Health-Related Quality of Life After Stroke
A Comprehensive Review

Tammy O. Tengs, ScD; Michelle Yu, MPH; Elvina Luistro, BA

Background and Purpose—We performed a comprehensive review of all quality-of-life (QOL) estimates for stroke appearing in the peer-reviewed literature between 1985 and 2000. We examine variation in QOL weights and the rigor of methods used to assess QOL and discuss the implications for cost-utility assessment and resource allocation decisions.

Methods—Through a systematic search, we identified 67 articles that met our inclusion criteria. A team of trained researchers read each article and followed detailed guidelines to extract QOL weights and other parameters. This effort yielded 161 QOL estimates for stroke-related health states. All estimates were measured on a 0 to 1 scale, with 0 representing the worst outcome and 1 representing the best.

Results—QOL estimates range from −0.02 to 0.71 (n=67) for major stroke, from 0.12 to 0.81 (n=14) for moderate stroke, from 0.45 to 0.92 (n=38) for minor stroke, and from 0.29 to 0.903 (n=42) for general stroke. Although QOL should decrease with severity, there were many instances in which the QOL for major stroke as reported by one study exceeded the QOL for moderate stroke as reported by another. The same reversal was found for moderate and minor stroke, and it occurred even when both authors used similar assessment methods and subject populations. Authors of cost-utility and decision analyses rarely base their choice of QOL weights on their own primary data (19%). When obtaining weights from secondary sources, some authors (23%) chose QOL weights for a severity of stroke that did not match the severity for which they sought data.

Conclusions—QOL estimates for stroke vary greatly and are not always estimated in sound fashion. This impedes the comparability and quality of the cost-effectiveness studies that use these QOL weights and hampers good resource allocation decisions. (Stroke. 2001;32:964-972.)

Key Words: quality of life ■ review literature ■ stroke assessment

The number of people who survive stroke and live with its consequences is increasing. The case fatality rate for stroke has declined over the past few decades,1 and 85% of people who experience a stroke now survive.2 Today there are nearly 4 million people in the United States who have survived a stroke and are living with the sequelae.

This increase in survival has necessitated a new approach to measuring the health outcomes associated with stroke prevention, treatment, and rehabilitation. Survival rates, once a helpful measure of preventive or therapeutic success, are no longer adequate for the task. The quality-adjusted life-year (QALY) has emerged as a common metric useful for capturing both health-related quality of life (QOL) and length of life in the same measure.3 With QALYs, each year of life after a stroke is adjusted for its quality. For example, if major stroke is assumed to have a QOL of 0.3, then 6 years lived after a major stroke would be counted as 2 QALYs (because 6×0.3=2). If the stroke is prevented and the person lives, for example, 10 years in full health, or 10 QALYs (10×1=10), then the incremental public health gain from preventing the stroke might be assessed as 8 QALYS (10−2=8). If, in this hypothetical example, we had ignored QOL, we might have estimated that preventing a major stroke results in 4 additional years of life (10−6=4). Thus, the life-years gained (4) by preventing a major stroke often differ from the QALYs gained (8) because the latter includes the QOL implications as well as the gain in survival when a stroke is prevented. Although this example is for prevention, QALYs can also be used to quantify the value of any medical therapy.

QOL weights used in the calculation of QALYs are generally assessed in 1 of 3 ways. The first is direct elicitation. This typically involves asking patients or community members to subjectively rate the QOL of a health state on a 0 to 1 scale, with 0 representing a lower bound such as death and 1 representing an upper bound such as perfect health. An experimenter assessing QOL for minor stroke might begin by describing minor stroke as follows:

Received September 13, 2000; final revision received December 6, 2000; accepted December 19, 2000.
From the Health Priorities Research Group, University of California, Irvine.
Reprint requests to Tammy O. Tengs, ScD, Director, Health Priorities Research Group, School of Social Ecology, University of California, Irvine, CA 92697-7075. E-mail tengs@uci.edu
© 2001 American Heart Association, Inc.
Stroke is available at http://www.strokeaha.org

See Editorial Comment, page 972

http://stroke.ahajournals.org/ Downloaded from
Imagine that mild stroke has made your right arm and leg weak and a little hard to move. You can bathe or get dressed without help. You can also walk at a normal speed, but with a slight limp, and feed yourself without difficulty. You can think and write clearly. Even though your speech is a little slurred, people can easily understand you. You can have sex as usual and do most of your usual chores and sports, but sometimes with a little difficulty.

