A Brief Review of Single-Item and Multi-Item Quality-of-Life Measures for Medicare Patients

John H. Wasson, MD

Abstract: Using responses to HowsYourHealth.org from 9068 patients aged 65 years or older, I illustrate measurement compromises for quality-of-life assessment and management. Key words: health status measurement, HowsYourHealth.org, Medicare, patient-reported outcome measures, quality-of-life measurement, what matters

HISTORY

Single-item measures for patient care

My colleagues and I began developing patient-reported measures in the 1980s after we documented a significant mismatch between what matters to primary care patients and what clinicians knew (Nelson et al., 1983). These single items measured physical, emotional, and social health, pain, overall health, and quality of life (QOL) (Nelson et al., 1990). Ease of administration and interpretation in patient care prompted international dissemination (de Azevedo-Marques & Zuardi, 2011; van Weel et al., 1995). A controlled trial involving 45 primary care clinicians and patients 70 years or older showed that when our measures were linked to patient education materials and explicit directives to clinicians, care improved (Wasson et al., 1999). We developed similar measures for other clinical situations (Wasson et al., 1994, 2000).

We recently prospectively tested a “What Matters Index” (WMI) that unambiguously identifies fundamental, remediable needs for each patient and directs the delivery of services to patient categories based on their risk for subsequent costly care (Wasson et al., 2018a, 2018b). The WMI is based on only 5 patient-reported measures: pain, emotional issues, polypharmacy, adverse medication effects, and low confidence in managing health problems. The methods for scoring each measure and reasons for inclusion are included in Table 1 (Wasson, 2017).
Table 1. Measures Included in a What Matters Index

<table>
<thead>
<tr>
<th>Patient-Reported Measure</th>
<th>Documented Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Health confidence</strong></td>
<td>Health confidence is an easy-to-measure proxy for patients’ capacity to manage their health problems. This capacity is the final common pathway for successful chronic care management. A low level of patient self-management capacity predicts poor involvement in self-care and is associated with increased use of costly health care services. Patients’ emotional problems and level of pain have a significant impact on the attainment of health confidence over time. These measures are fundamental to the human condition regardless of disease. A combination of pain and emotional problems has dramatic impact on health, and both pain and emotional problems may respond to simple behavioral interventions. In several forms, pain and emotional problems are already assessed as vital signs in clinical settings. Number of medications taken and adverse effects can give rise to harmful interactions and reduced adherence. Polypharmacy and medication side effects account for a very large percentage of preventable hospital and emergency department uses.</td>
</tr>
<tr>
<td>How confident are you that you can manage and control most of your health problems? <em>(Not very confident or somewhat confident scored as 1, very confident scored as zero.)</em></td>
<td></td>
</tr>
<tr>
<td><strong>Pain</strong></td>
<td>Pain during the past four weeks how much bodily pain have you generally had? <em>(Extreme or moderate pain scored as 1; none, very mild, or mild scored as zero.)</em></td>
</tr>
<tr>
<td><strong>Emotions</strong></td>
<td>Emotions during the past four weeks how much have you been bothered by emotional problems such as feeling anxious, irritable, depressed, or sad? <em>(Extremely or quite a bit scored as 1; not at all, a little, or somewhat scored as zero.)</em></td>
</tr>
<tr>
<td><strong>Polypharmacy</strong></td>
<td>How many prescription medicines are you taking more than three days a week? <em>(More than 5 scored as 1; 5 or less scored as zero.)</em></td>
</tr>
<tr>
<td><strong>Adverse effects from medicines</strong></td>
<td>Do you think any of your pills are making you sick? <em>(Yes or maybe scored as 1; no scored as zero.)</em></td>
</tr>
</tbody>
</table>


HowsYourHealth.org is an Internet-based application that integrates our single-item measures and the WMI into a comprehensive health checkup. It automatically provides clinicians a report about patients’ function, diagnosis, symptoms, health habits, preventive needs, capacity to self-manage chronic conditions, and their experiences of care (Nelson et al., 2015; Wasson et al., 2011). It has been freely available for almost 2 decades and provides the data for this report.

