Reprinted from MEDICAL CARE, September, 1998
Vol. 36, No. 9
Lippincott-Raven Publishers Printed in U.S.A.

Performance of a Self-Administered Mailed Version of the Quality of Well-Being (QWB-SA) Questionnaire Among Older Adults

ELENA M. ANDRESEN, PHD,*† BARBARA M. ROTHENBERG, MPA,†‡ AND ROBERT M. KAPLAN, PHD§

OBJECTIVES. **The Quality of Well-Being questionnaire is a measure of health-related quality of life (HRQoL) that has several desirable properties. Its widespread use has been hindered because it is difficult to administer. To overcome this limitation, a new self-administered form has recently been developed. This study examined the feasibility of using the Quality of Well-Being–Self-Administered (QWB-SA) questionnaire in an older population.**

METHODS. **The Quality of Well-Being–Self-Administered questionnaire was sent to 430 community-dwelling individuals aged 65 years and older who were randomly selected from primary care physicians' offices. Response patterns, scaling distributions, and the acceptability of the survey were examined for all respondents. The results of the QWB-SA questionnaire were compared to the Sickness Impact Profile (SIP) and the Medical Outcomes Study 36-item Short-Form Health Survey (SF-36) for those individuals who also had completed the latter two surveys approximately 10 months earlier and whose health had not changed substantially in the meantime.**

RESULTS. **Three hundred and one older adults (70%) responded. The mean QWB-SA questionnaire score was 0.7035. The scores were not skewed, and there were no floor or ceiling effects. The mean time to complete the QWB-SA questionnaire was 14.2 minutes, which was significantly shorter than for the SIP (19.3 minutes) but significantly longer than for the SF-36 (12.5 minutes). Subjects rated their satisfaction with the QWB-SA questionnaire somewhat lower than for the SIP and similar to SF-36. Correlations between the QWB-SA questionnaire and the SIP and SF-36 were moderate and were generally stronger for measures of physical health than for other domains such as mental health.**

CONCLUSIONS. **The self-administered QWB questionnaire was acceptable to older respondents, and it correlated with other measures of health-related quality of life. It can be considered as a candidate for some research applications among older adults.**

Key words: health status; gerontology; survey methods. (Med Care 1998;36:1349–1360)

*From the Department of Community Health, Saint Louis University School of Public Health, St. Louis, Missouri.

†From the Department of Community and Preventive Medicine, University of Rochester Medical Center, Rochester, New York.

‡From the Rochester Health Commission, Rochester, New York.

§From the Division of Health Care Sciences, 0622, Department of Community and Family Medicine, University of California, San Diego, La Jolla, California.

Supported in part by the Rochester Area Pepper Center at the University of Rochester, New York (PHS Grant No. 2 P60 AG10463-06).

Address correspondence to: Elena M. Andresen, PhD, Department of Community Health, Saint Louis University, 3663 Lindell Blvd., St. Louis, MO 63108-3342.

Received April 29, 1997; initial review completed June 1, 1997; final acceptance February 13, 1998.

Researchers considering one or more outcome measures for research with older adult groups can select from a number of objective measures (eg, mortality, hospitalization, cost) or more subjective measures, including self-reported health status or health-related quality of life (HRQoL). A recent report by the Institute of Medicine emphasized the importance and need for continued exploration of HRQoL measures among older adults.[1] The present study examined the performance of a self-administered mailed version of the Quality of Well-Being (QWB-SA) questionnaire, one of a number of surveys that can be used to measure HRQoL.

The Quality of Well-Being questionnaire is a preference-weighted measure. It combines three scales of functioning with a measure of symptoms and problems to produce a point-in-time expression of well-being that runs from 0 (for death) to 1.0 (for asymptomatic full function). With the QWB, all observations of study enrollees can be expressed numerically whether they die, experience a deterioration or improvement in health, or remain stable. In addition, QWB scores can be translated into quality-adjusted life years (QALYs) and used for policy analysis. The QWB is a preference-based instrument that uses weights derived from the general population. The preference weights estimate valuation of wellness without consideration of risk.[2,3]

The QWB has been validated in both population and clinical studies. For example, estimates from the US National Health Interview Survey compared three problems: sinusitis, diabetes, and chronic lung disease. In each of three age groups, sinus disease was shown to be a less serious problem than diabetes, which in turn had less impact than emphysema.[4] A variety of analyses have evaluated the relation between the QWB and biologic markers of the human immunodeficiency virus.[5] Other applications of the QWB include chronic obstructive pulmonary disease, cystic fibrosis, diabetes mellitus, atrial fibrillation, arthritis, cancer, and depression.[6–12] Further, the method has been used for modeling the allocation of health resources and has served as the basis for an experiment on rationing of health care by the State of Oregon.[13]

These data recommend the QWB as a generic measure of HRQoL. Some researchers have reported problems in using the QWB, however, because of its complexity and the need for substantial interviewer training.[14–16] A new self-administered format, the QWB-SA, may allow for widespread use of this outcome measure.[17] We assessed the feasibility and acceptability of using a mailed version of the QWB-SA among older adults and compared the QWB-SA to the Sickness Impact Profile (SIP), the Medical Outcomes Study 36-item Short-Form Health Survey (SF-36), and two National Center for Health Statistics (NCHS) disability days measures. We explored the pattern of skipped items, issues of completion time, and overall acceptance of the QWB-SA. We expected that the QWB would have a more normal distribution than the SIP or SF-36 and that there would be substantial agreement among the three questionnaires.