Alternatively, the experimenter might simply ask the respondent to "consider what life must be like following minor stroke" without offering a description. Either way, after characterizing the health state, an experimenter using direct elicitation methods might then use one of several elicitation techniques, such as the standard gamble, time tradeoff, or rating scale, to elicit a subjective estimate of QOL for that health state.

The standard-gamble method of direct elicitation is grounded in von Neumann-Morgenstern utility theory. It involves asking a respondent to make tradeoffs between a particular health state and a hypothetical gamble involving some chance of a better or worse outcome. For example, suppose that a respondent indicates that she is indifferent between a minor stroke and the following hypothetical gamble: a 0.8 chance of perfect health (which has a utility of 1) along with a 0.2 chance of death (which has a utility of 0). The standard gamble method assumes that if the respondent is truly indifferent between these 2 options, then the "expected utility" for these options is equivalent. Because the expected utility for the gamble is 0.8 \(\text{[calculated as } (0.8 \times 1) + (0.2 \times 0)]\), then the respondent’s indifference implies that their subjective utility for a minor stroke must also be 0.8. A different respondent might indicate indifference between minor stroke and the following gamble: a 0.7 chance of perfect health plus a 0.3 chance of death. This person would have an assessed utility for minor stroke of 0.7.

The time-tradeoff method of direct assessment involves asking a respondent to make hypothetical tradeoffs between a shorter life span in perfect health versus a longer life span in the health state in question. For example, if a respondent indicates that she is indifferent between living 90% of her remaining life span in perfect health versus living 100% of her remaining life span with the aftermath of a minor stroke, then this would imply that she assigns the health state of minor stroke a QOL weight of 0.9.

Finally, the rating-scale method of direct assessment involves asking the respondent to indicate where the health state lies on a 0 to 100 scale ranging from death to perfect health. The rating scale might be a category-rating scale divided into discrete intervals or a visual analog scale that is not divided into intervals.

The second general category of QOL assessment methods is indirect elicitation. To assess QOL indirectly, an experimenter would use an established scale such as the Health Utilities Index or the Quality of Well Being Scale. The experimenter would ask respondents to rate the stroke health state on several health-status dimensions, such as speech, cognition, or ambulation. These ratings are then linked or "mapped" to weights collected previously from community members for the same health-state dimension levels (although not for stroke specifically), and an overall QOL is derived.

The third QOL assessment method is judgment. This involves the experimenter or a panel of experts offering their opinion of the average QOL experienced by persons with stroke.

Regardless of how they are assessed, QOL weights are often used to weight life-years to calculate QALYs. The incremental gain in QALYs is then used as a measure of the effectiveness of an intervention. The primary advantage of the uniform QALY metric is that it allows decision makers to compare the relative value of seemingly incomparable medical and public health interventions for different diseases and conditions. QALYs have both theoretical and practical limitations, however. One theoretical limitation is that their use implies that 2 QALYs are twice as good as 1 QALY, and when one person gains 10 QALYs, this is equivalent to 10 people each gaining 1 QALY. Of course, this is a limitation only if this implicit assumption does not represent citizens’ or patients’ values. Note also that this “limitation” is not unique to QALYs. For example, even if we did not consider QOL and measured the value of an intervention simply by the gain in years of life, the same implicit assumption would apply. A theoretical limitation unique to QALYs is that they assume that the QOL a person assigns to a particular health state is independent of the length of time spent in that health state. Thus, if major stroke is assessed to have a QOL of 0.2, it is assumed that this same QOL would apply whether that person lived for 1 year or 5 years after major stroke. It is this assumption of “mutual utility independence” between length of life and QOL that allows QOL weights to be multiplied by life-years to calculate QALYs. Beyond the theoretical, a practical limitation of QALYs is that their accuracy depends entirely on the accuracy of the QOL and survival estimates on which they are based. Because of the unambiguity of death, mortality rates are fairly easy to measure. In contrast, QOL is much more difficult to measure.

