beneficiaries were the only group to consistently rate aspects of their health plans, medical care, personal doctors, and specialists lower than white beneficiaries. Black and Hispanic persons rated their health plans marginally higher than white persons but did not differ significantly for ratings of their medical care overall. Black persons were significantly more positive than white persons about their personal doctors, while Hispanic people were marginally more positive. Hispanic persons also provided marginally lower ratings of their specialists compared with white persons. Native Americans provided marginally lower ratings of their health plans than did white persons. We did not recommend using race and/or ethnicity variables in a national case-mix model because of the lack of consistency in their effects. We were also concerned that their effects might depend on local associations of cultural and socioeconomic characteristics with race and ethnicity that might vary from region to region.

All six of the contextual variables, which describe the ZIP Code area in which a person lives, had large between-plan differences. This is understandable because these variables represent averages over areas, and plans also tend to operate within areas. The ratio of within- to between-plan variances for these variables are larger than those for almost all of the individual-level variables. Therefore, the ZIP Code variables typically had an impact on case-mix adjustment when they were significantly related to CAHPS® scores.

Respondents from areas containing more educated residents were more likely to provide slightly lower ratings for health plans in all 3 years and for specialists in year one. There were marginally positive effects on health care ratings in year one for respondents from areas in which many residents were black or Hispanic. In years two and three, persons from areas with many Hispanic residents had higher ratings of plans and doctors, and in year two for care as well. Respondents from an urban area gave slightly lower health care ratings in years two and three but higher specialist ratings in year two. Although the effects of the racial/ethnic and poverty contextual variables are interesting, we are reluctant to use them now because the effects are inconsistent and for the same reasons as for the individual racial/ethnic variables.

For each of the individual-level variables—age, health status, and education—we calculated $F$-tests that compared the model with the variable entered in the linear (one-coefficient) specification to the model with the variable entered as a set of dummy variables. For the age and education variables, the test clearly rejected the simpler (linear) specification for each of the four rating scales (data not shown). The effect of age showed a clear trend for most levels and outcome variables, in which ratings increased with age. On the other hand, the steps in mean satisfaction were not equal for each increase in age category; instead, satisfaction appeared to level off in the older categories. Similarly, the trend in the individual education variable was toward lower satisfaction with more education, but the steps were not equal for each increase in education category. Therefore, the categorical effects were more accurate representations of age and education effects than linear variables. For health status, the linear trend toward lower ratings with worse health status (coded by higher numbers on the health-status response scale) appears to be an adequate description of the relationship.

## Regional Interactions

Analyses of interaction effects showed that there were strong regional interactions for health status, education, and ZIP Code percent Hispanic response for at least four of the nine outcome measures in the year three data. Health status had significant regional interactions for six of the variables and was one of the strongest predictors of the ratings. To evaluate the stability of these interaction effects across multiple years of CAHPS® Medicare analysis, we also used a model incorporating data from years two and three of the CAHPS® survey. We did not use data from year one because of differences in the format and questions of the survey in that year. Fitting a model with 2 years of data, we allowed for slopes on case-mix adjusters to vary (by including both the region-interaction effect and a region-by-year interaction). We estimated separate models for regional interactions for age, education, health status, proxy response, and ZIP Code percent Hispanic. Each of these models included an additional interaction term allowing these regional interaction slopes to vary by year. In all cases, we found no evidence of change in the interaction effects across the years. In particular, for education and health, the ratio of the overall effect to the interaction with year was large, indicating that the regional interactions were stable over the 2 years compared (and therefore likely to represent consistent patterns rather than random variations).

We also calculated $F$-tests of the significance of regional-interaction effects in an ANOVA model, treating plan effects as the random error term. This tests whether the effect of our case-mix adjusters varies by region (i.e., an interaction of each case-mix adjuster with region) more than would be expected if more or less favorably rated plans had been randomly distributed across regions. The regional interaction effect is tested against the plan interaction effect. With this test, health status, education, and proxy response have significant differences across a region, while the region-specific ZIP Code Hispanic effect is no longer significant. This suggests that there is substantial variation in the Hispanic coefficient from plan to plan, so although the average coefficient differs across regions, it does not differ more than it would if plans had been randomly assigned to regions.

The absolute and relative magnitude of the regional effects varies substantially from year to year. Nevertheless, we suggest that it is useful to include interaction terms for health and education by region; the two variables that appeared to have the most consistent interregional variability.

## Impact of Case-Mix Adjustment

To assess the effects of adjustment on the ratings of plans, relative to the unadjusted ratings, we compared adjusted ratings with unadjusted ratings, using several measures of the differences. The results of the impact analyses were comparable for the 3 years of data. Considering the ratios of adjustment to unadjusted standard deviations for each variable, the largest impact of adjustments is on "getting care you need" and the smallest is for "ease of getting referrals." The standard deviation of plan means is only slightly smaller for the various adjusted means than for the unadjusted means.