Methods

Development and Content of the Self-Administered Quality of Well-Being Questionnaire

The format of the QWB-SA includes five sections. The first section asks about acute and chronic symptoms. First, respondents are asked, in a yes/no format, whether they have each of 18 chronic symptoms or problems. Examples are blindness or severely impaired vision in both eyes (a separate item asks about one eye) and speech problems. Part I also asks about 25 acute physical symptoms (eg, headache, coughing, or wheezing) and 11 mental health symptoms (eg, spells of feeling upset). The format of these items requests respondents to think back over the last 3 days and indicate whether the symptom was absent or present yesterday, 2 days ago, and/or 3 days ago. A summary of the symptom items is shown in Table 1. Parts II through V use a similar 3-day recall format: Part II asks about self-care. It includes two items asking if the respondent has been in an institution (eg, hospital) and whether they need help with caring for themselves. Part III asks about mobility (eg, use of public transportation or driving). Part IV asks about physical functioning, such as walking and confinement to a bed or chair. Performance of usual activity, such as work, school, or housework, is evaluated in Part V.

The current version of the QWB uses a list of 26 symptom/problem complexes, some of which include multiple, unrelated symptoms. In developing the QWB-SA, a group of clinicians reviewed the list and rewrote it to approximate what appears in standardized history and systems reviews. This involved disaggregating some of the

TABLE 1. Missing Responses on Quality of Well-Being–Self-Administered Questions for 301 Older Adults*

Question Content	% Missing	Question Content	% Missing
Yes/no questions		Bowel symptoms	7.0
Speech problems	0.7	Urinary symptoms	6.3
Deformity	2.3	Bladder control symptoms	7.0
Tiredness, weakness	2.3	Genital symptoms	7.3
Weight gain/loss	2.0	Broken bone	4.7
Under/overweight	4.6	Swelling symptoms	6.3
Problems chewing	1.0	Fever	4.7
Hearing loss	0.3	Seizures, unconscious	3.3
Skin problems	0.3	Balance gait symptoms	7.6
Eczema/rash	2.3	Sleep problems	5.6
Dentures	5.0	Nervous, shaky	5.6
Use oxygen	6.3	Upset, blue	4.0
Use prosthesis	8.6	Anxiety	5.3
Use glasses, contacts	4.3	Feeling lack of control	7.0
Use hearing aid	7.0	Lonely	6.0
Use magnifying glass	6.0	Frustration, irritation	6.0
Use neck, back, leg brace	6.0	Hangover	4.7
Multiple-stage questions		Sexual problems	14.6
Vision both eyes†	3.0	Cognitive symptoms	5.3
Vision one eye†	8.6	Obsessive symptoms	5.6
Back symptoms‡	7.0	Any medications	10.0
Hips, sides symptoms‡	10.0	Dietary restrictions	7.6
Hands, feet symptoms‡	8.3	Appetite symptoms	4.7
Use wheelchair‡	13.3	Institutionalized	11.6
Other(s) control its movement	13.3	Personal care needs	12.0
3-day recall questions		Drive a car	12.3
Uncorrected vision	7.3	Use public transportation	15.0
Eye symptoms	7.0	*Transportation problems*	15.0
Headache	7.6	Climbing stairs	14.0
Ear symptoms	7.0	Walking problems	15.0
Ear bleeding, discharge	7.0	Limp, use walking aid	14.0
Nasal symptoms	6.6	Trouble kneeling, etc.	13.6
Throat symptoms	5.3	Trouble lifting, carrying	13.6
Toothache/jaw pain	5.0	Other movement problems	16.9
Problems lips/tongue/gums	5.3	Stay in bed, etc.	12.6
Cough/wheeze	6.6	Limited usual activities	4.3
Breathing symptoms	6.3	Limited social activities	4.3
Chest symptoms	7.3	Change plans due to health	7.0
Stomach symptoms	7.3		

*The wording of these questions is available by request from the authors.
†These are multiple-stage yes/no questions.
‡These are multiple-stage 3-day recall questions.

complexes, which had included unrelated symptoms that had similar preference weights, or amplifying them (eg, distinguishing between blindness in one and both eyes). This format is easier for respondents to answer for the identically weighted items. In addition, five new items were included in the mental health section, for a total of 11 items, and three self-rated health questions were added as well. The final QWB-SA has 58 symptoms/problems and 71 questions overall.

For this application, the symptoms/problems weights were based on the standardized QWB preference weighting system.[2,3] The QWB-SA scoring algorithm selects the symptom or problem that gives the lowest preference weight. Nearly all of the symptoms/problems were components of previously standardized QWB symptom states. The improved symptoms assessment not only better reflects health status, it resembles more closely a clinical review of symptoms, thus increasing the clinical utility of the QWB-SA. In the few cases where new symptoms were added, the standardized QWB weight, ie, the average weight across all symptoms and problems, was applied. This allowed for maximum correspondence between the QWB-SA and the interviewer-administered form.

