Analysis of the Performance Characteristics of the University of Washington Quality of Life Instrument and Its Modification (UW-QOL-R)

Ernest A. Weymuller, Jr, MD; Ramsey Alsarraf, MD, MPH; Bevan Yueh, MD, MPH; Frederic W.-B. Deleyiannis, MD, MPhil, MPH; Marc D. Coltrera, MD

Background: During a 5-year period, we analyzed 3 patient subsets from the University of Washington Quality of Life (UW-QOL) Registry and published the results. In each instance, editorial review has raised legitimate concerns regarding the UW-QOL instrument that deserve public comment. We present our response to these criticisms. Since our original publication (1993), we have added domains to the original UW-QOL instrument. These additions reflected our concern that we might be missing important elements in the spectrum of disease-specific response to treatment. Using the data we have accumulated in the last 5 years, we present an analysis of the internal consistency of the UW-QOL. We have identified those domains that are responsive (or not responsive) to treatment effect and have revised the UW-QOL accordingly to create the UW-QOL-R, which is recommended for future use.

Design: The project began January 1, 1993, after approval by the UW Human Subjects Committee. Critical comments offered by external review were collated and responded to. Internal consistency was evaluated by interitem correlation matrix (Cronbach $\alpha$) testing.

Subjects: All new patients presenting to the UW Medical Center (Seattle) with a diagnosis of head and neck cancer were asked to participate in a prospective analysis of QOL changes during and after treatment.

Intervention: Patients completed the pretreatment QOL questionnaire on the day of their initial workup. The format for the pretreatment test was an interviewer-supervised self-administered test; the subsequent tests were self-administered and were completed at 3, 6, 12, 24, and 36 months. Other data entered for each patient included site, stage, treatment, histologic classification, reconstruction, and current status. A QOL registrar was responsible for patient follow-up, data collection, and collation. All data were entered into the departmental relational database.

Results: Criticisms by external review included the following: “it is improper to call it [UW-QOL] a measure of quality of life”; “the summary scale is problematic because it implies that each of the subscales are weighted or ‘valued’ equally”; “some domain questions relate to surgery specific issues... while others are specific to radiation”; “we were confused by the scoring”; and “the UW-QOL index does not specifically address the psychological impact of the disease and its treatment.” After evaluation of internal consistency, the UW-QOL was modified by removing 2 domains that correlated poorly with the others. This resulted in a 10-item instrument (UW-QOL-R) with an overall internal consistency score of 0.85.

Conclusions: The UW-QOL can be effectively and accurately used to compare treatment effects in the management of head and neck cancer. With this revised instrument, the 10 items appear to measure the domains of overall QOL in a highly consistent and reliable fashion over time.

PATIENTS, MATERIALS, AND METHODS

PATIENTS
The project began January 1, 1993, after approval by the UW Human Subjects Committee. All new patients presenting to the UW Medical Center (Seattle) with a diagnosis of head and neck cancer were asked to participate in a prospective analysis of QOL changes during and after treatment. Patients completed the pretreatment QOL information on the day of their initial workup. The format for the pretreatment test was an interviewer-supervised self-administered test; the subsequent tests were self-administered and were completed at 3, 6, 12, 24, and 36 months. Other data entered for each patient included site, stage, treatment, histologic classification, reconstruction, and current status. A QOL registrar was responsible for patient follow-up, data collection, and collation. All data were entered into the departmental relational database.

DEFINITION OF TERMS
Composite UW-QOL score is the arithmetic mean of the 10 individual domain scores (maximum score, 100). Domain score was determined by offering participating patients a set of options (Likert scale) for each domain. The maximum (best) score is 100, the minimum is 0. As an example, the domain pain offers the following options: 100, I have no pain; 75, there is mild pain not requiring medication; 50, I have moderate pain that requires regular medication (codeine or nonnarcotic); 25, I have severe pain controlled only by narcotics; and 0, I have severe pain not controlled by antibiotics. Beginning in 1997, global score was determined by asking all patients to “consider everything that contributes to your personal well-being—how would you rate your overall quality of life during the past seven days.” The possible responses were excellent, very good, good, fair, poor, or very poor.

