A Comparison of Three Health Status Measures in Primary Care Outpatients

David Edelman, MD, G. Rhys Williams, MS, Margaret Rothman, PhD, Gregory P. Samsa, PhD

Our objective was to compare a brief, relatively new global health status measure, the Health Utilities Index Mark II (HUI), to two commonly applied health status measures (Medical Outcomes Study 36-Item Short-Form Health Survey [SF-36] and the Sickness Impact Profile [SIP] in a general medical outpatient population. Using a cross-sectional survey, we surveyed 160 patients in the General Medical Clinic of the Durham Veterans Affairs Medical Center. Each subject answered demographic questions and then completed the three health status measures. The mean time taken to complete the measures was 3, 10, and 20 minutes for the HUI, SF-36, and SIP, respectively ($p < .0001$). The HUI exhibited a modest "floor" effect; that is, scores were concentrated near the sicker of the scale. In contrast, responses to the SIP were heavily concentrated near the healthier end of the scale. Spearman correlation coefficients between the HUI and scales within the other two measures ranged from .54 (SF-36 mental health) to 0.69 (SF-36 physical functioning). Subjects accepted all measures well. These three health service measures varied in their distribution of responses and time required to complete. Users should consider the degree of sickness of the population to be assessed when choosing a measure.

KEY WORDS: quality of life; Health Utilities Index Mark II; health status measure; ambulatory care.

J GEN INTERN MED 1999;14:759-762.

Health-related quality of life (HRQOL) is an important patient outcome for many conditions and treatments. The instruments used to assess HRQOL are often referred to as health status measures. While HRQOL has mostly been assessed as an outcome in clinical research,^{1,2} use of health status measures has spread to clinical practice³ and public health surveillance.⁴⁻⁶ Other proposed uses include case-mix adjustment in the managed care setting.^{7,8} In addition to the widening number of uses for health status measures, there are a variety of health status measures available. It is important for diverse types of users to be able to make intelligent choices among the instruments based on the users' own needs.

Our objectives in this study were to compare the score distributions and convenience of administration among

three health status measures applied in a general medical clinic. We compared two well-studied measures, the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) and the Sickness Impact Profile (SIP), to the Health Utilities Index Mark II (HUI), a new and brief measure.

METHODS

Patients

We performed a cross-sectional study in the General Medicine Clinic in the Durham Veterans' Affairs Medical Center (VAMC). We interviewed a convenience sample of 173 patients on consecutive clinic days in June and July of 1997. A research assistant experienced in HRQOL questionnaire administration identified a convenience sample of three to eight patients per day to become subjects in the study. Halfway through the protocol, in order to increase completion rates, the research assistant began to contact potential subjects by telephone 4 to 7 days in advance of their appointment. She asked patients so contacted to arrive at the clinic 1 hour prior to their scheduled appointment to complete the study questionnaires. Those patients who did this were reimbursed \$25 for completion of all questionnaires. All patients capable of responding verbally to simple questions were eligible. Patients were excluded only if they declined participation.

Questionnaires and Scoring

The SF-36 is a 36-item questionnaire developed and validated in a cohort of medical outpatients.⁹⁻¹³ The questionnaire is divided into eight separate subscales; commonly used subscales with well-documented construct validity are physical functioning, role functioning (physical), role functioning (emotional), pain, energy, general health perceptions, social functioning, and mental health. For all scales, higher scores reflect better health. Scores are scaled from 0 to 100.

The SIP and its development are also described elsewhere.¹⁴ Briefly, the SIP was derived theoretically and empirically validated in a population of mixed outpatients, inpatients, and institution dwellers. In its full form, the SIP is a 136-item questionnaire that is scored into 12 subscales. Several of these can be combined into two domains, physical and psychological; there is also a single "overall" score.^{14,15} All scales are scored 0 to 100, with 0 being perfect health.

The HUI and its development are documented.^{16,17} Briefly, the HUI was developed to be used in evaluating HRQOL in children with leukemia, using their parents as proxy respondents. It has been used in adults in a variety of settings, including the assessment of the effect of chronic

Received from the Center for Health Services Research in Primary Care, Durham Veterans Affairs Medical Center, and Division of General Internal Medicine, Duke University Medical Center, Durham, NC (DE); Division of Outcomes Research, Janssen Pharmaceuticals, Titusville, NJ (GRW, MR); and The Center for Clinical Health Policy Research, Durham NC (GPS).