Sampling and Data Collection

The subjects were selected from the eligible patients at each of four physicians' offices cooperating with the University of Rochester's Office of Clinical Practice Evaluation, a department that offers quality assurance support to local primary care practices.[18] The details of sampling and survey methods have been described previously.[19] Briefly, physicians screened the computer-generated lists of older patients from their own practices to exclude those who were living in long-term care facilities or who were not able to communicate in writing because of language barriers or extreme physical frailty (eg, some stroke patients). Subjects then were randomly selected, drawing equally from those aged 65 to 74 years and from those 75 years and older to ensure an adequate sample of older patients. To compare respondents and nonrespondents, demographic information (age and gender) was recorded from physicians' records on all potential subjects.

The SIP and SF-36 were mailed to potential subjects in a 24-page booklet, with the questionnaires in random order.[20–22] The booklets also included two questions about restricted activity and bed days from the National Health Interview Survey.[23,24] The original mailing, which included a cover letter from the patient's primary care physician, was followed by a reminder postcard, a second mailed questionnaire, and a telephone call for those who did not respond. Approximately 10 months later, the QWB-SA nine-page booklet was mailed to subjects using the same methods. To enhance the sample size of the QWB-SA study group, we added a sample of older adults from the same medical practices using the same selection methods ($n = 57$).

Other Instruments and Questions

The SIP is a 12-scale, 136-item survey that yields two dimension scores (Physical and Psychosocial) and a global score. All scores range from 0 to 1, usually expressed as a percentage, with higher scores indicating more severe disability or decreasing HRQoL. The self-administered version used for this study included adaptations from Hedrick et al[25] and Chapko et al[26] and asked that respondents answer "true" or "not true" to each item. The SF-36 contains eight individual and two summary scales derived from longer instruments used in the Medical Outcomes Study; the scales yield scores from 0 to 100, where higher scores represent better health status.[22,27]

For this pilot version of the QWB-SA, subjects also were asked two global health questions. The first question asked them to rate their health as excellent, very good, good, fair, or poor. A second question asked subjects to rank their average health on a scale of 0 to 100 (in 5-point increments). In addition, respondents were asked to record the time it took them to complete the QWB-SA. To compare the acceptability of the QWB-SA, SIP, and SF-36, each time the questionnaires were administered subjects were asked "How would you rate your satisfaction with this (first/second) questionnaire? (Very satisfied, somewhat satisfied, neither satisfied nor dissatisfied, somewhat dissatisfied, very dissatisfied)."

The first global health question was used to select subjects who reported the same level of global health as on the earlier SIP/SF-36 survey. The results of the SIP, SF-36, and QWB-SA were compared only for subjects who had identical answers to this question both times the surveys were administered. Ideally, of course, all three instruments would have been administered at the same

time. Because they were not, however, as a conservative measure only those individuals who reported no health change were included in this analysis. This step was taken in an effort to distinguish between measurement issues and the effects of changes in health status.

Analysis Plan

Primary analyses were conducted using SPSS 7.0.[28] Chi-square analysis was used to compare categorical information for simple frequencies. For continuous variables, Student's t tests were used to test for differences between groups. Ninety-five percent confidence intervals (95% CI) for proportions were calculated based on methods described by Fleiss.[29] Paired t tests were used to compare the subjects' reported completion times. The Wilcoxon signed rank test was used to test for differences between paired satisfaction scores. We reported that an instrument had a scaling problem ("floor" or "ceiling" effects) if at least 20% of subjects were at either extreme. To assess the relation between the QWB-SA and subjects' self-rated general health (grouped as excellent, very good, good, fair/poor), we used ANOVA with planned linear contrasts and the QWB-SA as the dependent variable. Pearson Product-Moment or Spearman's correlations were used to assess construct validity by comparing the QWB-SA to the SIP and SF-36 and to NCHS Disability Days.

Results

Profile of Respondents

A total of 430 older adults were mailed QWB-SA surveys between March and June of 1996. Two hundred and ninety-six people responded to the multiple mailing strategy, and 37 of the 134 nonrespondents were contacted by telephone, resulting in five more returned surveys. The total response was thus 301, or 70.0% (95% CI: 65.4%, 74.0%). More than 80% of the respondents returned the survey within 2 weeks of the first mailing. For a comparison of respondents and nonrespondents, see Table 2. Of those who returned a QWB-SA, 239 also had completed the SIP and SF-36 approximately 10 months earlier (75.6% response to the earlier survey).[19] The two groups of QWB-SA respondents were quite similar (see Table 2).

QWB-SA Scores

The mean QWB-SA score was 0.7035 overall (SD = 0.099; range, 0.25–1.0) and was virtually identical for women (mean = 0.7035) and men (mean = 0.7034) and for both age strata. The mean QWB-SA score was 0.7051 for subjects aged 65 to 74 years, and it was 0.7018 for those aged 75 years and older. Only three subjects received maximum QWB-SA scores of 1.0 (1%), well below our threshold for classifying the scale as having scaling limits because of extreme scores. The distribution of the QWB-SA was fairly normal, with minimal skewness (-0.143, Standard Error = 0.140) and kurtosis (1.437, Standard Error = 0.280). QWB-SA scores decreased steadily with declining health on the self-reported health question. The mean score for 49 subjects who reported they were in excellent health was 0.7525 (SD = 0.104), and it was 0.6449 (SD = 0.096) for seven subjects who said their health was poor ($P < 0.01$). Average QWB-SA scale scores for age-specific and gender-specific groups and among self-reported health categories are shown in more detail in Table 3, and the overall distribution of scores is shown in Figure 1. The overall F-test showed a