Whatever their limitations, the QALY is now the measure of choice for cost-effectiveness analyses, and the US Panel on Cost-Effectiveness in Health and Medicine3 has recommended its use. Cost-effectiveness analyses that use QALYs are often called cost-utility analyses. There has been a large increase in the number of these analyses in recent years as payers have become increasingly interested in “value for money.” The ratio of the net direct healthcare costs of an intervention to the net QALYs saved by that intervention provides a helpful measure of value. Because QOL weights are used in cost-utility analyses, and because these analyses in turn affect resource allocation decisions, it is important that QOL weights be measured accurately.

The purpose of this review is to examine variation in QOL weights for stroke, the rigor of methods used to assess QOL, and the possible implications for cost-utility assessment and resource allocation decisions.

**Methods**

QOL weights for stroke were obtained from published journal articles. We used a variety of methods to identify articles that
First, we examined each of the stroke-related cost-effectiveness analyses, decision analyses, and QOL studies in our extensive files. We also searched the National Health Service (NHS) Economic Evaluation Database on the keywords “stroke” and “QALY” and examined the bibliographies of review articles. In addition, we searched MEDLINE from 1985 to 2000 for stroke-related articles that reported QALYs or contained other keywords such as time tradeoff, standard gamble, rating scale, or one of the health status instruments, such as the Health Utilities Index. Finally, when extracting data from each document, we noted citations for other articles that were likely to have stroke-related QOL data and retrieved those as well.

Our inclusion criteria for source documents were as follows: articles had to be (1) written in English, (2) peer reviewed, (3) full length, and (4) report QOL weights for stroke measured on a 0 to 1 or 0 to 100 scale. Articles reporting QOL weights for cerebrovascular diseases other than stroke were excluded, as were technical reports, which are often not peer reviewed, and conference abstracts, which are not full length. We also excluded articles that reported multidimensional health status (eg, SF-36, Sickness Impact Profile) when they did not report a final overall QOL estimate scaled in such a way that it met our inclusion criteria. In short, the QOL weights we surveyed had to be used or usable as part of a QALY measure.

We extracted QOL weights from articles using a 3-step process. First, 1 researcher read each document and entered 8 items directly into a database: (1) a description of the health state; (2) QOL weight; (3) method used to elicit or derive weights (eg, time tradeoff, Quality of Well-Being Scale); (4) respondent type (eg, patients, community members), sex, age, and sample size; (5) lower and upper bounds of the QOL scale (eg, death to excellent health); (6) whether the study was a QOL study, cost-utility analysis, or decision analysis; (7) whether the QOL assessments were primary (ie, original judgments or based on a sample of subjects) or secondary (ie, taken directly from or based loosely on another source); and, when available, (8) Rankin Scale value representing stroke severity. After the first researcher completed his or her work, a second researcher also read the document and reviewed the entered data. The 2 researchers then met to discuss any necessary revisions. Finally, the lead author reviewed the document and all data extracted and made additional revisions when necessary. We considered a procedure of completely independent extraction by 2 reviewers but rejected this procedure as unnecessary to maintain quality owing to the relative clarity and simplicity of the recording task and given that the lead author reviewed every article. To ensure consistency, the research team referred to detailed instructions and met weekly to discuss data extraction.

Of the 208 articles that we retrieved, 67 met our inclusion criteria. Of these, 13 articles focused primarily on the assessment of the QOL of stroke, 39 were cost-utility analyses of interventions to prevent or treat stroke, and 15 were decision analyses or other studies intended to inform medical decision making. Some documents contained more than 1 QOL estimate, so collectively these 67 documents reported 161 QOL estimates related to stroke.*

QOL weights for major stroke range from 0.02 (indicating that major stroke is worse than death) up to 0.71; QOL weights for moderate stroke range from 0.12 to 0.81, for minor stroke from 0.45 to 0.92, and for general stroke from 0.29 to 0.903. Figure 1 shows the frequency distribution of all QOL weights for each level of severity. The distributions for

*These 161 QOL weights, along with the corresponding health state, the method used to assess QOL, subject population, lower and upper bounds of the QOL scale, and study type, are available on request from the lead author.
Tengs et al  Quality of Life After Stroke  967

