

Functional Health Outcomes as a Measure of Health Care Quality for Medicare Beneficiaries

Arlene S. Bierman, William F. Lawrence, Samuel C. Haffer, and Carolyn M. Clancy

Objective. The Medicare Health Outcomes Survey (HOS), a new quality measure in the Health Plan Employer Data and Information Set, is designed to assess physical and mental functional health outcomes of Medicare beneficiaries enrolled in Medicare+Choice organizations. We discuss the rationale for the HOS measure together with methodologic challenges in its use and interpretation, using descriptive data from the baseline Medicare HOS to illustrate some of these challenges.

Data Source/Study Design. The 1999 Cohort 2 Medicare HOS baseline data were used for a cross-sectional descriptive analysis. A random sample of 1,000 beneficiaries from each health plan with a Medicare+Choice contract was surveyed ($N = 156,842$; 282 organizations included in these analyses).

Principal Findings. The HOS measure is designed to assess a previously unmeasured dimension of quality. Plan-level variation was seen across all baseline measures of sociodemographic characteristics and illness burden. At the individual level socioeconomic position as measured by educational attainment was strongly associated with functional status. The least educated beneficiaries had the highest burden of illness on all measures examined, and there was a consistent and significant gradient in health and functional status across all levels of education. In analyses stratified by race and ethnicity, socioeconomic gradients in function persisted.

Conclusions. Despite limitations, by focusing attention on the need to improve functional health outcomes among elderly Medicare beneficiaries enrolled in Medicare+Choice, the HOS can serve as an important new tool to support efforts to improve health care quality. The HOS provides valuable information at the federal, state, and health plan levels that can be used to identify, prioritize, and evaluate quality improvement interventions and monitor progress for the program overall as well as for vulnerable subgroups. To interpret the HOS as a quality measure individual- and plan-level differences in functional status and illness burden, as well as methodologic issues in health status measurement, need to be recognized and addressed.

Key Words. Functional status, managed care, Medicare, outcomes measurement, quality of care

A major goal of health care for elderly Medicare beneficiaries is to maintain independence, prevent functional decline, and improve health-related quality of life. However, these outcomes are not routinely assessed to determine whether this goal is achieved. In recent years much attention has been focused on the fact that receipt of health care services of suboptimal quality is a common problem for Medicare beneficiaries (Asch, Sloss, Hogan, et al. 2000; Jencks, Cuerdon, Burwen, et al. 2000). Deficits in quality of care for Medicare beneficiaries have been found in fee-for-service and managed care and in inpatient and ambulatory settings (Asch, Sloss, Hogan, et al. 2000; Ayanian et al. 1999; Jencks, Cuerdon, Burwen, et al. 2000; Seddon, Ayanian, Landrum, et al. 2001). Gaps in quality have been identified in the use of preventive services such as mammography and pneumococcal vaccine, management of chronic conditions such as heart failure and diabetes, management of acute events such as myocardial infarction, and management of common geriatric syndromes such as falls and incontinence (Ayanian et al. 1999; Blustein 1995; Hyman and Pavlik 2001; *Morbidity and Mortality Weekly Report* 2001; Kell, Drass, Bausell, et al. 1999; O'Connor, Quinton, Traven, et al. 1999; Seddon, Ayanian, Landrum, et al. 2001; Soumerai, McLaughlin, Gurwitz, et al. 1999). Improved clinical management in these areas could lead to better functional health outcomes. Moreover, the underuse of effective interventions that can improve functional status is widespread (Boult, Boult, Morishita, et al. 2001; Reuben, Frank, Hirsch, et al. 1999). The availability of a quality measure aimed at assessing functional health outcomes would be an important tool for efforts to assess and improve health care quality.

A recent Institute of Medicine (IOM) report emphasizes that a goal of the U.S. health care system is "to improve the health and functioning of the people of the United States," and asserts that gaps in quality of care are so large as to constitute a "chasm" (IOM 2001). The Centers for Medicare & Medicaid Services

The views expressed here are those of the authors and do not necessarily represent the position of the Agency for Healthcare Research and Quality, the Centers for Medicare & Medicaid Services, or the Department of Health and Human Services.

Address correspondence to Arlene S. Bierman, M.D., M.S., Center for Outcomes and Effectiveness Research, Agency for Healthcare Research and Quality, 6010 Executive Boulevard, Suite 300, Rockville, MD 20852. William F. Lawrence, M.D., is with the Center for Outcomes and Effectiveness Research, Agency for Healthcare Research and Quality. Samuel C. Haffer, Ph.D., is with the Center for Beneficiary Choices, Centers for Medicare & Medicaid Services (formerly HCFA). Carolyn M. Clancy, M.D., is with the Center for Outcomes and Effectiveness Research, Agency for Healthcare Research and Quality. This article, submitted to *Health Services Research* on August 21, 2001, was revised and accepted for publication on October 31, 2001.

(CMS) have a major focus on improving the quality of care purchased by Medicare for beneficiaries in both fee for service and managed care. Through the Health Care Quality Improvement Project, CMS has had some notable successes in quality improvement (CMS 2001; Kiefe, Allison, Williams, et al. 2001; Marciniak, Ellerbeck, Radford, et al. 1998). In 1997 CMS contracted with the National Committee on Quality Assurance (NCQA) to adapt the Health Plan Employer Data and Information Set (HEDIS) for use in Medicare managed care. The complex constellation of chronic illness, comorbidity, and functional impairment among the elderly presents unique challenges in developing quality measures for this population.

Recognition of functional status as an important outcome of care and concern about how managed care would affect this outcome led to the development of the Medicare Health Outcomes Survey (HOS), a new quality measure in Medicare HEDIS (Ware, Bayliss, Rogers, et al. 1996). The HOS is designed to measure change in self-reported function over two years using the summary scores from the Medical Outcomes Survey Short Form 36 (SF-36) (Cooper, Kohlmann, Michael, et al. 2001; Stevic, Haffer, Cooper, et al. 2000). In winter 2002 CMS will publicly release the results of the HOS measure. The case mix-adjusted proportion of enrollees whose two-year change in physical and mental function was worse compared with those whose function was the same or better will be reported. In this article we discuss the rationale for the HOS measure along with methodologic challenges in its use and interpretation, using descriptive data from the baseline Medicare HOS to illustrate some of these issues.