TABLE 2. Characteristics of Respondents and Nonrespondents

Subjects	Age (yr) [mean (SD)]	Gender (% male)	QWB-SA Score [mean (SD)]
Nonrespondents	76.2 (8.3)	35.7	Not applicable
All respondents (n = 301)	74.7 (6.6)	41.2	0.7035 (0.099)
p	0.06	0.28	Not applicable
Respondents who completed all three questionnaires (n = 239)	74.6 (6.39)	43.1	0.708 (0.10)
Respondents who completed only the QWB (n = 62)	75.1 (7.60)	33.9	0.687 (0.09)
p	0.57	0.197	0.14

TABLE 3. QWB-SA Scale Scores

	Number of Persons	Mean Score (SD)
Men overall	124	0.7034 (0.108)
Ages 65–74	69	0.7065 (0.114)
Ages 75+	55	0.6996 (0.100)
Women overall	177	0.7035 (0.092)
Ages 65–74	84	0.7039 (0.084)
Ages 75+	93	0.7018 (0.100)
Self-reported health*		
Excellent	49	0.7525 (0.104)
Very good	92	0.7215 (0.096)
Good	101	0.6912 (0.083)
Fair	43	0.6448 (0.087)
Poor	7	0.6449 (0.096)
All subjects	301	0.7035 (0.099)

* ANOVA, $P < 0.01$; nine subjects skipped this question.

difference in QWB scores among respondents selecting different categories of self-rated health ($F\ 3/288 = 13.28$, $P < 0.001$). The linear contrast was highly significant ($F\ 1/288 = 39.69$, $P < 0.001$).

Acceptability of the QWB-SA

Subjects reported that they completed the QWB-SA in an average of 14.2 minutes (SD = 10.3; $n = 289$ for this question). The median time was 10 minutes, and 75% of subjects completed it in 15 minutes or less (95% CI: 69.6%, 79.6%). Among the subjects who reported completion times on all three surveys ($n = 213$), the mean for the QWB-SA was approximately 5 minutes less than for the SIP (14.1 minutes and 19.3 minutes, respectively; $P < 0.01$) and 2 minutes longer than for the SF-36 (14.4 minutes versus 12.5 minutes; $P < 0.01$).

The survey was generally well-received by respondents. Sixty percent of the 293 subjects who answered the question about "satisfaction" with the QWB-SA reported that they were very or somewhat satisfied (95% CI: 54.2%, 65.4%). Only nine people reported any level of dissatisfaction; the remainder was neither satisfied nor dissatisfied. Among 219 subjects who recorded their opinion on all three surveys, the QWB-SA rated somewhat lower than the SIP (60% versus 69% very or somewhat satisfied, $P = 0.04$) and similar to the SF-36 (67%; $P = 0.13$).

Patterns of Missing Responses

No QWB-SA survey item was answered by every respondent. Most items asked respondents to answer questions about their last 3 days. If the symptom or problem was not present, the directions explicitly told subjects to check the category "no days." In general, the 3-day recall questions posed more of a problem for subjects than the 18 single-category (yes/no) questions (eg, impaired vision, hearing loss). Answers on yes/no questions were missing for between 0.3% (hearing loss, skin problems) to 8.6% (use of prosthesis) of respondents. The question on use of a prosthesis

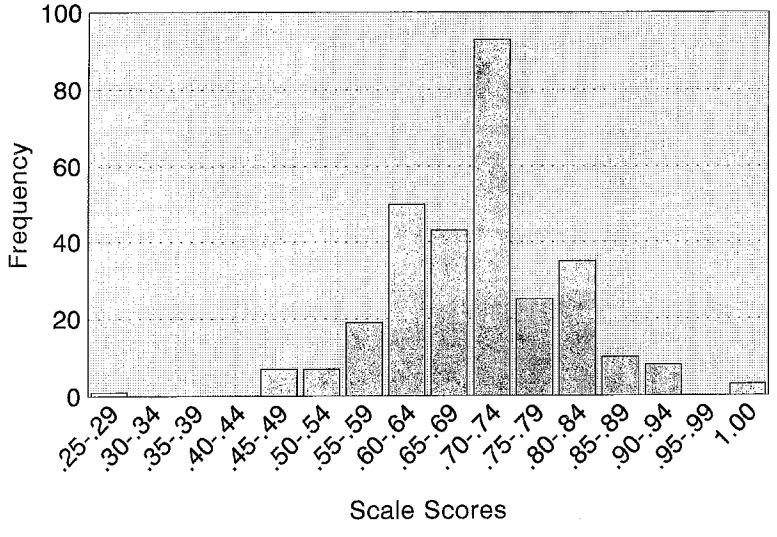

FIG. 1. Distribution of 301 QWB-SA scores.

may have caused some confusion for individuals who did not have such a device. For two items that had two-step answers (eg, blindness or severely impaired vision in both eyes followed by a question about one eye), the second question was more commonly skipped. Although the directions asked subjects to complete each item, it would be understandable if this sort of sequence were skipped sometimes. For example, a 3-day recall question asks if the subject has missing or paralyzed hands, feet, or arms: one subject said this was true, but 16 other subjects for whom this was not true skipped the follow-up question asking if they had missing or paralyzed fingers or toes.