**Multi-item measures for outcome assessment**

Over a similar time, multi-item measures for assessing outcomes of health care have gained widespread acceptance. The earliest examples include various “short-form” (SF) measures used in the Rand Health Insurance Experiment and the Medical Outcomes Study (Brook et al., 1979; Tarlov et al., 1989). The Consumer Assessment of Healthcare Providers and Systems (CAHPS) is
commonly used for judging health care quality (Agency for Healthcare Research and Quality, n.d.). Innumerable multi-item measures claim specificity for diseases, symptoms, and components of health services. Multiple items generally distinguish smaller differences between interventions than single items. Therefore, they have been widely disseminated in the United States for health services evaluation and reimbursement. However, a significant sacrifice for this increased “granularity” are the costs of distribution and the nuisances of computation and interpretation. As an example, in their detailed analysis of SF-36 variants, Selim et al. (2018) needed complex mathematical techniques to compare management strategies: scoring, transforming, norming, imputation, and variance assessment. The authors then assumed that their calculated effect sizes reflect an important improvement in a patient’s QOL. This assumption is highly debatable. Multi-item measures principally designed for psychometric integrity and outcome assessments have long been known to impose implementation and interpretive problems that run counter to what matters to a patient and a clinician trying to serve that patient’s needs (Feinstein, 1992).

PURPOSE OF THE REPORT

The purpose of this report is to illustrate measurement compromises for QOL assessment and management.

METHODS

Patients

Responses to HowsYourHealth.org from 9068 patients aged 65+ years are the basis for this report. Fifty-four percent of the patients are female. Eleven percent are poor: they self-report at least some of the time having difficulty paying for essentials such as food, clothing, or housing. In total, 1706 patients have 3 or more of 5 common chronic conditions and 3499 are taking 3 or more prescription medications. The 5 chronic conditions are hypertension (n = 4793), arthritis (n = 4026), atherosclerotic cardiovascular disease (n = 1804), diabetes (n = 1673), and respiratory disease (n = 1325).

Quality of life

Patients self-report their overall QOL: “How have things been going for you in the past four weeks: very well—could hardly be better (27%), pretty good (48%), good and bad parts about equal (19%), pretty bad (5%), very bad—could hardly be worse (1%).”

Self-reported items

• The WMI measures.
• Multi-item comparison measures. A mix of measures is typically included in the SF-36 variants. For example, the SF-12 contains measures for overall health, emotional state, performance of physical activities, engagement in daily and social activities, and severity of pain (RAND Health, 2018). For this analysis, indicators for social determinants of health—a measure for social support and another for poverty—are added to the customary domains. This multi-item comparator is abbreviated in the following results as “6 FNX limitations, OH, Poverty.”
• A sum of the 5 chronic conditions listed previously.

Analyses

I present the Medicare patients’ “pretty bad” and “very bad” responses about their QOL as an indication of “poor QOL” (vs all other QOL responses).

I display the area under the receiver operating characteristic curve (AUROC) to examine the association of prototypic, multi-item measures with patient-reported QOL. The AUROC represents the probability that a randomly chosen patient who reports poor quality is correctly rated by a prototypic, multi-item measure compared with a randomly chosen patient who had neutral or positive QOL. When the AUROC is 0.70 to 0.79, its discrimination is considered fair, and when 0.80 to 0.89, good.
I include in the AUROC displays only 4 prototypic measures because their range subsumes that observed for other combinations. I also limit the displays to “poor QOL” because the results of an analysis for “good QOL” are not different (“very well” and “good” responses vs all others).

RESULTS

Poor QOL and other self-reported items
Table 2 displays the percentage of Medicare patients responding to each item when they rate their QOL either “poor” or “not poor.” As expected, negative responses to any item are strongly associated with poor overall QOL. Also listed is the Spearman correlation between the full scale of an item and the 5 possible responses for QOL ranging from “very well—could hardly be better” to “very bad—could hardly be worse.”

QOL and prototypic comparators
The Figure (left diagram) illustrates the association of the 5-item WMI, a sum of 5 chronic conditions and an 8-item comparator modeled after the SF-36. The better discriminators between QOL states will have a higher AUROC. The 8-item comparator is the best discriminator for poor QOL and the sum of 5 diagnoses is clearly the worst. The correlation of a full 5-response range of QOL reflects this difference ($r = 0.60$ vs $r = 0.28$).