Address correspondence and reprint requests to Dr. Edelman: Durham Veterans Affairs Medical Center, Health Services Research (152), 508 Fulton St., Durham, NC 27705.

outpatient otolaryngological problems on HRQOL and the measurement of HRQOL in patients who have recently suffered a cerebrovascular accident.^{18,19} It is a 7-item multiattribute scale that yields a single number when scored. This number ranges from 0 to 1, with 1 being perfect health. Because our study population consisted of older, chronically ill adults, we omitted the "fertility" attribute, leaving a 6-item multiattribute scale. Table 1 shows the HUI.

Administration of Measures

After obtaining written informed consent, the research assistant administered all questionnaires to all subjects. The questionnaires were administered in a random order to avoid any bias arising from consistently administering one questionnaire before the others. Subjects also completed a brief demographic survey. Subjects were asked at the end of each questionnaire to rate the difficulty of completing the questionnaire on a 5-point Likert scale. The research assistant measured the time required to complete each questionnaire. Approximately 15% of the measures were self-administered.

Analysis

We compared the HUI to the physical functioning and mental health subscales of the SF-36 and the physical

and psychological domains of the SIP. We also compared the HUI to the overall SIP score. Because scores were not distributed normally, we calculated with Spearman coefficients. We compared time to completion of the three measures by using analysis of variance (ANOVA). Likert scale data were presented descriptively. We performed all analyses using the SAS statistical analysis system (SAS Institute, Cary, NC).

RESULTS

We approached 209 patients for the study. Four refused, and 32 did not attempt one or more of the health status measures due to time constraints. Thirteen patients omitted questions on one or more of the measures, leaving a final sample size of 160 patients.

Respondents' median age was 60 (Interquartile Range 49–71, rang 27–100). Twenty-two percent of respondents were smokers, 55% reported having hypertension, 39% reported having heart disease, and 17% reported a history of strokes or transient cerebral ischemia. Almost all respondents (96%) were male.

The distributions of questionnaire scores are presented in Figure 1. The median SF-36 physical functioning score was 55. The median SF-36 mental health score was 76. The median SIP physical scale score was 5.8, and the distribution was skewed toward better health—i.e., many scores

Table 1. The Health Utilities Index, Mark II (HUI)

Attribute*	Description	Score
Sensation	Normal	1
	Requires aid to see or hear or speak	2
	Limited sight, hearing or speech after correction	3
	Blind, deaf or mute	4
Mobility	Walks, bends, lifts, jumps, and runs normally	1
	Limitations in above but moves independently	2
	Can move with equipment (cane, walker) but without another person's help	3
	Requires help from another person to get around	4
	Unable to use arms and legs; bedbound	5
Emotion	Generally happy	1
	Occasionally angry, irritable, anxious, depressed, or suffers nightmares	2
	Often angry, irritable, anxious, depressed, or suffers nightmares	3
	Almost always angry, irritable, anxious, depressed, or suffers nightmares	4
	Usually in psychiatric institution	5
Cognition	Normal	1
	Memory impaired as judged by companions	2
	Memory severely impaired, requires constant support	3
	Unable to retain any new information	4
Self-care	Normal eating, bathing, dressing, toileting	1
	Independent eating, bathing, dressing, toileting but with difficulty	2
	Requires equipment to eat, bathe, dress, toilet	3
	Requires another person to eat, bathe, dress, toilet	4
Pain	No pain	1
	Occasional pain, relieved by over-the-counter medicines, no disruption of usual activities	2
	Frequent pain, not on narcotics, occasional disruption of usual activities	3
	Frequent pain, on narcotics or near complete disruption of usual activities	4
	Constant pain, not relieved by medication	5

*The fertility scale has been removed for our patient population.

reflected the best health the scale could measure. The SIP psychological scale is also skewed toward absence of illness, with a median of 8.7. The median HUI score was 0.29, and the distribution was skewed toward poorer health—i.e., many scores reflected the poorest health the scale could measure. The HUI scores were moderately correlated with SF-36 and SIP scores, with Spearman ρ ranging from .69 (SF-36 physical functioning) to .54 (SF-36 mental health).

The HUI took an average of 3 minutes to complete, the SF-36 took 10 minutes, and the SIP took 20 minutes ($p < .0001$). The difficulty, as measured on a 5-point Likert scale (1 = no difficulty, 5 = too difficult to complete) was 1.28, 1.45, and 1.39 for the HUI, SF-36, and SIP, respectively ($p = .20$).

DISCUSSION

The distributions of responses to the three health status measures were noticeably different. The spectrum of distributions seen in our study replicates the experience of

clinicians when they ask questions regarding quality of life while taking a clinical history. For example, if a patient is seen with acute sinusitis, a clinician is likely to ask, "Has this kept you from going to work?" However, this question is unlikely to be used to define the important elements of the health status of patients with terminal illness. A more relevant question for these patients might be, "Can you get out of bed?" The latter question, however, is insensitive to important quality of life issues for the sinusitis patient. By analogy, the HUI is probably best at evaluating HRQOL in a relatively healthy population, and the SIP likely is more able to evaluate HRQOL among more severely ill patients.