In the section of the QWB-SA asking about acute and chronic symptoms during the past 3 days, all items were missing at least 3% of subject responses. Questions with the lowest percentage of missing responses included loss of consciousness (3.3% of subjects skipped this question), loss of appetite or overeating (4.7%), and toothache or jaw pain (5.0%). The symptom with the highest missing response was one on loss of sexual interest or performance (14.6%). Two self-care items also had more missing responses: placement in a hospital, nursing home, or rehabilitation center (11.6%) and a general question on the need for help in personal care needs (12.0%). Responses to questions on mobility and physical activity were missing for 12.3% to 16.9% of respondents. Three "usual activity" questions had fewer response problems and were missing for 4.3% to 7.0% of subjects. Table 1 lists the percentage of missing responses for each QWB-SA symptom.

In all, 150 subjects (49.8%) skipped at least one symptom on the QWB-SA 3-day recall section, and the mean number of missing items was 4.7 items (SD = 9.3). The number of items missed was related to gender but not to age. Men skipped an average of 6.3 (SD = 11.9) of the multiple-day questions compared with only 3.5 (SD = 6.6) for women ($P < 0.05$). Only among women did older respondents miss more items: women aged 65 to 74 years missed an average of 2.0 items (SD = 4.5), whereas women aged 75 years and older missed an average of 4.9 items (SD = 7.8; $P < 0.01$). For men, the corresponding figures were 6.2 items (SD = 12.9) and 6.3 items (SD = 10.6), respectively.

Respondents also skipped questions on their earlier survey.[19] Among the 239 subjects who returned both their SIP/SF-36 and QWB survey packets, 73 persons said that they had had at least 1 bed-day in the last 12 months due to illness or injury, but only 65 (89.0%) actually reported how many days they had been in bed. Because the traditional SIP scoring assumes a "no" for any item that is not directly answered in the affirmative, no subjects had missing SIP scale scores. Not surprisingly, the item missed most frequently was one about sexual activity ($n = 18$, 11.7% missing).

The SF-36 scales can be scored if up to 50% of the items in a scale are missing by using mean item replacement[22]; even so, no scale was complete for 100% of subjects. The scales with the lowest completion levels were the Role—Physical (6.7% missing), and Role—Emotional and Physical Functioning scales (6.3% missing). The two summary scales rely on all of the individual items, and each was missing for 10.9% of subjects. Individual SF-36 questions that were missed most often were "limitations; vigorous activities" (7.5% missing), and "limitations; walking one block" (5.4%).

Floor and Ceiling Effects

For the SIP, a substantial percentage of subjects received individual scale scores of 0 (ranging from 30.5% for the Social Interaction scale to 86.9% for the Work scale), indicating no measurable negative health state (a "ceiling" effect). The Physical Dimension score was 0 for 27.3% of subjects, and the Psychosocial Dimension score was 0 for 22.0% of subjects. Only the SIP total score did not suffer from a strong ceiling effect (7.8% received scores of 0); however, there was no discernible problem with extreme scores in the negative health range (ie, no "floor" effect, or scores of 100%). In contrast, the SF-36 showed floor effects for one scale (Role—Physical, with 21.6% of respondents scoring 0) and ceiling effects on Role—Physical (48.0%), Social Functioning (57.5%), Role-Emotional (72.0%), and Pain (20.8%). The standardized summary scales both demonstrated very normal scale ranges, with no subject receiving either 0 or 100 scores, which is a result of the scaling algorithm (27). Approximately 70% of subjects reported that they had 0 activity days on either the NCHS bed-day or restricted activity day question.

Correlations Among the QWB-SA, SIP, and SF-36

Correlations among the QWB-SA and other measures of HRQoL are shown in Table 4. In comparing these measures, only the 125 subjects

TABLE 4. QWB, SIP, SF-36, and Disability Days: Pearson Product-Moment Correlations*,† Among Measures for Older Adults