An attribute of an AUROC is the insight it provides about the marginal advantage (or disadvantage) of adding or omitting measures. For example, if to the WMI any other additional functional limitation is added (from a response for a significant limitation in social support, social activity, daily activity, or physical activity), the AUROC is only marginally increased to 0.90 from 0.87. In other words, so many measures are correlated with QOL (Table 2), and therefore each other, that after a point the addition of measures is not meaningful.

Past hospital use and comparators
The Figure (right diagram) also includes the AUROC for identifying the 1884 Medicare

<table>
<thead>
<tr>
<th>Problem Categories</th>
<th>Poor Quality of Life (n = 531)</th>
<th>Not Poor Quality of Life (n = 8537)</th>
<th>Correlation* With Overall Quality of Life (Number of Response Categories)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall health b</td>
<td>82%</td>
<td>13%</td>
<td>0.69 (5)</td>
</tr>
<tr>
<td>Daily activities b</td>
<td>58%</td>
<td>7%</td>
<td>0.58 (5)</td>
</tr>
<tr>
<td>Social activities b</td>
<td>55%</td>
<td>5%</td>
<td>0.56 (5)</td>
</tr>
<tr>
<td>Emotional state b,c</td>
<td>53%</td>
<td>6%</td>
<td>0.56 (5)</td>
</tr>
<tr>
<td>Health confidence c</td>
<td>81%</td>
<td>37%</td>
<td>0.45 (3)</td>
</tr>
<tr>
<td>Physical ability b</td>
<td>65%</td>
<td>16%</td>
<td>0.45 (5)</td>
</tr>
<tr>
<td>Pain b,c</td>
<td>69%</td>
<td>21%</td>
<td>0.43 (5)</td>
</tr>
<tr>
<td>Social support b</td>
<td>38%</td>
<td>10%</td>
<td>0.38 (5)</td>
</tr>
<tr>
<td>6 or more medications a</td>
<td>54%</td>
<td>23%</td>
<td>0.36 (4)</td>
</tr>
<tr>
<td>Possibly sick from medications a</td>
<td>54%</td>
<td>17%</td>
<td>0.31 (3)</td>
</tr>
<tr>
<td>Poverty b</td>
<td>42%</td>
<td>9%</td>
<td>0.30 (2)</td>
</tr>
<tr>
<td>3 or more chronic conditions</td>
<td>41%</td>
<td>18%</td>
<td>0.28 (5)</td>
</tr>
</tbody>
</table>

* Spearman’s correlation for 5 response categories of Overall Health with an item’s response categories.
  b Included in an 8-item comparator.
  c Included in the What Matters Index.
Quality-of-Life Measures for Medicare Patients

Figure. The AUROC represents the probability that a randomly chosen patient who reports negative quality is correctly rated by a combination of measures compared with a randomly chosen patient who had neutral or positive quality of life. When AUROC is 0.70 to 0.79 its discrimination is considered fair, and when 0.80 to 0.89, good. AUROC indicates area under the receiver operation characteristics curve; WMI, What Matters Index.

patients who used the hospital in the past year. Although they discriminate better than chance, none of the comparators is acceptable for predicting hospital utilization by individual patients. This finding, based on retrospective data, is consistent with prospective prediction of hospital and emergency care in Medicaid and private practice patients. Regardless of the measures or algorithms used for prediction, designations of being at risk for costly care are inherently inaccurate (Wasson et al., 2018a).

DISCUSSION

This brief report illustrates several different methods for assessing and improving, whenever possible, a patient’s QOL.

- Simply ask patients about their quality of life and inquire about what they would need to maintain or improve it. These efficient, direct queries bring to the fore a concern for every patient’s QOL.
- Administer a minimized few memorable measures for which a consistent remedy can be predefined (Wasson et al., 2018b). This efficient approach, represented by the WMI, can immediately address important patient needs and indirectly measure and monitor a patient’s QOL.
- Use longer multi-item measures that require additional scoring and interpretation before action can be taken. An association with QOL is likely if the right measures are chosen.

My colleagues and I previously documented better value—lower costs for similar results—of single-item compared with multi-item instruments for assessing practice quality and patient engagement (Ho et al., 2013; Wasson, 2013). This brief report illustrates broader advantages of parsimony in measurement of what matters to patients.
REFERENCES