ANALYSIS OF INTERNAL CONSISTENCY
Analysis of internal consistency was performed by calculating the interitem correlation coefficients (Pearson r) for each of the time points of the 36-month period and the overall collection of data points. Cronbach α coefficients were then calculated using the SPSS computer software program (SPSS Inc, Chicago, Ill). Least correlated items were then eliminated from the revised instrument, and these α coefficients were then recalculated for the UW-QOL-R.

RESPONSES TO EDITORIAL CRITICISM

Criticism: “It is improper to call it [UW-QOL] a measure of quality of life despite its title . . . the authors should consider the use of the term disease-specific functional status instead of quality of life.”

Response: In his text, Guide to Clinical Trials, Spilker comments that there is no ideal test at present to evaluate quality of life. A test that could become a gold standard would be rapid to complete; be reproducible; be valid either in a single patient population or across a large number of diseases; be widely accepted; not require excessive training of staff to administer; be easy to interpret; [and] yield objective results. . . . The view of some researchers is that finding a test that meets all of these criteria is as difficult as finding the Holy Grail. It is an even greater challenge to devise a test that would be applicable to patients in different national cultures.

In his text, Quality of Life Assessments in Clinical Trials, Spilker says, The conceptual formulation which has emerged, and which is gaining acceptance, defines quality of life functionally by patients’ perception of performance in four areas: physical and occupational function, psychological state, social interaction, and somatic sensation. In this model the patient serves as his own control, the comparisons being made against expectation of function.

Spilker also notes that The simplest classification of quality of life tests divides them into indexes, profiles, and batteries. . . . An index is a test that yields a single number at its conclusion. It usually evaluates multiple domains and often tests multiple components of each domain. The test may include measures of the quantity as well as the quality of life.

This description most closely fits the UW-QOL instrument.

Functional disability scales (which are discussed as a subset of QOL assessment by Spilker) are more strictly used for the periodic assessment of physical disabilities. . . . Increasingly, clinicians and researchers need reliable and validated measures of functional disability to measure clinical progress, evaluate programs and establish appropriate eligibility for social and insurance programs.

The UW-QOL instrument does not fit these criteria. The inclusion of a global question in the process of QOL analysis is an important adjunct. According to Spilker, “Some authors and tests focus on objective criteria to define and measure quality of life, whereas others stress the measurement of subjective aspects of this concept. Using both approaches is best.” Spilker says, “Some authors and tests focus on objective criteria to define and measure quality of life, whereas others stress the measurement of subjective aspects of this concept. Using both approaches is best.”

Spilker also says that “Global quality of life questions do not need to be, and in fact cannot be, validated.” We believe that the inclusion of a global QOL question in the UW-QOL meets the criteria identified by Spilker, who confirms the
Revised University of Washington Quality of Life Questionnaire (UW-QOL-R)

This questionnaire asks about your health and quality of life over the past seven days. Please answer all of the questions by checking one box for each question.

1. Pain. (Check one box: ☐)
☐ I have no pain.
☐ There is mild pain not needing medication.
☐ I have moderate pain—requires regular medication (codeine or nonnarcotic).
☐ I have severe pain controlled only by narcotics.
☐ I have severe pain, not controlled by medication.

2. Appearance. (Check one box: ☐)
☐ There is no change in my appearance.
☐ The change in my appearance is minor.
☐ My appearance bother me but I remain active.
☐ I feel significantly disfigured and limit my activities due to my appearance.
☐ I cannot be with people due to my appearance.

3. Activity. (Check one box: ☐)
☐ I am as active as I have ever been.
☐ There are times when I cannot keep up my old pace, but not often.
☐ I am often tired and have slowed down my activities although I still get out.
☐ I don’t go out because I don’t have the strength.
☐ I am usually in bed or chair and don’t leave home.

4. Recreation. (Check one box: ☐)
☐ There are no limitations to recreation at home or away from home.
☐ There are a few things I can do but I still get out and enjoy life.
☐ There are many times when I wish I could get out more, but I’m not up to it.
☐ There are severe limitations to what I can do, mostly I stay at home and watch TV.
☐ I can’t do anything enjoyable.

5. Swallowing. (Check one box: ☐)
☐ I can swallow as well as ever.
☐ I cannot swallow certain solids foods.
☐ I can only swallow liquid food.
☐ I cannot swallow because it “goes down the wrong way” and chokes me.