The largest adjustments upward are comparable to one standard deviation of the plan means for most measures. The largest adjustments downward are usually much smaller, half as big or less. This

suggests that there are a few plans with unusually adverse case mix, from the standpoint of the effect of case-mix on consumer assessments.

Comparison of the ratio of the standard deviation of adjustments to the standard deviation of unadjusted means across regions suggests that the impact of adjustment may be somewhat larger in some regions than in others. Generally, the ratio is above average in the Pacific, New England, and Upper Midwest Regions, below average in New York and New Jersey, Mid-Atlantic, and South Atlantic, and mixed in Northwest and Southwest.

To quantify the effect of case-mix adjustment on the ranking of plans, we calculated the Kendall Tau correlation coefficient between the adjusted and unadjusted plan ratings. This measure is related to the fraction of pairs of plans that switched ordering as a consequence of case-mix adjustment, where the denominator is the total number of pairs of plans. (The Kendall Tau statistic stretches this quantity to a scale from -1 to +1, to make it comparable to other correlation coefficients.)

The Kendall Tau statistics for overall rating of plans in the 3 years were 0.92, 0.89, 0.91, indicating that the percentages of pairs of plans whose ordering would be changed using that adjustment model were 3.9, 5.5, and 4.5 percent, respectively. Generally, where the ratio of the standard deviation of the adjustment divided by the standard deviation of the adjusted mean is larger, the Kendall Tau is smaller and the fraction of pairs that would be switched is larger. Nonetheless, the unadjusted and adjusted means give between-plan comparisons that are in agreement, most of the time, in every region.

## SUMMARY AND CONCLUSIONS

Previous studies as well as the analyses presented here support the continued use of perceived health status and age in CAHPS® case-mix adjustment models. Although education does not explain a large proportion of the variance in the dependent variables assessed, there is more interplan variability in education than in age or health, and as a consequence, education predicts more interplan variability than either health status or age in some models.

Response by a proxy is not an important predictor of responses, either for individual-level analyses or for assessments of interplan variability. We suggest including the proxy variable primarily because of concerns about the potential effects of cognitive impairment on reports about plan experiences in this population and the likelihood that proxy respondents describe experiences with the health plan differently than enrollees would. Thus, the Medicare adjustment model now includes health status, age, education, and a variable indicating whether a proxy answered the survey. We also recommend including interaction terms for health and education by region because they are the two variables that appeared to have the most consistent interregional variability.

In general, the case-mix adjustments are not large and do not greatly change the picture of which plans are high- or low-rated. It is noteworthy, however, that the largest adjustments are quite substantial, so there are at least a few plans for which, under our models, an important part of their measured satisfaction can be attributed to case mix rather than to actual plan performance.

# REFERENCES

Aharony, L., and Strasser, S.: Patient Satisfaction: What We Know About and What We Still Need to Explore. *Medical Care Review* 50(1):49-79, Spring 1993.

Cioffi, M.J., Cleary, P.D., Ding, L., et al.: *Analysis of Case-Mix Strategies and Recommendations for Medicare Managed Care* CAHPS®. HCFA Report. December 1998.

Cleary, P.D. and McNeil, B.J.: Patient Satisfaction as an Indicator of Quality Care. *Inquiry* 25(1):25-36, Spring 1988.

Cleary, P.D., Zaslavsky, A.M., and Cioffi, M.: Sex Differences in Assessments of the Quality of Medicare Managed Care. *Women's Health Issues* 10(2):70-79, 2000.

Cleary, P.D., Edgman-Levitan, S., McMullen, W., and Delbanco, T.L.: The Relationship Between Reported Problems and Patient Summary Evaluations of Hospital Care. *Quality Review Bulletin* 18(2):53-59, February 1992.

Corder, L.S., Woodbury, M.A., and Manton, K.G.: Proxy Response Patterns Among the Aged: Effects on Estimates of Health Status and Medical Care Utilization from the 1982-1984 Long-Term Care Surveys. *Journal of Clinical Epidemiology* 49(2):173-182, 1996.

Epstein, A.M., Hall, J.A., Tognetti, J., et al.: Using Proxies to Evaluate Quality of Life: Can They Provide Valid Information About Patients' Health Status and Satisfaction with Medical Care. *Medical Care* 27(3):S91-S98, March 1989.

Greenley, J.R., Young, T.B., and Schoenherr, R.A.: Psychological Distress and Patient Satisfaction. *Medical Care* 20(4):373-385, April 1982.

