Measuring Quality of Life in Stroke

R. de Haan, MS; N. Aaronson, PhD; M. Limburg, MD, PhD; R. Langton Hewer, MD, PhD; and H. van Crevel, MD, PhD

Background and Purpose: Little attention has been focused on quality of life in stroke outcome research. The purpose of this review is to outline the meaning of the concept, describe important methodological issues and methods of assessment, review existing quality of life measures, and discuss criteria for selecting an appropriate instrument.

Summary of Review: The following 10 quality of life instruments were reviewed: COOP Charts; Euroqol; Frenchay Activities Index; Karnofsky Performance Status Scale; McMaster Health Index Questionnaire; Medical Outcomes Study 20-Item Short-Form Health Survey; Nottingham Health Profile; Quality of Life Index; Quality of Well-being Scale; and the Sickness Impact Profile. They were evaluated in terms of length, time needed to complete, content, scoring, and psychometric characteristics.

Conclusions: Emphasis should be placed on further psychometric evaluation of existing quality of life measures rather than on generating new instruments. There is particular need for supplementary data on the responsiveness of the instruments to changes in patients' clinical status over time. The choice of a suitable quality of life instrument should be based not only on psychometric properties but also on careful consideration of the research question, the relevance to the objectives of the study, the feasibility of the instrument, and the specific characteristics of the stroke patients under investigation. (Stroke 1993;24:320–327)

KEY WORDS • quality of life • stroke outcome

Despite the growing interest in quality of life (QL) issues in clinical research and practice, little attention has been paid to evaluating systematically the QL of stroke patients.1 QL investigations are useful not only in gaining a better understanding of patients' reactions to illness and for enhancing supportive care2 but also for evaluating the efficacy of therapeutic interventions. Although there is an association between neurological deficits and QL, they are not synonymous. If, with any given intervention, a clinical benefit has been obtained as demonstrated in terms of improved neurological function, an evaluation of the effects on patients' daily functioning, subjective health, and well-being is still highly relevant.

The objectives of this paper are 1) to describe the QL concept as it is applied in health care settings; 2) to outline the main methodological problems in measuring QL; 3) to review existing QL measures; 4) to identify the most relevant criteria for selecting suitable instruments; and 5) to propose directions for future QL studies in the field of stroke research.

The Concept of Quality of Life

Over the years QL has been defined in many different ways. One speaks, for example, of need satisfaction,3 health-related subjective experiences,4 or psychosocial and physical well-being.5 One way of resolving the problem of definition is to leave this task to the individual patient by simply asking the question: “How would you rate your present QL?”6–8 A single-item QL measure, however, is in itself not very reliable or valid and is of little analytic value. More specific information is required to interpret global ratings of a concept as complex as QL. For this reason, most researchers today adopt a multidimensional approach to QL assessment.9

A broad consensus has emerged that at least four dimensions should be included in a QL assessment: physical, functional, psychological, and social health. The physical health dimension refers primarily to disease-related and treatment-related symptoms. Functional health comprises self-care, mobility, and physical activity level, as well as the capacity to carry out various roles in relation to family and work. Cognitive functioning, emotional status (especially poststroke depression), and general perceptions of health, well-being, life satisfaction, and happiness are the central components of the psychological life domain. Social functioning includes the assessment of qualitative and quantitative aspects of social contacts and interactions.

These four QL domains are also reflected, in part, in the International Classification of Impairments, Disabilities, and Handicaps (ICIDH) of the World Health Organization.10 Impairments reflect organ dysfunctions or abnormalities of body structure. Disabilities refer to the consequences of impairments in terms of the pa-
tant’s functional performance. Handicaps are concerned with the social disadvantages resulting from impairments and disabilities. This overlap between QL and the organ, person, and societal levels of the ICIDH is to be expected because both are closely linked to the World Health Organization definition of health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.”

Various studies of QL in stroke indicate that physical disabilities often have a negative impact on QL, but this is not always the case. Stroke patients with little or no physical dysfunction can also experience compromised QL.17-19 Psychological status appears to be as important as physical disability in altering an individual QL.

In summary, we stress the importance of conceptualizing QL as a broad spectrum of consequences of disease, including elements of impairments, disabilities, and handicaps, as well as patients’ perceived health status and well-being.

Methodological Issues

QL measures can be differentiated into 1) generic scales or health profiles, 2) disease-specific scales, and 3) scale batteries.15,16 Generic scales are not developed for a specific target population and may be suitable for use with many patient populations. A major strength of these generic scales is the possibility of detecting the relative effects of disease and treatment on different life domains. Another advantage is that they allow comparison of QL results across patient populations. Generic scales, however, are not always sufficiently focused on the specific problems of any given patient population.

Disease-specific QL scales exist for stroke,17 rheumatic disorders,18 cardiovascular diseases,19 and cancer.20 Such scales do not allow cross-disease comparisons but are often more sensitive to the QL issues particularly relevant to a specific population of patients.

A battery of scales for measuring particular dimensions or aspects of health can also be used, such as instruments to assess cognitive functioning, depression, activities of daily living (ADL), or social functioning. The use of such a series of specific and often unidimensional measures allows one to assess each relevant life domain in depth. This approach has the disadvantages that results may not be comparable across studies and that the patient burden associated with completing such a battery may be unacceptably high.