6. Chewing. (Check one box: ☐)
☐ I can chew as well as ever.
☐ I can eat soft solids but cannot chew some foods.
☐ I cannot chew soft solids.

7. Speech. (Check one box: ☐)
☐ My speech is the same as always.
☐ I have difficulty saying some words but I can be understood over the phone.
☐ Only my family and friends can understand me.
☐ I cannot be understood.

8. Shoulder. (Check one box: ☐)
☐ I have no problem with my shoulder.
☐ My shoulder is stiff but it has not affected my activity or strength.
☐ Pain or weakness in my shoulder has caused me to change my work.
☐ I cannot work due to problems with my shoulder.

9. Taste. (Check one box: ☐)
☐ I can taste food normally.
☐ I can taste most foods normally.
☐ I can taste some foods.
☐ I cannot taste any foods.

10. Saliva. (Check one box: ☐)
☐ My saliva is of normal consistency.
☐ I have less saliva than normal, but it is enough.
☐ I have too little saliva.
☐ I have no saliva.

Which issues have been the most important to you during the past 7 days?

Check ☐ up to 3 boxes.
☐ Pain
☐ Appearance
☐ Speech
☐ Activity
☐ Shoulder
☐ Recreational
☐ Swallowing
☐ Taste
☐ Saliva

GENERAL QUESTIONS
Compared to the month before you developed cancer, how would you rate your health-related quality of life? (check one box: ☐)
☐ Much better
☐ Somewhat better
☐ About the same
☐ Somewhat worse
☐ Much worse

In general, would you say your health-related quality of life during the past 7 days has been: (check one box: ☐)
☐ Outstanding
☐ Very good
☐ Good
☐ Fair
☐ Poor
☐ Very poor

Overall quality of life includes not only physical and mental health, but also many other factors, such as family, friends, spirituality, or personal leisure activities that are important to your enjoyment of life. Considering everything in your life that contributes to your personal well-being, rate your overall quality of life during the past 7 days. (check one box: ☐)
☐ Outstanding
☐ Very good
☐ Good
☐ Fair
☐ Poor
☐ Very poor

Please describe any other issues (medical or nonmedical) that are important to your quality of life and have not been adequately addressed by our questions (you may attach additional sheets if needed).

Figure 1. Revised University of Washington Quality of Life instrument and global quality of life questions.

appropriateness of using a Likert scale to address the issue of global QOL.

Criticism: “The summary scale in a QOL instrument is problematic because it implies that each of the subscales are weighted or ‘valued’ equally, which is probably not a good assumption in a multi-scale instrument.”

Response: We believe that the opinion of Spilker once again holds.

Another limitation of health status questionnaires is the unsolvable issue of whether each item should be equally weighted. From a clinical perspective, not all activities of daily living are equal for a patient, and the technology of deriving weights leaves the clinician dissatisfied.”

Spilker also asks,

How does one weight the individual domain scores so as to arrive at a reasonable overall quality of life score? At the present time no studies resolve the issue. It may be that the relative weightings of quality of life domain scores are themselves variable over time, and hence not amenable to fixed weighting. . . . many researchers are now more confident in the use of quality of life subscores as probes (not as diagnostics), and suggest that an assessment of quality of life may include both an overall score as defined precisely for the instruments being used and component subscores. From an analytic point of view, this makes it possible to begin to dissect out component factors of quality of life and the variable impact treatment may have on each.”

Our conclusion is that importance weighting may add a level of complexity that is generally not worth the trouble, since we do not believe that the process of weighting is well defined or appropriate in this setting. When specific research inquiry warrants importance weighting, it is reasonable to ask the patient how he or she would weight the importance of a particular domain. This tech-

©2001 American Medical Association. All rights reserved.
University of Washington Quality of Life (UW-QOL) Instrument Internal Consistency*