Hall, J.A., Feldstein, M., Fretwell, M.D., et al.: Older Patients' Health Status and Satisfaction with Medical Care in an HMO Population. *Medical Care* 28(3):261-269, March 1990.

Hall, J.A., Roter, D.L., Milburn, M.A., and Daltroy, L.H.: Why Are Sicker Patients Less Satisfied with Their Medical Care? Tests of Two Explanatory Models. *Health Psychology* 17(1):70-75, 1998.

Hall, J.A., Milburn, M.A., and Epstein, A.M.: A Causal Model of Health Status and Satisfaction with Medical Care. *Medical Care* 31(1):84-94, January 1993.

Hays, R.D., Vickrey, B.G., Perrine, H.K., et al.: Agreement Between Self-Reports and Proxy Reports of Quality of Life in Epilepsy Patients. *Quality of Life Research* 4(2):159-168, 1995.

Hays, R.D., Shaul, J.A., Williams, V.S.L., et al.: Psychometric Properties of the CAHPS 1.0 Survey Measures. *Medical Care* 37(3) Supp:MS22-MS31, 1999.

Kane, R.L., Maciejewski, M., and Finch, M.: The Relationship of Patient Satisfaction with Care and Clinical Outcomes. *Medical Care* 35(7):714-730, 1997.

Kippen, L.S., Strasser, S., and Joshi, M. : Improving the Quality of the NCQA Annual Member Satisfaction Survey Version 1.0. *American Journal of Managed Care* 3 (5):719-730, 1997.

Lavizzo-Mourey, R., Zinn, J., and Taylor, L.: Ability of Surrogates to Represent Satisfaction of Nursing Home Residents with Quality of Care. *Journal of the American Geriatrics Society* 40(1):39-47, 1992.

Lee, Y., and Kasper, J.D.: Assessments of Medical Care by Elderly People: General Satisfaction and Physician Quality. *Health Services Research* 32(6):741-757, February 1998.

Magaziner, J., Itkin, Z.S., Gruber-Baldini, A.L., et al.: Proxy Reporting in Five Areas of Functional Status: Comparison with Self-Reports and Observations of Performance. *American Journal of Epidemiology* 146(5):418-428, September 1997.

Magaziner, J., Simonsick, E.M., Kashner, T.M., and Hebel, J.R.: Patient-Proxy Response Comparability on Measures of Patient Health Status and Functional Status. *Journal of Clinical Epidemiology* 41(11):1065-1974, 1988.

Magaziner, J., Pear, B.S., Hebel, J.R., and Gruber-Baldini, A.: Use of Proxies to Measure Health and Functional Status in Epidemiologic Studies of Community-Dwelling Women Aged 65 Years and Older. *American Journal of Epidemiology* 143:283-292, 1996.

Marshall, G.N.., Hays, R.D., and Mazel, R.: Health Status and Satisfaction with Health Care: Results from the Medical Outcomes Study. *Journal of Consulting and Clinical Psychology* 64(2):380-390, 1996.

Roberts, R.E., Pascoe, G.C., and Attkisson, C.C.: Relationship of Service Satisfaction to Life Satisfaction and Perceived Well-Being. *Evaluation and Program Planning* 6(3):373-383, 1983.

Rothman, M.L., Hedrick, S.C., Bulcroft, K.A., et al.: The Validity of Proxy Generated Scores as Measures of Patient Health Status. *Medical Care* 29(2):115-124, February 1991.

Rubenstien, L.Z., Schairer, C., Wieland, D., and Kane, R.: Systematic Biases in Functional Status Assessment of Elderly Adults: Effects of Different Data Sources. *Journal of Gerontology* 39(6):686-691, 1984.

Sprangers, M.A.G., and Aaronson, N.K.: The Role of Health Care Providers and Significant Others in Evaluating the Quality of Life of Patients with Chronic Disease: A Review. *Journal of Clinical Epidemiology* 45(7):743-760, 1992.

Ware, J.E., Kosinski, M., and Keller, S.D.: A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. *Medical Care* 34(3):220-233, 1996.

Weinberger, M., (ed.): Consumer Assessment of Health Plans Study (CAHPS®). *Medical Care* 37(3)Supp, 1999.

Weiss, G.L.: Patient Satisfaction with Primary Medical Care: Evaluation of Sociodemographic and Predispositional Factors. *Medical Care* 26(4):383-392, April 1988.

Zapka, J.G., Palmer, R.H., Hargraves, J.L., et al.: Relationships of Patient Satisfaction with Experience of System Performance and Health Status. *Journal of Ambulatory Care Management* 18(1):73-83, 1995.

Zaslavsky, A.M.: Issues in Case-Mix Adjustment of Measures of the Quality of Health Plans. *Proceedings, Government and Social Statistics Sections*. American Statistical Association. Alexandria, VA. 1998.