		SF-36												
										Summary Scales		NCHS Days*		
	QWB-SA	Physical Functioning	Role-Physical	Social Functioning	Mental Health	Role-Emotional	Energy Vitality	Pain	General Health	Physical	Mental	Bed	Restricted	Age
QWB-SA	—	0.511	0.281	0.277	0.383	0.167	0.468	0.410	0.492	0.467	0.223	-0.252	-0.339	-0.071
SIP total score	-0.448	-0.805	-0.636	-0.538	-0.414	-0.555	-0.671	-0.542	-0.613	-0.745	-0.378	0.284	0.239	0.365
Sleep and rest	-0.309	-0.487	-0.404	-0.382	-0.389	-0.400	-0.537	-0.313	-0.433	-0.452	-0.385	0.254	0.129	0.183
Eating	-0.120	-0.204	-0.223	-0.278	-0.081	-0.352	-0.147	-0.099	-0.148	-0.171	-0.209	0.162	0.191	0.001
Work	-0.110	0.045	-0.025	0.022	0.051	0.131	-0.088	-0.130	-0.134	-0.114	0.089	0.211	0.155	-0.125
Home management	-0.333	-0.803	-0.509	-0.433	-0.277	-0.501	-0.568	-0.439	-0.490	-0.664	-0.262	0.162	0.073	0.438
Recreation and pastimes	-0.335	-0.671	-0.626	-0.492	-0.237	-0.414	-0.589	-0.505	-0.489	-0.696	-0.244	0.203	0.245	0.390
Physical dimension	-0.420	-0.864	-0.524	-0.393	-0.285	-0.381	-0.545	-0.461	-0.562	-0.742	-0.163	0.139	0.124	0.368
Ambulation	-0.395	-0.830	-0.559	-0.314	-0.203	-0.350	-0.500	-0.468	-0.521	-0.749	-0.082	0.208	0.172	0.389
Mobility	-0.372	-0.741	-0.371	-0.379	-0.333	-0.304	-0.535	-0.395	-0.538	-0.606	-0.213	0.008	-0.056	0.325
Body care and movement	-0.379	-0.776	-0.460	-0.412	-0.279	-0.391	-0.449	-0.362	-0.476	-0.628	-0.195	0.089	0.198	0.260
Psychosocial dimension	-0.401	-0.498	-0.528	-0.502	-0.581	-0.621	-0.554	-0.459	-0.513	-0.474	-0.573	0.223	0.216	0.196
Social interaction	-0.336	-0.498	-0.385	-0.405	-0.385	-0.324	-0.433	-0.356	-0.450	-0.453	-0.314	0.158	0.100	0.258
Alertness behavior	-0.237	-0.271	-0.384	-0.434	-0.497	-0.592	-0.401	-0.273	-0.301	-0.234	-0.573	0.064	0.153	0.081
Emotional behavior	-0.367	-0.412	-0.367	-0.253	-0.402	-0.354	-0.477	-0.485	-0.484	-0.449	-0.318	0.249	0.318	0.089
Communication	-0.262	-0.353	-0.389	-0.312	-0.314	-0.381	-0.277	-0.242	-0.303	-0.333	-0.291	0.131	0.149	0.223
NCHS disability days*														
Bed days	-0.252	-0.195	-0.260	-0.060	-0.030	-0.048	-0.241	-0.103	-0.190	-0.240	-0.022	—	0.570	0.025
Restricted days	-0.339	-0.227	-0.346	-0.206	-0.145	-0.144	-0.338	-0.307	-0.208	-0.321	-0.106	0.570	—	-0.046
Age	-0.071	-0.360	-0.329	-0.143	-0.044	-0.300	-0.281	-0.192	-0.089	-0.299	-0.107	0.025	-0.046	—

*Spearman's ρ was calculated for all correlations with NCHS disability day scores.
†Correlation coefficients at or above the level of 0.3 are significant at $P < 0.01$ (two-tailed test).

with identical responses about their overall health (classified as excellent, very good, good, fair, or poor) at both survey times were included. Correlations between the QWB-SA and the physical health components of the SIP and SF-36 were moderate. For example, the correlation between the QWB-SA and the SIP Physical Dimension was r = -0.42, whereas it was 0.47 with the SF-36 Summary Physical scale. The correlations between the QWB-SA and mental health measures were somewhat weaker (r = -0.40 and 0.22 for the SIP Psychosocial Dimension and SF-36 Summary Mental Health scale, respectively). Generally, the correlations were stronger between the QWB-SA and the SF-36 than between the QWB-SA and the SIP. The correlations between the NCHS measures and the QWB-SA scores were weak or modest (r = -0.25 for bed days and r = -0.34 for restricted activity days).

Discussion

The mean QWB-SA scores for these older adults were similar to other published reports of research among older adult populations using the interviewer-administered QWB. However, a formal comparison of both instruments in the same subjects would be valuable. Estimates of average scores using the traditional interviewer format for the QWB include a Medicare demonstration project in North Carolina.[30,31] The mean QWB score for that population was 0.70 (SD = 0.11). The demographic makeup of the population was somewhat similar to the current sample, with a mean age of 75.2 years and a similar proportion of women (61%); however, the study group from North Carolina was 31.8% African-American elderly adults compared with only 2% in the current study. In another study using the same traditional QWB version but with selected healthier older adults, Andresen et al[14] reported a higher mean score of 0.73 (SD = 0.87) among community-dwelling adults with a younger mean age of 72.5 years. In that study, the QWB also was compared to both the SIP and three scales of the SF-36. The correlation with the overall SIP score was -0.52, and correlations with the SF-36 Physical Function, Role—Physical, and General Health scales were more modest, ranging from r = 0.36 to 0.39. In the population-based Beaver Dam study, mean QWB scores were 0.73 (SD = 0.10) for the population aged 45 years and older.[32] The mean QWB-SA score in a study of 123 women and 95 men (mean age = 48 years) in San Diego clinics was 0.706 (SD = 0.108).[33]

A very thorough study of potential HRQoL measures for rheumatoid arthritis research compared 28 different measures, including an arthritis-specific QWB, among 303 patients in a randomized controlled trial.[15] The patients' mean QWB score of 0.60 was substantially lower than for the current study or for other nonpatient groups described above, despite the younger age of the arthritis patients (18 to 65 years old). Another formal comparison of outcome measures including the QWB was conducted for a randomized controlled trial among patients with angina pectoris[16]; the version tested for the trial asked about symptoms during the prior 16 days, a more intimidating task than in the current QWB-SA. The mean QWB score for 59 patients (mean age = 65 years) was 0.68 (SD = 0.10), and the correlation between the SIP total score and the QWB score was r = -0.55, which was somewhat stronger than in the present study.

The interviewer-administered QWB recently was adapted and studied as an outcome measure for residents of long-term care facilities[34]; it showed promise for this frail population. The adult subjects of our study were selected to include older ages, but they were community-dwelling adults, not frail elderly individuals living in nursing home settings. The performance of the QWB-SA among institutionalized individuals and/or among extremely frail older adults has yet to be determined.