HEALTH OUTCOMES AS A QUALITY MEASURE

Most current activities in quality measurement, including those of Medicare HEDIS, derive from broad interest in the systematic application of scientific evidence to daily practice and build on Donabedian's framework for quality assessment (Donabedian 1980, 1988). This framework emphasizes the relationships among the structure, processes, and outcomes of health care. Historically most quality improvement measures have focused on assessing the structure or processes of care. The past decade, however, has witnessed substantial stakeholder interest in assessing the outcomes, or "end results," of health care services (Clancy and Eisenberg 1998) To be valid, structure and process measures need to be associated with outcomes of care. Conversely, for an outcome measure to be valid it

needs to be associated with the structure and processes of care (Brook, McGlynn, and Cleary 1996).

In addition to using the traditional endpoints of mortality and physiologic measures, health services researchers have developed short, valid tools to assess additional dimensions of health and well-being important to patients. Health perceptions are among the best predictors of future use of services and mortality (Bierman et al. 1999; Connelly, Philbrick, Smith, et al. 1989; Wilson and Kaplan 1995). Health status measures assess the net effect of one or all health problems and treatments on multiple domains of health. Available instruments include measures of overall health as well as assessments of the effect of a single disease entity. The content of items in these instruments, whether generic or disease specific, focuses on standardized reports of an individual's ability to pursue a variety of activities. A generic patient-reported measure needs to consider function across multiple domains of health (Guyatt, Feeny, and Patrick 1993; Patrick and Chiang 2000) to be applicable to a wide variety of health states. The SF-36, a frequently used measure, assesses physical, emotional, role, and social functioning; vitality; bodily pain; and general health perceptions (Ware 1993). The physical component summary (PCS) and mental component summary (MCS) scales are the two summary scales derived from the SF-36 (Ware, Kosinski, Bayliss, et al. 1995). These measures are widely used in outcomes research and are beginning to be applied to quality measurement.

Initial versions of HEDIS relied almost exclusively on structure and process measures. Measures of preventive care predominated (Schneider, Riehl, Courte-Wienecke, et al. 1999). Newer versions of HEDIS have placed increasing emphasis on measures of the quality of care for chronic illness and have added disease-specific outcomes measures. Medicare HEDIS 2001 includes structural measures (e.g., board certification of physicians; arrangements with public health, educational, and social service organizations; availability of language interpretation services) and process measures for preventive services (e.g., mammography screening, influenza and pneumococcal vaccinations) and chronic diseases (e.g., hemoglobin A1C checked, eye exam performed in diabetics, beta blockers prescribed after myocardial infarction). The three disease-specific outcomes measures are all physiologic endpoints (control of hypertension, glucose in diabetics, and hyperlipidemia after a myocardial infarction). The HOS is the first generic outcomes measure to be included in HEDIS.

CRITERIA FOR ADOPTION OF A QUALITY MEASURE

The conceptual framework for assessing quality of care is not new, but there is now an expectation that this information will be publicly reported and provided in a format that allows comparison to inform consumer choice. NCQA has facilitated the development of criteria for measures used for comparative purposes (NCQA 1998). These include relevance, scientific soundness, and feasibility. Relevance includes the prevalence and seriousness of the condition; the type of measure used (outcome vs. process); the significance to one or more stakeholder groups; the financial effect on health systems, providers, or consumers; priorities for using resources to maximize health; the ability of providers to control the dimension of care; potential for improvement; and variation across plans (if there is little variation among providers, the measure will not permit differentiation). Scientific soundness includes clinical evidence linking the measure and outcomes, reproducibility, validity (both clinical and logical sensibility), and accuracy. Measures must not be affected by factors beyond the control of health care providers or those factors must be known and measurable, and the properties of the measure should not be influenced by the use of different data sources. Feasibility addresses measure specification, the burden imposed by data collection, confidentiality of personal health information, data availability, lack of known biases that could influence responses, and potential for independent confirmation. The feasibility of obtaining requisite data is balanced by the strength of payers' interest in the specific measure at hand. Before being adopted into HEDIS the HOS measure was evaluated and approved by the Clinical Performance Measurement Committee of NCQA using these criteria.

FUNCTIONAL STATUS AS A QUALITY MEASURE

A benefit of using functional status as a quality measure is that it captures what is important to the patient. Outcomes measures that reflect the patient's perspective on the success of treatments are especially useful for patients with chronic conditions because traditional measures such as laboratory tests or x-rays provide imprecise information about the effectiveness of treatment and do not enable clinicians and patients to make decisions that involve tradeoffs between length and quality of life. For example, radiographic findings on a knee joint in a patient with osteoarthritis do not matter nearly as much to the patient as the degree to which the person can walk and to which he or she is free from pain (Creamer,

Lethbridge-Cejku, and Hochberg 2000). This focus on what is important to the patient is one of the hallmarks of provision of quality care (IOM 2001).

Generic functional status instruments sacrifice the depth of measurement of biologic information of disease-specific clinical measures for the breadth of covering a wide variety of health domains. Nevertheless, generic measures have been shown to be responsive to changes caused by health care interventions (QualityMetric, Inc. 2000). Generic and disease-specific measures have been found to provide complementary information (Cunningham, Bozzette, Hays, et al. 1995). The SF-36 has shown differences resulting from clinical interventions including coronary angioplasty (Krumholz, McHorney, Clark, et al. 1996), joint-replacement surgery (Fortin, Clarke, Joseph, et al. 1999; Stucki et al. 1995), and pharmacologic therapy for depression (Coulehan, Schulberg, Block, et al. 1997; Revicki, Simon, Chan, et al. 1998). A randomized controlled trial of outpatient comprehensive geriatric assessment coupled with an intervention to promote adherence to recommendations resulted in better SF-36 PCS and MCS scores at 15 months in the intervention group (Reuben, Frank, Hirsch, et al. 1999). Another randomized controlled trial of outpatient geriatric assessment and management found that experimental participants were less likely than controls to lose functional ability (Boult, Boult, Morishita, et al. 2001). In addition, generic instruments allow comparability between different health conditions such as estimates of functional deficits associated with specific conditions. The breadth of domains measured makes generic measures responsive to the presence of multiple health conditions (Lawrence, Fryback, Martin, et al. 1996; Mancuso, Peterson, and Charlson 2000; Mozes, Maor, and Shmueli 1999; Small, Birkett, Meyers, et al. 1996; Stier, Greenfield, Lubeck, et al. 1999; van Manen, Bindels, Dekker, et al. 2001), an advantage in measuring function in people with multiple chronic conditions, the norm among the elderly.