TABLE 1. Number of QOL Weights Elicited From Each Type of Respondent by Each Assessment Method

<table>
<thead>
<tr>
<th>Assessment Method</th>
<th>Respondent</th>
<th>Time Tradeoff</th>
<th>Standard Gambel</th>
<th>Rating Scale</th>
<th>Judgment</th>
<th>Health Status Instrument</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author</td>
<td>0</td>
<td>0</td>
<td>68</td>
<td>0</td>
<td>3</td>
<td>71 (44%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experts</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4 (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community</td>
<td>8</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>3</td>
<td>32 (20%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients</td>
<td>8</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>17 (11%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients and Community</td>
<td>17</td>
<td>9</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>28 (17%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patients and proxy</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4 (2%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not stated</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5 (3%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>42 (26%)</td>
<td>17 (11%)</td>
<td>20 (12%)</td>
<td>70 (43%)</td>
<td>9 (6%)</td>
<td>3 (2%)</td>
<td>161 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

QOL varies depending on the elicitation technique used. Figure 2 shows the variation for major stroke. The median estimates elicited were 0.51 with the standard-gamble method, 0.32 with time tradeoff, 0.23 with rating scale, and 0.39 with judgment. There was also considerable variation in QOL estimates obtained by different authors for the same health state using the same technique. This occurred even when other variables, such as the age of sample, the size of sample, the upper and lower bounds of the scale, and the type of respondents, were the same. For example, Samsa et al used the time-tradeoff method to elicit QOL estimates for major stroke and found a mean of 0.36, but Hallan et al used the same method and found a mean of 0.51.

Some authors described the severity of the stroke health state using the Rankin Scale. Rankin Scale values range from 0 (stroke with no symptoms) to 6 (fatal stroke). Because the score is numerical, we were able to calculate the correlation between Rankin severity and assessed QOL weight. The Pearson correlation was −0.74 (n=23), and the Spearman rank order correlation was −0.72. Squaring the first correlation revealed that 55% of the variance in QOL was explained by the severity of the health state.

QOL accuracy is especially important when weights are used as part of a QALY measure in an analysis intended to inform healthcare decisions (see Table 2). To better understand the quality of the data used in this context, we looked more closely at the 104 QOL weights used in cost-utility or decision analyses. We found that 20 of these estimates were based on primary data collected by the author of the analysis, 19 were based on the author’s own judgment, 34 were based loosely on 1 or more secondary sources, and 31 were taken directly from a secondary source.

In the 31 cases in which the author obtained the estimate directly from a secondary source, we retrieved and checked that source to assess the characteristics of the QOL information that it reported. The results are summarized in Table 3. Generally, the secondary source reported QOL estimates based on data (63%), but it was not uncommon for the cited source to contain only the judgment of the secondary author (37%). We also found that 74% of secondary estimates were for stroke of same severity, 23% were for stroke of different severity, and 1 QOL estimate for stroke was based on a QOL estimate for myocardial infarction. For example, Patel et al obtained their 0.4 QOL estimate for major stroke from Solomon et al. Solomon et al reported high-quality data for mild (0.5), moderate (0.4), and severe (0.08) stroke. Patel et al chose the Solomon figure for moderate stroke, and thus severity did not match in this case.

The 67 studies we reviewed varied in the specificity with which authors described the stroke health state and the degree of detail given about the QOL elicitation procedure. As noted earlier, 26% of health states were described as “stroke” with no indication of the severity of stroke. In 12% of cases, QOL was assessed for a particular time in the course of the disability, such as 3 months, 1 year, or 10 years after the event, but in the remaining 88%, no information about the timing was provided. Twenty-one percent of studies reported the age composition of the subject population from whom QOL weights were elicited, and 16% reported the sex composition. However, only 6% reported separate QOL weights for subgroups according to age or sex.
Discussion

The present report represents the most comprehensive review of QOL weights for stroke published to date. It reveals enormous variation in assessed QOL for the same health state and considerable variation in the rigor of methods used to estimate QOL. Because QOL weights are used in QALYs to characterize the effectiveness of various interventions for stroke, resource allocation decisions based on cost-utility analyses might well be affected. For example, when Derdeyn and Powers varied the QOL for stroke from 0.4 to 0.8, the cost-effectiveness ratio for carotid atherosclerotic screening went from $25,045 to $58,811. Because $50,000 is the threshold that is commonly chosen to separate interventions that are “worth” the cost from those that are not, the decision about whether to recommend the intervention was sensitive to the QOL estimate.