One of the problems identified in this study was that some QWB data were missing (see Table 1). This represents a problem for the QWB-SA that was not common for the original QWB, where the interviewer oversees data collection. Missing data are common, however, for other self-administered questionnaires in addition to the QWB-SA, and some of these subjects who completed an earlier survey also skipped questions on the SF-36 and SIP.[19]

The QWB-SA scoring algorithm assumed that missing responses were equivalent to no response. For example, if the respondent did not complete the question about "blindness in one eye," it was assumed that he or she did not have this symptom. We do not know how reasonable it is to assume that the missing responses are equivalent to negative responses. This scoring would be appropriate if respondents did not understand the directions and assumed that they

could leave an item blank if the symptom or problem did not apply. This may explain why high prevalence problems (eg, skin problems, hearing loss) have few missing data whereas low prevalence problems (eg, use of wheelchair, movement problems) have more missing data. Conversely, we should not assume that data are missing at random. Thus, assuming that a respondent does not have the problem if the response is missing could produce an upward bias. It is also important to emphasize, however, that the QWB scoring protocol does not use all of the symptom information. Only the symptom or problem that received the lowest score was used. Thus, in many cases, the missing data were likely to have little or no impact on scoring. For example, if a respondent reported cognitive symptoms, which receive a substantial weight, having any other symptom missing would not affect the score.

Further research is needed to determine the impact of missing data. These data also suggest that extra instruction may be required to reduce the amount of missing data. In clinical trials, it is common to check forms for completeness before allowing a participant to leave the premises. We encourage such reviews when they are feasible.

Because missing items are scored as 0 (or not present) for both the QWB-SA and the SIP, they do not present a problem in obtaining scores for all respondents. The SF-36 scoring algorithm, in contrast, excludes a respondent if more than one half of the items on any one scale are missing. This cautious approach is justified because of the smaller number of items on the SF-36, but it does mean that the SF-36 has a disadvantage in the present comparison of the completion of these three instruments. More questions tend to be missed on self-administered surveys, and unlike the SIP and QWB-SA, the SF-36 scales were not scoreable for 100% of subjects. A gender difference was detected in skipped items for the QWB-SA in this study, and patterns of missing items on HRQoL surveys warrant continued attention.

Acceptability, reported by these subjects in response to a single subjective question, was somewhat lower for the QWB-SA than for the SIP but similar to the SF-36. The issue of what subjects think of survey instruments is a complex one and deserves more attention than the simple closed-ended, categorical question asked in the present survey. Subject/patient input to survey development often is lacking, and few researchers ask and report about subjects' opinions, although at least one report about subject response to the SIP was favorable.[35] Opinions about the acceptability of measures should be included in future studies, and a recent Institute of Medicine report recommended that older adults provide input into survey measures.[1]

The pattern of correlations between the QWB-SA and the SIP and SF-36 supports the construct validity of the QWB-SA. Correlations between the QWB-SA and scales related to physical health tended to be higher than for scales of social and mental health, which would be expected given the more extensive measurement of physical health in the QWB. Comparisons between the QWB-SA and the SIP total score, the SF-36 physical health summary, and the SIP physical dimension scores achieved correlations in the moderate range ($r = 0.40$ and higher). Because of the 10-month time lag between the SIP/SF-36 and the QWB-SA surveys, these correlations may underestimate the construct validity of the QWB-SA. We used a conservative approach in our analysis and included only those subjects who indicated in a very general sense that their health was the same after 10 months. We would not expect correlations for 10 months to be stable. Health status represents at least two components: a stochastic transition between health states and measurement error. When correlations are not perfect for a 10-month interval, both of these characteristics are represented. Thus, we feel that the combination of direction and pattern of correlations supports the construct validity of the QWB-SA.

The QWB-SA is the only utility-based survey in the present comparison. It can be translated directly into quality-adjusted life years, which are used in cost–utility analysis and other types of research. It is also unique in explicitly incorporating death as a measurable outcome. The self-administered format of the QWB alleviates previously reported concerns about its complexity and the cost of interviewer administration. Selection among the three HRQoL surveys presented here should be made based on the specific population and outcome measurement needs. The QWB-SA now can be added as a reasonable choice for self-administered tools.

Acknowledgments

The authors thank their Rochester colleagues Dr. Robert Panzer and Dr. Paul Katz for their support of the research and acknowledge with special thanks Ms. Beth Sheehan for her help with the data collection and study management.

References

1. **Feasley JC, ed.** Health outcomes for older people: Questions for the coming decade. Institute of Medicine. Washington, DC: National Academy Press, 1996.

2. **Kaplan RM, Bush JW, Berry CC.** Health status: Types of validity and the index of well-being. Health Serv Res 1976;11:478.

3. **Kaplan RM, Anderson JP.** The general health policy model: An integrated approach. In: Spilker B, ed. Quality of life and pharmacoeconomics in clinical trials. New York NY: Raven, 1996:309.

4. **Erickson P, Kendall EA, Anderson JP, Kaplan RM.** Using composite health status measures to assess the nation's health. Med Care 1989;27(Suppl 3):S66.

5. **Kaplan RM, Anderson JP, Patterson TL, et al.** Validity of the Quality of Well-Being Scale for persons with HIV infection. Psychosomatic Med 1995;57:138.

6. **Kaplan RM, Atkins CJ, Timms R.** Validity of a Quality of Well-Being Scale as an outcome measure in chronic obstructive pulmonary disease. J Chron Dis 1984;37:85.