Challenges to Use and Interpretation of Functional Status Measures

Despite the attractiveness of using health and functional status measures as outcomes for quality measurement, there are many challenges in their use and interpretation. Underlying biologic factors, environmental factors such as social and economic supports, and individual factors such as personal motivation all play a role in determining functional status (Kaplan, Kravitz, and Greenfield 2000; Patrick and Chiang 2000; Wilson and Cleary 1995; Wilson and Kaplan 1995). Factors beyond the control of health plans, such as age, education level, and socioeconomic status, have been shown to influence functional status in dia-

betes (Caruso, Silliman, Demissie, et al. 2000) and prostate cancer (Penson, Stoddard, Pasta, et al. 2001). However, these factors have also been associated with the quality of care received (Fiscella et al. 2000). The degree to which health care can mediate these differences remains to be determined. Therefore, while case-mix adjustment that includes demographic factors may facilitate comparisons between plans by controlling for factors outside of a plan's control, it may also mask disparities in health care quality by ethnicity and socioeconomic status. Stratifying the measures by ethnicity and socioeconomic status may overcome this limitation; however, oversampling of under-represented groups may be needed to provide adequate sample size for comparison at the plan level.

In addition, the effect of racial, ethnic, and cultural differences on SF-36 scores are not well-established. The survey is available in numerous languages and has been validated internationally (Fukuhara, Ware, Kosinski, et al. 1998; Lewin-Epstein et al. 1998; Ware, Gandek, Kosinski, et al. 1998). In the United States comparisons between blacks and whites have yielded conflicting information about differences for particular subgroups. In breast cancer survivors (Ashing-Giwa, Ganz, and Petersen 1999) and patients evaluated for acute chest pain (Johnson, Goldman, Orav, et al. 1995) no differences were found between blacks and whites when socioeconomic status was accounted for, yet in a group of indigent blacks and whites some differences were found (Larson, Colangelo, and Goods 1998). A study of Chinese Americans suggested possible differences in question interpretation (Chang et al. 2000). More population-based work needs to be done to better understand how different racial and ethnic groups interpret these measures so that the effect of differences in racial/ethnic compositions between plans on outcomes measures can be examined.

Because the HOS measure is a longitudinal assessment of outcomes it is sensitive to attrition bias. Attrition bias in the HOS measure could occur as a result of disenrollment of beneficiaries, both involuntary and voluntary. Between 1999 and 2001, 1,700,000 Medicare beneficiaries were disenrolled from their managed care plans either because their plan withdrew from participation or reduced its service area (Gold 2001). Reduction in service area might affect the HOS measure in two ways. First, smaller sample size at two years would affect the precision of the estimates of functional change. Second, if beneficiaries in the market areas that were withdrawn were healthier or sicker at baseline than those in remaining areas, this could bias the two-year change scores. In addition, there is evidence that sicker enrollees may disenroll from plans (Morgan et al. 1997). Differences in disenrollment rates are potentially associated with differences in health plan quality. A plan that performs better in caring for and retaining sicker beneficiaries could have

worse two-year change scores than a plan whose sicker members disenrolled. Currently, disenrollees are not surveyed at two years. If two-year data were collected on disenrollees, it would help in both understanding and adjusting for these biases.

Case-mix adjustment is used to account for group-level differences in baseline health status between plans. Residual confounding in case-mix adjustment models is the norm and is likely to present a challenge to the interpretation of HOS results. In measuring health plan quality it is important to recognize that baseline health status of enrollees may reflect quality of care in prior years. Also, because functional impairment is a predictor of subsequent decline the higher the proportion of functionally impaired members in a plan, the greater the likelihood for poorer performance on change scores (Saliba, Elliot, Rubenstein, et al. forthcoming).

In interpreting HOS results and responsiveness of the measures it is necessary to consider effect size or minimally important difference (Guyatt 2000) in SF-36 scores. There may be statistically significant differences in scores, but when do these differences represent meaningful changes in health? Changes in scores with meaningful events have been examined to provide evidence of clinical importance (QualityMetric, Inc. 2000; Ware and Kosinski 2001); for example, having a migraine headache treated improves the PCS score about two points (QualityMetric, Inc. 2000) and aging one year for those over age 65 represents a loss of about one point on PCS (Ware and Kosinski 2001). It is also important to consider whether all domains of health that are important to people are captured. For example, sensory function is not directly measured by the SF-36, so maintenance of visual function in a diabetic through preventive eye care would only be measured through the effect on physical or mental function. In a study of responsiveness to change to interventions for the management of benign prostatic hypertrophy, disease-specific measures were more sensitive to change in symptoms than generic measures (Jenkinson, Gray, Doll, et al. 1997).

Patient preferences are likely to play a role in interpretation of HOS scores. Consider two interventions applied to equivalent groups of patients with the same baseline functional status; after the intervention one group has a PCS of 45 and an MCS of 55, and the other has a PCS of 55 and an MCS of 45. There is no clear answer which group received the best health benefit.

Administration and response issues may also affect scores. While the SF-36 takes an average of ten minutes to complete, people with less than a high school education (McHorney 1996), elderly people (Sherbourne and Meredith 1992), and those with cognitive impairments (McHorney, Teno, Lu, et al. 1990) tend to take longer, increasing the respondent burden and potentially decreasing re-

response rates in these groups. Thus, plans with low response rates could have artificially higher scores because of greater nonresponse of those with lower functional status. Furthermore, elderly respondents are more likely to require assistance or a proxy to respond, or have cognitive impairments that can affect validity (Hayes et al. 1995; McHorney 1996). Mode of administration affects SF-36 scores, with telephone-administered surveys producing higher scores than mail surveys (McHorney, Kosinski, and Ware 1994). These factors may also influence assessment of health plan performance.