In order to fully evaluate health status measures, it is necessary to compare the ability of various measures to discriminate among health states across a wide spectrum of burden of disease. In addition, an ideal measure of HRQOL would be able to distinguish changes within individuals across time. This sort of "evaluative" measure is especially important for the use of health status measures in the clinical setting. The sensitivity of the SF-36 in detecting

Comparison of Health Status Measures

FIGURE 1. Distributions for the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) physical functioning and mental health, Sickness Impact Profile (SIP) physical and psychological domains, SIP overall, and Health Utilities Index Mark II (HUI) scores. All distributions are presented with perfect health on the right.

longitudinal change within patients has been questioned.¹³ It would be important in evaluating a group of health status measures to determine the relative performance of the measures as evaluative tools, again using subjects across a wide spectrum of burden of illness.

Given the increasing diversity of users of health status measures, there is also an increasing divergence in the reasons for choosing one measure over another. Issues in choosing instruments would have to include not only the metrics of the instrument (such as validity, reliability, and sensitivity) but also subject satisfaction with the instrument. Time required for completion of the measure is also an important factor. Finally, our data suggest that users should choose a measure appropriate for the burden of disease in the target population.

REFERENCES

1. Guyatt GH, Patrick DL, Feeny DH. Postscript, proceedings of the international conference on the measurement of quality of life as an outcome in clinical trials. *Control Clin Trials*. 1991;12(suppl):266-69.
2. Ware JE. Standards for validating health measures: definition and content. *J Chronic Dis*. 1987;40:641-43.
3. Osoba D, Till JE, Pater JL, Young JR. Health-related quality of life: measurement and clinical application. A workshop report. *Can J Oncol*. 1995;5(1):338-43.
4. Hennessy CH, Moriarty DG, Zack MM, Scherr PA, Brackbill R. Measuring health-related quality of life for public health surveillance. *Public Health Rep*. 1994;109(5):665-72.
5. Garretsen HFL, Gilst ECH, Van Oers HAM. Collecting health information at a local level. *Health Promotion Int*. 1991;6:121-3.
6. Quality of life as a new public health measure—behavioral risk factor surveillance system, 1993. *MMWR*. 1994;43(20):375-80.
7. Peskin SR. Applications of QOL measurements: a managed care perspective. *Oncology*. 1995;9(11 suppl):127-8.
8. Fowles JB, Weiner JP, Knutson D, Fowler E, Tucker AM, Ireland M. Taking health status into account when setting capitation rates: a comparison of risk-adjustment methods. *JAMA*. 1996;276(16):1316-21.
9. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): conceptual framework and item selection. *Med Care*. 1992;30:473-83.
10. McHorney CA, Ware JE, Raczel AE. The MOS 36-Item Health Survey (SF-36): psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993;31:247-63.
11. Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic medical conditions. *JAMA*. 1989;262(7):907-13.
12. Komaroff AL, Fagioli LR, Doolittle TH, et al. Health status in patients with chronic fatigue syndrome and in general population and disease comparison groups. *Am J Med*. 1996;101:281-90.
13. Weinberger M, Kirkman MS, Samsa GP, et al. A nurse-coordinated intervention for primary care patients with non-insulin-dependent diabetes mellitus: impact on glycemic control and health-related quality of life. *J Gen Intern Med*. 1995;10(2):59-66.
14. Bergner M. The sickness impact profile: development and final revisions of a health status measure. *Med Care*. 1981;19(8):787-805.
15. DeBruin AF, DeWitte LP, Stevens F, Diederiks JPM. Sickness Impact Profile: the state of the art of a generic functional status measure. *Soc Sci Med*. 1992;35(8):1003-14.
16. Torrance GW, Feeny GH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark II. *Med Care*. 1996;34(7):702-22.
17. Saigal S, Feeny D, Rosenbaum P, Furlong W, Burrows E, Stoskopf B. Self-perceived health status and health-related quality of life of extremely low-birth-weight infants at adolescence. *JAMA*. 1996;276(6):453-9.
18. Wyatt JR, Niparko JK, Rothman M, deLissovoy G. Cost utility of the multichannel cochlear implants in 258 profoundly deaf individuals. *Laryngoscope*. 1996;106(7):816-21.
19. Grotta J, for the US and Canadian Lubeluzole Ischemic Stroke Study Group. Lubeluzole treatment of acute ischemic stroke. *Stroke*. 1997;28(12):2338-46.