Furthermore, overestimation of QOL weights for stroke can lead to underinvestment in stroke prevention and treatment relative to other diseases. This is because when QOL weights for stroke are too high, the QALYs lived after a stroke will be overestimated. As a result, the value of preventing or treating stroke will appear to be less than it is. Whereas the overestimation of stroke QOL can lead to the underestimation of the value of stroke care relative to care for other diseases, the misestimation of stroke QOL can lead to skewed priorities within the area of stroke care. For example, when one author characterizes major stroke with a QOL

TABLE 2. Source of QOL Weights

<table>
<thead>
<tr>
<th>Source</th>
<th>Study Type</th>
<th>Cost-Utility or Decision Analysis</th>
<th>QOL Study</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary data collected by author</td>
<td></td>
<td>20</td>
<td>57</td>
<td>77 (48%)</td>
</tr>
<tr>
<td>Judgment made by author</td>
<td></td>
<td>19</td>
<td>0</td>
<td>19 (12%)</td>
</tr>
<tr>
<td>Based loosely on secondary source</td>
<td></td>
<td>34</td>
<td>0</td>
<td>34 (21%)</td>
</tr>
<tr>
<td>Taken directly from secondary source</td>
<td></td>
<td>31</td>
<td>0</td>
<td>31 (19%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>104 (65%)</td>
<td>57 (35%)</td>
<td>161 (100%)</td>
</tr>
</tbody>
</table>

TABLE 3. Rigor and Consistency of QOL Weights Taken Directly From a Secondary Source

<table>
<thead>
<tr>
<th>Consistency</th>
<th>Secondary Source</th>
<th>Data</th>
<th>Author Judgment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke of same severity</td>
<td>Stroke of same severity</td>
<td>13</td>
<td>10</td>
<td>23 (74%)</td>
</tr>
<tr>
<td>Stroke of different severity</td>
<td>Stroke of different severity</td>
<td>6</td>
<td>1</td>
<td>7 (23%)</td>
</tr>
<tr>
<td>Nonstroke health state</td>
<td>Nonstroke health state</td>
<td>1</td>
<td>0</td>
<td>1 (3%)</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>20 (63%)</td>
<td>11 (37%)</td>
<td>31 (100%)</td>
</tr>
</tbody>
</table>
weight that exceeds the weight used by another author for moderate or minor stroke, then the relative ordering of any resulting cost-utility analyses could be affected. An intervention that is actually more cost-effective than another could appear to be less cost-effective. With healthcare decisions based increasingly on value for money, this could lead to poor coverage decisions that might ultimately hurt patients.

Arguably, the severity of stroke, defined by morbidity, should be the primary factor accounting for variation in QOL weights. When we compared QOL weights to Rankin scores, however, we found that severity accounted for only half of the variance in QOL. Furthermore, when we compared the complete distributions of QOL weights according to severity, we found a surprising similarity between the distributions for minor stroke and general stroke. This is problematic because the general category of all stroke presumably includes some fraction of major and moderate strokes as well. It follows logically that general stroke should have a much lower QOL distribution than minor stroke. We also found remarkable similarities between the distributions of QOL weights for moderate and major stroke. If it is genuinely worse to have a major stroke, then the differential should probably be greater than that observed.

The rigor of methods used to assess QOL is uneven. Although respected techniques such as the time tradeoff or standard gamble are often used, a large number of QOL estimates in the literature are based simply on author judgment. We also found that it is surprisingly common for authors to base their judgments of QOL not on data but on the judgments of other authors. Presumably, analysts make this choice out of the desire to ensure that their cost-utility ratios are easily compared with others in the literature. Unfortunately, basing judgments on judgments might also serve to perpetuate inaccuracies in QOL. This practice may have been more acceptable before 1993, when there were few empirical QOL studies on which to base judgments, but now there are a number of fine studies\(^7\) with large samples that might offer more accurate estimates. Regardless of whether an estimate is based on secondary data or a secondary judgment, it is important for the author of a cost-utility analysis to choose a QOL weight appropriate for the severity of stroke. Too often, estimates that were originally derived for stroke of one degree of severity are used later for another degree of severity.