<table>
<thead>
<tr>
<th>Follow-up Period, mo</th>
<th>UW-QOL</th>
<th>UW-QOL-R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretreatment (n = 456)</td>
<td>0.74</td>
<td>0.78</td>
</tr>
<tr>
<td>3 (n = 301)</td>
<td>0.83</td>
<td>0.87</td>
</tr>
<tr>
<td>6 (n = 309)</td>
<td>0.78</td>
<td>0.83</td>
</tr>
<tr>
<td>12 (n = 267)</td>
<td>0.81</td>
<td>0.84</td>
</tr>
<tr>
<td>18 (n = 216)</td>
<td>0.81</td>
<td>0.85</td>
</tr>
<tr>
<td>24 (n = 180)</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td>36 (n = 102)</td>
<td>0.82</td>
<td>0.83</td>
</tr>
<tr>
<td>Total (N = 1831)</td>
<td>0.81</td>
<td>0.85</td>
</tr>
</tbody>
</table>

*Data are Cronbach α scores determined by interitem correlation matrix for items of both instruments. Least correlated items (dryness and employment) were removed from UW-QOL instruments for revised (UW-QOL-R) internal consistency ratings.

The Cronbach α scores of internal consistency are quite high across the entire 36-month follow-up period for the original UW-QOL. For example, the range of these internal consistency values was 0.74 to 0.84, and the overall internal consistency score for the original UW-QOL was 0.81 (Table).

INTERNAL CONSISTENCY

We performed this analysis because, over time, we had added domains to the original UW-QOL. These additions reflected our concern that we might be missing important elements in the spectrum of disease-specific response to treatment. In assessing internal consistency, we have identified those domains that are responsive (or not responsive) to treatment effect.

Internal consistency refers to the reliability of each item or domain in a given instrument to provide a QOL measurement in a fashion that is similar to each other item or domain of that same instrument. In the process of this analysis, 2 of the 12 domains of this instrument were found to correlate poorly with the other 10 domains. These domains were “dryness” and “employment.” Both of these items had wide variations compared with the other measured QOL items and were found to show a bimodal rather than more normal distribution. We believe that it is likely that scores for both of these variables are influenced by factors external to the manner in which treatment affects the other QOL domains (ie, whether or not an individual underwent radiation therapy and whether or not an individual was employed to begin with) and thus do not tap the changes in QOL over time as accurately as the other domains. Thus, these items were removed from our revised instrument. With this revised instrument, there is no “cancellation effect,” because each of the 10 items appears to measure the domains of overall QOL in a highly consistent and reliable fashion over time (Figure 2).

The Cronbach α scores of internal consistency are quite high across the entire 36-month follow-up period for the original UW-QOL. For example, the range of these internal consistency values was 0.74 to 0.84, and the overall internal consistency score for the original UW-QOL was 0.81 (Table).

CONSIDERATION OF QOL STUDY COSTS

It is also important to consider the cost of QOL studies before embarking on a longitudinal project. In discussing clinical trials to evaluate QOL, Spilker indicates that
implicit in this strategy is the large volume of data that flows from a quality of life study. Several variables are measured at each encounter and patients are followed for a considerable time. In addition to the usual clinical information, several items of quality of life data will be collected. This has clear workload implication.\(^{6(20)}\)

We have found this statement to be painfully accurate. To pursue QOL data for longitudinal analysis, one must be committed to thorough data collection and account for the attendant costs. Collection of our data for 4 years generated costs in excess of $250,000 in personnel salary alone.

**CONCLUSIONS**

After administering this instrument to more than 500 patients, we can indicate that the UW-QOL meets the following desirable characteristics articulated by Spilker: (1) it is short and rapid to complete; (2) it is reproducible, reliable, and valid in a population of head and neck cancer patients; (3) it does not require excessive training to administer; and (4) it is easy to interpret and yields objective results (separation by site and stage).\(^{32}\)

We conclude that inclusion of a global measure of posttreatment QOL is a critical part of QOL assessment. Therefore, we recommend the UW-QOL-R for future use.

Accepted for publication September 22, 2000.

Dr Yueh is supported by a Career Development Award from the Health Services Research and Development Service of the Veterans Administration. Dr Yueh is also part of the Health Services Research and Development Service and Surgery Service, VA Puget Sound Health Care System.

Corresponding author: Ernest A. Weymuller, Jr, MD, University of Washington, Department of Otolaryngology-Head and Neck Surgery, Box 356515, Seattle, WA 98195 (e-mail: eaw@u.washington.edu).

**REFERENCES**