WHAT DO BASELINE DATA FROM HOS TELL US?

Selected descriptive data from the HOS 1999 cohort 2 baseline data are reported to illustrate individual- and plan-level differences in case mix. The HOS is a mail survey with telephone follow-up of nonrespondents of a random sample of 1,000 beneficiaries from each health plan with a Medicare+Choice contract. The HOS consists of a 95-item core. The core instrument is composed of the SF-36, a well-tested, valid, and reliable self-report tool for assessing the physical and mental health status of the general population, activities of daily living, chronic conditions, clinical symptoms such as chest pain and shortness of breath, as well as sociodemographic information including race/ethnicity, income, and education. The survey response rate was 66 percent. For individual-level analysis we included all community-dwelling beneficiaries aged 65 and older, excluding individuals enrolled in Evercare or Program of All-inclusive Care for the Elderly (PACE) plans or in hospice care ($N = 156,842$). Age- and gender-adjusted differences in health and functional status measures across five levels of educational attainment are reported. We also report differences in age- and gender-adjusted SF-36 PCS summary scores stratified by race and gender. SUDAAN was used to account for clustering within plans. For plan-level analyses we created variables to indicate the proportion of individuals in a given plan reporting specific demographic and health and functional status characteristics. We report the mean percentage and range across 282 Medicare+Choice plans for these variables.

How Different Are Plans?

Table 1 shows plan-level variation in demographics, chronic disease prevalence, symptoms, SF-36 scores and subscales, and comorbidity for 282 Medicare+Choice

Table 1: Demographics, Illness Burden, and Functional Status Among Medicare+Choice Plans (%)

	<i>Mean</i>	<i>Standard Error</i>	<i>Range</i>
Demographics			
Non-Hispanic White	79	14	11–96
Non-Hispanic Black or African American	7	8	0–60
Hispanic or Spanish	6	8	0–56
Asian or Pacific Islander	1	4	0–49
Income < \$10,000	16	7	6–54
Education < high school	31	9	14–74
Chronic Conditions			
Hypertension	54	4	43–72
Arthritis	49	4	37–67
Diabetes	17	3	10–37
Cardiovascular disease*	36	3	28–49
Congestive heart failure	7	2	3–14
Impairment			
Difficulty hearing	13	3	6–23
Difficulty seeing	6	2	2–18
Symptoms			
Urinary incontinence	27	4	14–39
Felt depressed or sad [†]	13	3	6–34
SF-36 Summary Scores			
PCS	40.8	1.6	35.3–44.8
MCS	52.5	1.1	45.5–55.1

*Reports at least one of the following conditions: ischemic heart disease, stroke, congestive heart failure, other heart condition.

[†]Felt sad or depressed much of the time in the past year.

organizations. For example, whereas the mean proportion of blacks enrolled was 7 percent, the range was 0 percent to 60 percent. The mean number of enrollees at the plan level reporting annual household income less than \$10,000 was 16 percent with a range of 6 percent to 54 percent. Mean PCS and MCS ranged ten points across plans. Large plan-level differences were seen for all measures of illness burden including self-reported chronic conditions such as hypertension, congestive heart failure, and diabetes; the proportion of enrollees reporting visual or hearing impairment or depressed mood; and mean number of comorbid conditions. While the number of plans at the extremes of the range may be low, these plan-level differences in demographics and illness burden are likely to be associated with two-year change scores.

How Different Are Populations?

Table 2 shows age- and gender-adjusted differences in health and functional status stratified by level of educational attainment, a measure of socioeconomic position. The least educated beneficiaries had the highest burden of illness on all measures examined, and there was a consistent and significant gradient in health and functional status across all levels of education. For example, 45.7 percent of beneficiaries with eight years or fewer of formal education, compared to 15.3 percent of beneficiaries who completed college, reported fair or poor health. Medicare beneficiaries need to make complex decisions about insurance coverage and health plan enrollment, make informed decisions about a growing array of therapeutic interventions, and become active participants in their care. All of these choices can ultimately affect health outcomes. Quality improvement interventions informed by knowledge of group-level differences in education and health literacy will need to be tailored to communicate these choices effectively.

Table 3 shows differences in SF-36 PCS summary scores stratified by income, race, and education. The population means represent Medicare beneficiaries enrolled in Medicare+Choice organizations, not the general U.S. elderly population, and may reflect selection bias in managed care enrollment as well as cultural differences in response. On average Asians had the highest mean SF-36 PCS scores (43.1), followed by whites (41.0) and Hispanics (40.5). The lowest mean PCS scores were seen among blacks (38.2). A gradient in PCS scores was seen across all levels of income and education for each ethnic group. For example, there was a seven-point difference in mean PCS scores for whites and a six-point difference for blacks between the lowest and highest levels of income. Similar differences were seen for education. Plans that serve socioeconomically disadvantaged beneficiaries are therefore likely to have a worse case mix. The HOS data have the potential to provide insights into the complex interactions among race, ethnicity, income and education, and health care delivery and health outcomes. They will also provide the opportunity to assess the effectiveness of quality improvement interventions for the most vulnerable beneficiaries.

Table 2: Education, Health Status, and Self-reported Chronic Disease Prevalence of Medicare Beneficiaries Aged 65 and Older in Medicare+Choice*