7. **Orenstein DM, Nixon PA, Ross EA, Kaplan RM.** The Quality of Well-Being in cystic fibrosis. Chest 1989;95:344.

8. **Kaplan RM, Hartwell SL, Wilson DK, Wallace JP.** Effects of diet and exercise interventions on control and quality of life in non-insulin-dependent diabetes mellitus. J Gen Intern Med 1987;2:220.

9. **Ganiats TG, Palinkas LA, Kaplan RM.** Comparison of Quality of Well-Being Scale and Functional Status Index in patients with atrial fibrillation. Med Care 1992;30:958.

10. **Kaplan RM, Anderson JP, Wu AW, Mathews WC, Kozin F, Orenstein D.** The Quality of Well-Being Scale: Applications in AIDS, cystic fibrosis, and arthritis. Med Care 1989;27(Suppl 3):S27.

11. **Kaplan RM.** Quality of life assessment for cost/utility studies in cancer. Cancer Treat Rev 1993;19(Suppl A):85.

12. **Kaplan RM.** Using general quality of life measures to assess outcomes in mental health. In: Miller N, ed. Cost/effectiveness of psychotherapy. Proceedings of an NIMH Conference, NIMH. 1998. In press.

13. **Kaplan RM.** Application of a general health policy model in the American health care crisis. J Royal Soc Med 1993;86:277.

14. **Andresen EM, Patrick DL, Carter WB, Malmgren JA.** Comparing the performance of health status measures for healthy older adults. J Am Geriatr Soc 1995;43:1.

15. **Bombardier C, Raboud J.** A comparison of health-related quality-of-life measures for rheumatoid arthritis research. The Auranofin Cooperating Group. Controlled Clin Trials 1991;12:243S.

16. **Visser MC, Fletcher AE, Parr G, Simpson A, Bulpitt CJ.** A comparison of three quality-of-life instruments in subjects with angina pectoris: The Sickness Impact Profile, the Nottingham Health Profile, and the Quality of Well-Being Scale. J Clin Epidemiol 1994;47:157.

17. **Kaplan RM, Ganiats TG, Rosen P, Sieber W, Anderson JP.** Development of a self-administered Quality of Well-Being scale (QWB-SA): Initial studies (Abstract). Qual Life Res 1995;4:443.

18. **Andresen EM, Bowley N, Rothenberg B, Panzer R, Katz P.** Test–retest performance of a mailed version of the SF-36 among older adults. Med Care 1996;11:1165.

19. **Andresen EM, Rothenberg BM, Panzer R, Katz P, McDermott MP.** Selecting a generic measure of health-related quality of life for use among older adults. Eval Health Profes 1998;21:244.

20. **Bergner M, Bobbitt RA, Carter WB, Gilson BS.** The Sickness Impact Profile: Development and final revision of a health status measure. Med Care 1981;19:787.

21. **Hays RD, Sherbourne CD, Mazel RM.** The Rand 36-item Health Survey 1.0. Health Econom 1993;2:217.

22. **Ware JE, Sherbourne CD.** The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473.

23. **Wagner EH, LaCroix AZ, Grothaus LC, Hecht JA.** Responsiveness of health status measures to change among older adults. J Am Geriatr Soc 1993;41:241.

24. **Scholes D, LaCroix AZ, Wagner EH, et al.** Tracking progress toward national health objectives in the elderly: What do restricted activity days signify? Am J Public Health 1991;81:485.

25. **Hedrick SC, Rothman ML, Chapko M, Inui TS, Kelly JR, Ehreth J.** Overview and patient recruitment in the Adult Day Health Care Evaluation study. Med Care 1993;31(9 Suppl):SS3.

26. **Chapko M, Rothman ML, Ehreth J, et al.** Data collection in the Adult Day Health Care Study. Med Care 1993;31(9 Suppl):SS15.

27. **Ware JE, Kosinski M, Keller SD.** SF-36 physical and mental health summary scales: A user's manual. Boston, MA: The Health Institute, 1994.

28. **SPSS, Inc.** SPSS 7.0 for Windows. Chicago, IL: SPSS, Inc., 1996.

29. **Fleiss JL.** Statistical methods for rates and proportions, 2nd ed. New York, NY: John Wiley & Sons, 1981: 14.

30. **Gerrity MS, Gaylord S, Williams ME.** Short versions of the Timed Manual Performance Test: Development, reliability, and validity. Med Care 1993;31:617.

31. **Patrick DL, Erickson P.** Health status and health policy: allocating resources to health care. New York, NY: Oxford University Press, 1993.

32. **Fryback DG, Dasbach EJ, Klein R, et al.** The Beaver Dam Health Outcomes Study: Initial catalog of health-state quality factors. Med Decis Making 1993;13:89.

33. **Kaplan RM, Sieber WJ, Ganiats TG.** Quality of Well-Being Scale: Comparison of the interviewer-administered version with a self-administered questionnaire. J Psychol Health. 1997;12:783–791.

34. **Hays RD, Siu AL, Keeler E, et al.** Long-term care residents' preferences for health states on the Quality of Well-Being Scale. Med Decis Making 1996;16:254.

35. **Carter WB, Deyo R.** The impact of questionnaire research on clinical populations: A dilemma in review of human subjects research resolved by a study of a study. Clin Res 1981;29:287.