A potential contributor to variation in QOL is heterogeneity in the upper and lower bounds of the QOL scale used to elicit weights from respondents. For example, a study participant invited to provide subjective QOL for minor stroke might give a higher answer if the scale is bounded at the top by normal health than if it is bounded by perfect health. This is because minor stroke is presumably closer to the former than the latter. We found that authors used a total of 15 different classes of upper bounds, including normal, good, full, perfect, usual, excellent, current, and best health. We also found 5 classes of lower bounds, including death, worst possible health state, and limitations in activities of daily living, with poor self-related health.

There are also a number of limitations in the way studies report QOL weights. Authors of primary studies rarely report the sex or age composition of their subject groups. Furthermore, those that do generally do not report subgroup means for stroke QOL. With increasing national interest in learning the relative cost-effectiveness of medical interventions in different populations, it might be important in the future to have these specific weights. For example, some studies\(^77,78\) have shown that estrogen can serve as an endogenous neuroprotective agent to prevent stroke in postmenopausal women. In assessing the cost-effectiveness of estrogen, it would be best to have QOL weights for stroke specific to women. This is particularly important because some evidence suggests that women may assign a lower QOL to major stroke than men.\(^79\)

In addition to omitting information about the sex and age of subject populations, authors collecting primary QOL data rarely report the timing of the assessment relative to the stroke event. As a consequence, authors of cost-utility and decision analysis invariably use the same constant QOL weight over all years after stroke. This would be appropriate if 1 of 2 assumptions were true: either QOL does not change over time, or QOL does change but the weight is right on average. The first assumption can be challenged as suspect because there is evidence that among those who survive stroke, morbidity often improves with time.\(^79–83\) The second assumption would be reasonable if the QOL weight used in the analysis were obtained from a sample of patients who were well distributed according to time since event, so that any weights derived would be averaged across time. Because authors of primary studies rarely mention the timing in the course of the disease, this is difficult to determine.

When QOL weights are derived from samples, authors generally report the mean, or sometimes the median, of that sample. Only 43% of the primary studies that we reviewed reported the range or variance. If analysts are to ever combine similar estimates with meta-analysis, information about the variance will be important. Furthermore, because individuals have differences in their subjective feelings about the QOL of stroke, the mean does not give the full picture. Samsa et al,\(^17\) for example, reported that the mean QOL for major stroke was 0.3. However, 45% of their sample indicated that death was preferable to life after major stroke, so that the assessed QOL was less than 0. It is important to report, as Samsa et al did, both the QOL mean and information about the distribution. The omission of information about variability can be important when it leads to resource allocation decisions that put at a disadvantage those with different, but equally legitimate, subjective attitudes toward the QOL of stroke.

This review reveals enormous variation in QOL weights for stroke. We also found large variation in the rigor of methods used to assess QOL. When authors of cost-utility analyses use different QOL weights for the same health state, the resulting ratios become incomparable. Similarly, when authors make unwise decisions in choosing secondary QOL weights, the resulting cost-utility ratios become inaccurate. Incomparability and inaccuracy might affect decisions that could, in turn, adversely affect patient health. A large national repository of rigorously assessed QOL weights for stroke and other diseases, as has been suggested by the US Public Health Service Panel on Cost-Effectiveness in Health and Medicine,\(^3\)
would be a welcome contribution. In the meantime, authors of cost-utility and decision analyses should exercise sound judgment in choosing the QOL weights on which they base their analyses.

Acknowledgments

Part of this research was funded by the California Tobacco-Related Disease Research Program (grant number 6P7-3005). In addition to the coauthors, Karen Lee, Amy Wallace, Kari Parsons, Edward Reynaud, and Sharon Brown helped with earlier efforts to code QOL estimates related to stroke, and we appreciate their assistance. Laurie Chen, Nora Williams, Jenifer Maple, and Arnold Liao contributed to the development of the QOL extraction protocol. Benson Chen, Chris Nguyen, Joy Chen, Penny Liu, Linda Torossian, and Jeff Kwan gathered the articles. Finally, we appreciate the efforts of the talented group of authors who performed the original studies reviewed herein.