	<i>8th Grade</i> n = 20,191	<i>< High School</i> n = 28,478	<i>High School Graduate</i> n = 55,133	<i>Some College</i> n = 31,843	<i>College Graduate</i> n = 21,197
Mean SF-36 PCS (standard error)	37.8 (0.08)	39.2 (0.07)	41.0 (0.05)	41.9 (0.07)	44.1 (0.08)
Mean SF-36 MCS (standard error)	49.6 (0.07)	51.2 (0.06)	52.9 (0.04)	53.7 (0.06)	54.7 (0.07)
Fair/poor health	45.7	36.1	25.9	21.0	15.3
<i>Chronic Conditions (%)</i>					
Hypertension	57.9	57.4	53.7	51.5	47.2
Arthritis	53.5	49.9	46.8	47.3	43.3
Diabetes	23.2	20.1	16.6	15.1	12.3
Cardiovascular disease	21.5	22.8	19.9	20.2	16.7
Congestive heart failure	9.6	8.8	6.9	6.5	5.0
<i>Symptoms (%)</i>					
Urinary incontinence	28.5	27.8	26.1	27.1	25.5
Felt depressed or sad [†]	21.9	16.1	10.9	9.4	6.8
Back pain [‡]	17.3	13.5	9.8	9.1	6.4
Shortness of breath [§]	15.8	11.6	8.0	7.1	4.5
Chest pain with exercise [¶]	6.0	3.9	2.6	2.4	1.5
<i>Sensory Impairment (%)</i>					
Difficulty hearing	21.5	16.4	12.2	9.9	6.9
Difficulty seeing	11.4	6.8	5.2	4.3	3.8
<i>Activities of Daily Living Limitations (%)</i>					
Any difficulty bathing	21.3	15.6	12.5	11.2	9.0
Any difficulty dressing	18.0	12.6	10.4	9.7	7.8
<i>Comorbidity (%)</i>					
Three or more chronic conditions	58.7	54.8	48.6	48.5	41.8
Mean no. of comorbid conditions	3.3	3.0	2.7	2.7	2.4

*Adjusted for age and gender. For all variables, chi-square test of independence $p < .0001$.

[†]Felt sad or depressed much of the time in the past year.

[‡]Low back pain interferes with usual daily activities all or most of the time.

[§]Shortness of breath all or most of the time when walking less than one block.

[¶]All or most of the time.

Table 3: Income, Education, and Mean SF-36 PCS Scores of Elderly Medicare Beneficiaries in Medicare+Choice*

	<i>Non-Hispanic White</i> n = 129,523	<i>Non-Hispanic Black or African American</i> n = 9,220	<i>Hispanic or Spanish</i> n = 8,051	<i>Asian or Pacific Islander</i> n = 2,486
Overall	41.0	38.2	40.5	43.1
<i>Income</i>				
Poor (< \$10k)	38.1	36.7	38.6	42.0
Low (\$10k–20k)	39.4	38.0	39.3	41.2
Middle (\$20k–30k)	41.1	40.3	41.8	43.3
Upper middle (\$30k–50k)	42.9	42.0	43.9	43.7
High (> \$50k)	45.1	42.5	44.1	45.3
<i>Education</i>				
< 8th grade	37.9	36.3	38.7	41.0
< high school	39.3	37.7	40.3	42.7
High school graduate	41.1	39.6	41.6	43.5
Some college	41.9	40.6	42.6	44.1
College graduate	44.2	42.7	44.1	43.6

*Adjusted for age and gender.

DISCUSSION

Quality of care is a multidimensional construct, and measures that capture these multiple dimensions are necessary. The HOS measure assesses an important dimension of quality of care for Medicare beneficiaries, the effect of health care on functional status. The HOS provides valuable new health status information on beneficiaries enrolled in Medicare+Choice including the health needs of high-risk Medicare beneficiaries. The HOS measure may serve to focus attention on the need to develop evidence-based strategies to improve functional health outcomes in the Medicare population. The focus of this article is quality measurement. However, health and functional status information on enrollees can also be used to inform the design of health services and risk-adjustment strategies for payment, although HOS data are not being used for this purpose.

Many questions about the HOS measure remain unanswered. Can functional health outcomes be improved? Growing evidence from the literature suggests so. Will the SF-36 be sufficiently responsive to change in functional status? How will the HOS measure perform in different populations? How will this influence results? How should case mix be adjusted for? The HOS data provide an unprecedented look at the functional status of beneficiaries in Medicare+Choice

and the evidence to begin to answer these questions. It is unclear how current uncertainty and instability in the Medicare+Choice market will affect the application of the HOS measure.

Working with the managed care industry and other partners, CMS plans to use the HOS data in three distinct ways: to provide comparative information for beneficiaries to make informed decisions when choosing a health plan, encourage competition-based quality improvement, and assess the performance of health plans by integrating valid and reliable performance measures into the accreditation scoring process. Peer-review organizations use results from the HOS as an additional data source to assist in monitoring the quality of care provided at Medicare+Choice organizations. They may initiate quality improvement projects with Medicare+Choice organizations, collaborate with organizations that approach them with quality improvement project ideas using HOS data, and provide technical assistance to health plans in interpreting their plan-specific data. Additionally, peer-review organizations in Arizona, Florida, New York, Ohio, and Michigan have partnered with interested health plans in each state in a national pilot project that tests the utility of using HOS measures to help identify cohorts of potentially depressed patients and improve the recognition and treatment of depression in primary care.

CMS is exploring the feasibility of collecting similar information for traditional fee-for-service Medicare. A pilot project begun in 1998 explored the methodologic and technical issues surrounding the implementation of the HOS in fee-for-service Medicare. The HOS instrument and protocol were tested in a sample of beneficiaries at the group-practice level, small-area level, as well as within a national random sample. Results of this project are now being analyzed.

Brook, McGlynn, and Cleary (1996) underscore that, "It will never be possible to produce an error-free measure of quality of care." The HOS is one quality measure among many measures of quality included in HEDIS. By focusing attention on the need to design interventions to improve functional health outcomes the HOS is an important new tool for efforts to improve quality of care for the Medicare program. Health plans can use HOS data to target interventions to improve functional status. For example, improved pain management among individuals with arthritis, better management of congestive heart failure, or diagnosis and management of urinary incontinence are all interventions that could potentially improve functional health outcomes. Analysis of HOS longitudinal data may provide insights on the relative contribution of patient-level factors, such as socioeconomic status, and plan characteristics to functional status. The HOS will also provide valuable information at the federal (CMS) and state (peer-review organi-

zations) levels to help identify and prioritize quality improvement interventions. New models of health care delivery will be needed to improve functional status among Medicare beneficiaries, and these models need to address differences in health and functional status among different populations of elderly beneficiaries. The HOS can be considered a "first-generation" generic quality measure of health status to support these efforts.

REFERENCES

- Asch, S. M., E. M. Sloss, C. Hogan, R. H. Brook, and R. L. Kravitz. 2000. "Measuring Underuse of Necessary Care Among Elderly Medicare Beneficiaries Using Inpatient and Outpatient Claims." *Journal of the American Medical Association* 284 (18): 2325-33.
- Ashing-Giwa, K., P. A. Garz, and L. Petersen. 1999. "Quality of Life of African-American and White Long Term Breast Carcinoma Survivors." *Cancer* 85 (2): 418-26.
- Ayanian, J. Z., J. S. Weissman, S. Chasan-Taber, and A. M. Epstein. 1999. "Quality of Care by Race and Gender for Congestive Heart Failure and Pneumonia." *Medical Care* 37 (12): 1260-69.
- Bierman, A. S., T. A. Bubolz, E. S. Fisher, and J. H. Wasson. 1999. "How Well Does a Single Question About Health Predict the Financial Health of Medicare Managed Care Plans?" *Effective Clinical Practice* 2 (2): 56-62.
- Blustein, J. 1995. "Medicare Coverage, Supplemental Insurance, and the Use of Mammography by Older Women." *New England Journal of Medicine* 332 (17): 1138-43.
- Boult, C., L. B. Boult, L. Morishita, B. Dowd, R. L. Kane, and C. F. Urdangarin. 2001. "A Randomized Clinical Trial of Outpatient Geriatric Evaluation and Management." *Journal of the American Geriatric Society* 49 (4): 351-59.
- Brook, R. H., E. A. McGlynn, and P. D. Cleary. 1996. "Quality of Health Care. Part 2: Measuring Quality of Care." *New England Journal of Medicine* 335 (13): 966-70.
- Caruso, L. B., R. A. Silliman, S. Demissie, S. Greenfield, and E. H. Wagner. 2000. "What Can We Do to Improve Physical Function in Older Persons with Type 2 Diabetes?" *The Journal of Gerontology Series A, Biological Sciences and Medical Sciences* 55 (7): M372-77.
- Centers for Medicare and Medicaid Services. 2001. "Medicare Priorities." [Online report of the Health Care Quality Improvement Project. <http://www.hcfa.gov/quality/download/3k1.pdf>.]
- Chang, D. F., C. A. Chun, D. T. Takeuchi, and H. Shen. 2000. "SF-36 Health Survey: Tests of Data Quality, Scaling Assumptions, and Reliability in a Community Sample of Chinese Americans." *Medical Care* 38 (5): 542-48.
- Clancy, C. M., and J. M. Eisenberg. 1998. "Outcomes Research: Measuring the End Results of Health Care." *Science* 282 (5387): 245-46.
- Connelly, J. E., J. T. Philbrick, G. R. Smith, Jr., D. L. Kaiser, and A. Wymer. 1989. "Health Perceptions of Primary Care Patients and the Influence on Health Care Utilization." *Medical Care* 27 (Suppl. 3): S99-109.

- Cooper, J. K., T. Kohlmann, J. A. Michael, S. C. Haffer, and M. Stevic. 2001. "Health Outcomes. New Quality Measure for Medicare." *International Journal of Quality in Health Care* 13 (1): 9–16.
- Coulehan, J. L., H. C. Schulberg, M. R. Block, M. J. Madonia, and E. Rodriguez. 1997. "Treating Depressed Primary Care Patients Improves Their Physical, Mental, and Social Functioning." *Archives of Internal Medicine* 157 (10): 1113–20.
- Creamer, P., M. Lethbridge-Cejku, and M. C. Hochberg. 2000. "Factors Associated with Functional Impairment in Symptomatic Knee Osteoarthritis." *Rheumatology (Oxford)* 39 (5): 490–96.
- Cunningham, W. E., S. A. Bozzette, R. D. Hays, D. E. Kanouse, and M. F. Shapiro. 1995. "Comparison of Health-related Quality of Life in Clinical Trial and Nonclinical Trial Human Immunodeficiency Virus-infected Cohorts." *Medical Care* 33 (Suppl. 4): AS15–25.
- Donabedian, A. 1980. *Explorations in Quality Assessment and Monitoring*. Vol. 1. Chicago: Health Administration Press.
- Donabedian, A. 1988. "The Quality of Care. How Can It be Assessed?" *Journal of the American Medical Association* 260 (12): 1743–48.
- Fiscella, K., P. Franks, M. R. Gold, and C. M. Clancy. 2000. "Inequality in Quality: Addressing Socioeconomic, Racial, and Ethnic Disparities in Health Care." *Journal of the American Medical Association* 283 (19): 2579–84.
- Fortin, P. R., A. E. Clarke, L. Joseph, M. H. Liang, M. Tanzer, D. Ferland, C. Phillips, A. J. Partridge, P. Belisle, A. H. Fossel, N. Mahomed, C. B. Sledge, and J. N. Katz. 1999. "Outcomes of Total Hip and Knee Replacement: Preoperative Functional Status Predicts Outcomes at Six Months After Surgery." *Arthritis and Rheumatism* 42 (8): 1722–28.
- Fukuhara, S., J. E. Ware, Jr., M. Kosinski, S. Wada, and B. Gandek. 1998. "Psychometric and Clinical Tests of Validity of the Japanese SF-36 Health Survey." *Journal of Clinical Epidemiology* 51 (11): 1045–53.
- Gold, M. 2001. "Medicare+Choice: An Interim Report Card." *Health Affairs* 20 (4): 120–38.
- Guyatt, G. H. 2000. "Making Sense of Quality-of-life Data." *Medical Care* 38 (Suppl. 9): II175–79.
- Guyatt, G. H., D. H. Feeny, and D. L. Patrick. 1993. "Measuring Health-related Quality of Life." *Annals of Internal Medicine* 118 (8): 622–29.
- Hayes, V., J. Morris, C. Wolfe, and M. Morgan. 1995. "The SF-36 Health Survey Questionnaire: Is it Suitable for Use with Older Adults?" *Age and Aging* 24 (2): 120–25.
- Hyman, D. J., and V. N. Pavlik. 2001. "Characteristics of Patients with Uncontrolled Hypertension in the United States." *New England Journal of Medicine* 345 (7): 479–86.
- Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.
- Jencks, S. F., T. Cuerdon, D. R. Burwen, B. Fleming, P. M. Houck, A. E. Kussmaul, D. S. Nilasena, D. L. Ordin, and D. R. Arday. 2000. "Quality of Medical Care Delivered to Medicare Beneficiaries: A Profile at State and National Levels." *Journal of the American Medical Association* 284 (13): 1670–76.