References


52. Danese MD, Powe NR, Sawin CT, Ladenson PW. Screening for mild thyroid failure at the periodic health examination: a decision and cost-effectiveness analysis. JAMA. 1995;276:139–146.


Utility assessments and resource allocations are often based on quality-adjusted life-years (QALY). QALY estimates have increased in popularity because they provide a common metric that is useful for summarizing both health-related quality of life (QOL) and length of life in the same measure. These estimates allow decision makers to compare the effectiveness of different interventions. Tengs and colleagues reviewed the prior use of QOL estimates that are used in the calculation of QALY and the problems associated with these estimates. These problems are particularly important given that QALY estimates are used for stroke resource allocation decisions.

The present review by Tengs et al clearly documents that QOL estimates vary greatly by stroke severity, such that there are many instances in which the QOL estimate for major stroke as reported by one study exceeded, for example, the QOL estimate for moderate stroke as reported by another. An additional problem revealed in the review by Tengs et al is that a majority (81%) of cost-utility authors used QOL estimates from secondary sources and potentially did not use appropriate QOL weights for their particular study sample.

A few additional issues need to be considered when QOL estimates are used for cost-utility analyses. First, although QALY values provide uniform metrics, QOL estimates obtained from stroke patients, community members, and experts are known to vary. Patients who have not experienced a stroke (see Solomon et al) or individuals at risk for future stroke (see Samsa et al), for example, respond with low QOL estimates for physical impairments. Yet it is clear that patients who actually experience a high level of impairment as a result of a stroke provide high estimates of their QOL. It is recommended that researchers and policy makers be careful not to overgeneralize the use of QOL weights. QOL estimates obtained from stroke patients, for example, should only be used to calculate QALY estimates for stroke patients, not community members, and vice versa.

Second, it is troubling that QOL estimates are least likely to be obtained from the patients themselves than from any other source. Health-related QOL has been defined as a multidimensional conceptualization that includes the physical, mental, and social aspects of health (see Sherbourne et al and Bosworth et al) and requires a subjective rating. Given the subjective nature of QOL, it is unclear how QOL estimates of experimenters or a panel of experts would generalize to either a stroke patient or a community member’s QOL perceptions.

Third, among those patients who may have experienced a stroke, investigators need to consider the time between the event and assessment. It may be that as the stroke patient recovers or adapts to a given level of physical function, the associated QOL is likely to increase. If QOL varies with time, the point at which the utility is solicited could greatly affect the “optimal” treatment strategy identified.

Fourth, even after adjustment for physical functioning and disease severity, significant depressive symptoms and decreased social support are related to lower health-status utility. These results suggest that besides simply considering disease severity, both physical and psychological health need to be considered in decision modeling of stroke practice patterns and outcomes. Similarly, although research has been conducted to examine the relationship between QOL and both demographic factors (such as race, sex, and age) and socioeconomic status (eg, education and income), these factors are rarely considered in QALY calculations.

These problems with QOL estimation can have large financial and resource implications. Varying the QOL value for stroke, for example, from 0.4 to 0.8 changes the cost-effectiveness ratio for carotid atherosclerotic screening from about $25 000 to $38 000. Thus, incorrectly estimating QOL values may lead to inappropriate allocation of resources for prevention and treatment of stroke.

Tengs and colleagues should be commended for summarizing the current state of the use of QOL estimates for QALY measures and subsequent cost-effectiveness analyses for stroke. This literature remains quite problematic, and researchers are urged to use primary data collected for their particular purpose. When primary data are unavailable, researchers are encouraged to (1) look to data that come from similar populations under investigation in terms of patients’ disease severity and demographic and socioeconomic characteristics, (2) not overgeneralize QOL weights, and (3) consider the elapsed time between the stroke and QOL assessment. These recommendations will improve the validity of QALY estimates.

Hayden B. Bosworth, PhD, Guest Editor
Health Service Research and Development
Durham Veterans Affairs Medical Center
Departments of Medicine and Psychiatry
Duke University Medical Center
Durham, North Carolina

References