- Jenkinson, C., A. Gray, H. Doll, K. Lawrence, S. Keoghane, and R. Layte. 1997. "Evaluation of Index and Profile Measures of Health Status in a Randomized Controlled Trial. Comparison of the Medical Outcomes Study 36-item Short-form Health Survey, EuroQol, and Disease Specific Measures." *Medical Care* 35 (11): 1109-18.
- Johnson, P. A., L. Goldman, E. J. Orav, T. Garcia, S. D. Pearson, and T. H. Lee. 1995. "Comparison of the Medical Outcomes Study Short-form 36-item Health Survey in Black Patients and White Patients with Acute Chest Pain." *Medical Care* 33 (2): 145-60.
- Kaplan, S. H., R. L. Kravitz, and S. Greenfield. 2000. "A Critique of Current Uses of Health Status for the Assessment of Treatment Effectiveness and Quality of Care." *Medical Care* 38 (Suppl. 9): II184-91.
- Kell, S. H., J. Drass, R. B. Bausell, K. A. Thomas, M. A. Osborn, and D. Gohdes. 1999. "Measures of Disease Control in Medicare Beneficiaries with Diabetes Mellitus." *Journal of the American Geriatric Society* 47 (4): 417-22.
- Kiefe, C. I., J. J. Allison, O. D. Williams, S. D. Person, M. T. Weaver, and N. W. Weissman. 2001. "Improving Quality Improvement Using Achievable Benchmarks for Physician Feedback: A Randomized Controlled Trial." *Journal of the American Medical Association* 285 (22): 2871-79.
- Krumholz, H. M., C. A. McHorney, L. Clark, M. Levesque, D. S. Baim, and L. Goldman. 1996. "Changes in Health After Elective Percutaneous Coronary Revascularization. A Comparison of Generic and Specific Measures." *Medical Care* 34 (8): 754-59.
- Larson, C. O., M. Colangelo, and K. Goods. 1998. "Black-white Differences in Health Perceptions Among the Indigent." *Journal of Ambulatory Care Management* 21 (1): 35-43.
- Lawrence, W. F., D. G. Fryback, P. A. Martin, R. Klein, and B. E. Klein. 1996. "Health Status and Hypertension: A Population-based Study." *Journal of Clinical Epidemiology* 49 (11): 1239-45.
- Lewin-Epstein, N., T. Sagiv-Schifter, E. L. Shabtai, and A. Shmueli. 1998. "Validation of the 36-item Short-form Health Survey (Hebrew Version) in the Adult Population of Israel." *Medical Care* 36 (9): 1361-70.
- Mancuso, C. A., M. G. Peterson, and M. E. Charlson. 2000. "Effects of Depressive Symptoms on Health-related Quality of Life in Asthma Patients." *Journal of General Internal Medicine* 15 (5): 301-310.
- Marciniak, T. A., E. F. Ellerbeck, M. J. Radford, T. F. Kresowik, J. A. Gold, H. M. Krumholz, C. I. Kiefe, R. M. Allman, R. A. Vogel, and S. F. Jencks. 1998. "Improving the Quality of Care for Medicare Patients with Acute Myocardial Infarction: Results from the Cooperative Cardiovascular Project." *Journal of the American Medical Association* 279 (17): 1351-57.
- McHorney, C. A. 1996. "Measuring and Monitoring General Health Status in Elderly Persons: Practical and Methodological Issues in Using the SF-36 Health Survey." *Gerontologist* 36 (5): 571-83.
- McHorney, C. A., M. Kosinski, and J. E. Ware, Jr. 1994. "Comparisons of the Costs and Quality of Norms for the SF-36 Health Survey Collected by Mail Versus Telephone Interview: Results from a National Survey." *Medical Care* 32 (6): 551-67.
- McHorney, C. A., J. Teno, J. Lu, C. D. Sherbourne, and J. E. Ware, Jr. 1990. "The Use of Standardized Measures of Functional Status and Well-being Among Cognitively Impaired and Intact Elders: Results from the Medical Outcomes Study." Presented at the 43rd Annual Scientific Meeting of the Gerontological Society of America, Boston.

- Morbidity and Mortality Weekly Report. 2001. "Influenza and Pneumococcal Vaccination Levels Among Persons Aged 65 Years—United States, 1999." *Morbidity and Mortality Weekly Report* 50 (25): 532–37.
- Morgan, R. O., B. A. Virnig, C. A. DeVito, and N. A. Persily. 1997. "The Medicare-HMO Revolving Door—The Healthy Go in and the Sick Go out." *New England Journal of Medicine* 337 (3): 169–75.
- Mozes, B., Y. Maor, and A. Shmueli. 1999. "The Competing Effects of Disease States on Quality of Life of the Elderly: The Case of Urinary Symptoms in Men." *Quality of Life Research* 8 (1-2): 93–99.
- O'Connor, G. T., H. B. Quinton, N. D. Traven, L. D. Ramunno, T. A. Dodds, T. A. Marciniak, and J. E. Wernberg. 1999. "Geographic Variation in the Treatment of Acute Myocardial Infarction: The Cooperative Cardiovascular Project." *Journal of the American Medical Association* 281 (7): 627–33.
- Patrick, D. L., and Y. P. Chiang. 2000. "Measurement of Health Outcomes in Treatment Effectiveness Evaluations: Conceptual and Methodological Challenges." *Medical Care* 38 (Suppl. 9): II14–25.
- Person, D. F., M. L. Stoddard, D. J. Pasta, D. P. Lubeck, S. C. Flanders, and M. S. Litwin. 2001. "The Association Between Socioeconomic Status, Health Insurance Coverage, and Quality of Life in Men with Prostate Cancer." *Journal of Clinical Epidemiology* 54 (4): 350–58.
- QualityMetric, Inc. 2000. "The SF-36 Health Survey: A Summary of Responsiveness to Clinical Interventions." Report prepared for The Health Assessment Lab and the National Committee for Quality Assurance under contract No. 500-97-P001, entitled *Implementing HEDIS in Medicare Managed Care*, sponsored by the Health Care Financing Administration, U.S. Department of Health and Human Services. Lincoln, RI: QualityMetric, Inc.
- Reuben, D. B., J. C. Frank, S. H. Hirsch, K. A. McGuigan, and R. C. Maly. 1999. "A Randomized Clinical Trial of Outpatient Comprehensive Geriatric Assessment Coupled with an Intervention to Increase Adherence to Recommendations." *Journal of the American Geriatric Society* 47 (3): 269–76.
- Revicki, D. A., G. E. Simon, K. Chan, W. Katon, and J. Heiligenstein. 1998. "Depression, Health-related Quality of life, and Medical Cost Outcomes of Receiving Recommended Levels of Antidepressant Treatment." *Journal of Family Practice* 47 (6): 446–52.
- Saliba, D., M. Elliot, L. Z. Rubenstein, D. H. Solomon, R. T. Young, C. Y. Kamberg, et al. Forthcoming. "The Vulnerable Elders Survey (VES-13): A Tool for Identifying Vulnerable Elders in the Community." *Journal of the American Geriatric Society* (in press).
- Schneider, E. C., V. Riehl, S. Courte-Wienecke, D. M. Eddy, and C. Sennett, National Committee for Quality Assurance. 1999. "Enhancing Performance Measurement: NCQA's Road Map for a Health Information Framework." *Journal of the American Medical Association* 282 (12): 1184–90.
- Seddon, M. E., J. Z. Ayanian, M. B. Landrum, P. D. Cleary, E. A. Peterson, M. T. Gahart, and B. J. McNeil. 2001. "Quality of Ambulatory Care After Myocardial Infarction Among Medicare Patients by Type of Insurance and Region." *American Journal of Medicine* 111 (1): 24–32.

- Sherbourne, C. D., and L. S. Meredith. 1992. "Quality of Self-report Data: A Comparison of Older and Younger Chronically Ill Patients." *Journal of Gerontology* 47 (4): S204-211.
- Small, G. W., M. Birkett, B. S. Meyers, L. M. Koran, A. Bystritsky, and C. B. Nemeroff, Fluoxetine Collaborative Study Group. 1996. "Impact of Physical Illness on Quality of Life and Antidepressant Response in Geriatric Major Depression." *Journal of the American Geriatric Society* 44 (10): 1220-25.
- Soumerai, S. B., T. J. McLaughlin, J. H. Gurwitz, S. Pearson, C. L. Christiansen, C. Borbas, N. Morris, B. McLaughlin, X. Gao, and D. Ross-Degnan. 1999. "Timeliness and Quality of Care for Elderly Patients with Acute Myocardial Infarction Under Health Maintenance Organization Vs Fee-for-service Insurance." *Archives of Internal Medicine* 159 (17): 2013-20.
- Stevic, M., S. C. Haffer, J. K. Cooper, R. W. Adams, and J. A. Michael. 2000. "How Healthy Are Our Seniors? Baseline Results from the Medicare Health Outcomes Survey." *Journal of Clinical Outcomes Management* 7 (8): 39-42.
- Stier, D. M., S. Greenfield, D. P. Lubeck, K. A. Dukes, S. C. Flanders, J. M. Henning, J. Weir, and S. H. Kaplan. 1999. "Quantifying Comorbidity in a Disease-specific Cohort: Adaptation of the Total Illness Burden Index to Prostate Cancer." *Urology* 54 (3): 424-29.
- Stucki, G., M. H. Liang, C. Phillips, and J. N. Katz. 1995. "The Short Form-36 Is Preferable to the SIP as a Generic Health Status Measure in Patients Undergoing Elective Total Hip Arthroplasty." *Arthritis Care Research* 8 (3): 174-81.
- U. S. National Committee on Quality Assurance. 1998. "Desirable Attributes of HEDIS Measures." [Online report. <http://www.ncqa.org/Programs/HEDIS/desirable%20attributes.html>.]
- van Manen, J. G., P. J. Bindels, E. W. Dekker, C. J. Ijzermans, B. J. Bottema, J. S. van der Zee, and E. Schade. 2001. "Added Value of Co-morbidity in Predicting Health-related Quality of Life in COPD Patients." *Respiratory Medicine* 95 (6): 496-504.
- Ware, J. E., Jr. 1993. *SF-36 Health Survey: Manual and Interpretation Guide* Boston: The Health Institute.
- Ware, J. E., M. S. Bayliss, W. H. Rogers, M. Kosinski, and A. R. Tarlov. 1996. "Differences in 4-year Health Outcomes for Elderly and Poor, Chronically Ill Patients Treated in HMO and Fee-for-service Systems. Results from the Medical Outcomes Study." *Journal of the American Medical Association* 276 (13): 1039-47.
- Ware, J. E., Jr., B. Gandek, M. Kosinski, N. K. Aaronson, G. Apolone, J. Brazier, M. Bullinger, S. Kaasa, A. Leplege, L. Prieto, M. Sullivan, and K. Thunedborg. 1998. "The Equivalence of SF-36 Summary Health Scores Estimated Using Standard and Country-specific Algorithms in 10 Countries: Results from the IQOLA Project. International Quality of Life Assessment." *Journal of Clinical Epidemiology* 51 (11): 1167-70.
- Ware, J. E., Jr., and M. Kosinski. 2001. *SF-36: Physical and Mental Health Summary Scales: A Manual for Users of Version 1*, 2nd edition. Lincoln, RI: QualityMetric, Inc.
- Ware, J. E., Jr., M. Kosinski, M. S. Bayliss, C. A. McHorney, W. H. Rogers, and A. Raczek. 1995. "Comparison of Methods for the Scoring and Statistical Analysis of SF-36 Health Profile and Summary Measures: Summary of Results from the Medical Outcomes Study." *Medical Care* 33 (Suppl. 4): AS264-79.

- Wilson, I. B., and P. D. Cleary. 1995. "Linking Clinical Variables with Health-related Quality of Life. A Conceptual Model of Patient Outcomes." *Journal of the American Medical Association* 273 (1): 59-65.
- Wilson, I. B., and S. Kaplan. 1995. "Clinical Practice and Patients' Health Status: How Are the Two Related?" *Medical Care* 33 (Suppl. 4): AS209-214.