Most of the available QL instruments depend on patient self-reports. The data can be collected by either structured interviews or written questionnaires. These methods of data collection are less suitable for patients with serious cognitive, speech, and language disorders. It is not surprising, therefore, that in stroke outcome studies patients with severe impairments have often been excluded from QL assessments.12,13,20-22 This creates the problem of not evaluating QL in a subgroup of patients in which this is highly relevant. An alternative approach is to ask physicians to rate their patients’ QL. However, this may create methodological difficulties. Some studies have shown low levels of interobserver reliability as well as disagreement between ratings provided by physicians and those of patients.23-26 There are indications that patients’ ratings of their overall QL reflect a broad range of physical, emotional, and social health concerns, whereas physicians’ ratings are more typically restricted to the physical health status of the patient.26

The QL of stroke patients with communication disorders can also be assessed by significant others such as partners or children. There are two variants of such measurements. The most straightforward is to ask the caregiver to rate the QL of the patient. The alternative, the so-called proxy measurement, involves asking the respondent to answer the questions as he or she thinks the patient would. Evidence regarding the quality of caregiver reports is mixed.27 Some suggest that they are reliable and valid.28 Others find evidence that the social network of the patient underreports health complaints,29 satisfaction, and emotional health.30 Studies on the interchangeability of patient and proxy responses, by definition, can be based only on score agreement between communicative patients and their caregivers. Thus, even in the case of complete concordance one cannot necessarily generalize the findings to the target population of noncommunicative patients.

Surprisingly few QL studies have included control groups or other normative data in their designs or analysis.13,14,31 In stroke research, the use of age-matched control groups should be viewed as an essential component of a QL investigation. Such a control group allows one to distinguish between QL effects related to the disease and its treatment and those attributable to the aging process per se.

Alternative approaches to QL assessments involve the use of utility measures. Utility refers to the value attached to specific levels of health. For clarity of presentation we will focus only on the use of utility scales in clinical outcome studies (see References 32 and 33 for a discussion of utility analysis in clinical decision making). Utility scales, as a special class of generic instruments, focus on a broad range of life domains. The scales classify patients into various health states. The utilities of these health states typically have been assessed previously by a representative sample of the general public, by patient samples, or by a panel of experts and are expressed as a single numeric value on a scale ranging from 0.0 (death or worst possible outcome) to 1.0 (complete health or best possible outcome). These aggregated and weighted utility scores can be used as an evaluative end point in clinical trials. An attractive feature of the utility approach is that it yields a single summary score. This facilitates comparison of results across studies and the performance of cost–utility analyses. In these latter economic studies, survival data are combined with health ratings to yield quality-adjusted life years.34 This index reflects the gained years of life resulting from intervention, corrected by the utility of that life. By comparing alternative treatments on their costs and effects, cost–utility analysis can provide a useful basis for allocation of scarce health care resources.35,36

Various methodological problems in utility measurements have yet to be resolved. Torrance33 argues that individuals differ considerably in their health state preferences but that group mean utility values, required in most applications, are fairly stable. However, research is needed to clarify differences in utility values generated by different measurement methods. Moreover, differences in valuation of health states by different populations such as physicians, patients, and the
general public, as well as the agreement of utility measures with other QL measures, merit further study.37

Ultimately, the decision as to whether to use a more classic, psychometric approach (e.g., questionnaires) or a utility approach to QL assessment depends on the specific research question being addressed. If one is interested primarily in generating a detailed picture of the impact of disease and treatment on patients' QL, then a multidimensional classic approach may be the most appropriate. If, on the other hand, the intent is to adjust survival data or to generate cost–utility information, then the utility approach to QL assessment would be preferred. Of course, the use of one approach does not preclude the other. With sufficient resources, both approaches could be incorporated within a single study.

Review of Selective Quality of Life Instruments

By means of a Medline search (key words: cerebrovascular disorders, stroke outcome, quality of life, health status, reliability, validity), reference tracing, and a recent bibliography,38 we identified nine generic QL measures that could be used in stroke research. These include the COOP Charts39,40; the McMaster Health Index Questionnaire41–45; the Nottingham Health Profile46–50; the Sickness Impact Profile51–56; the Medical Outcomes Study 20-Item Short-Form Health Survey57–59; the Karnofsky Performance Status Scale;60,61 the Quality of Life Index62–65; the EuroQol66,67; and the Quality of Well-being Scale.68–71 Only one stroke-specific QL measure, the Frenchay Activities Index,17,72 was located.

Each of these measures has been evaluated in terms of its reliability and validity. Two types of reliability can be evaluated: homogeneity (or internal consistency) and stability. Homogeneity refers to the statistical coherence of the scale items and is commonly assessed by the Cronbach’s α coefficient.73 This coefficient is based on the average correlation of items within a scale and reflects the extent to which the items measure a common entity. The stability of a self-rating scale can be assessed by test–retest reliability, which is based on the measurement of the same person on two occasions with the same instrument. Interobserver reliability refers to the score agreement between different independent observers measuring a clinical phenomenon with the same instrument at the same time.

Validity reflects the degree to which a scale measures what it is intended to measure. Three types of validity can be distinguished: content, criterion, and construct.71 Content validity implies a nonstatistical, subjective evaluation of the extent to which a measure covers the substantive issues of interest. For this review the content validity of the measures has not been evaluated. Rather, we have simply described the QL domains included in the scales. For criterion validity one examines the relation between QL scores and some superior criterion of QL. Because no gold standard exists for QL, this type of validity data is seldom available. More common are studies of the construct validity of a measure. Support of construct validity is provided if instruments purported to assess the same QL domains correlate significantly with one another (convergent validity), whereas instruments intended to assess different QL domains exhibit low correlations (discriminant validity).

Finally, the clinical validity of an instrument can be assessed by examining the extent to which QL scores 1) are able to distinguish between patients with different diagnoses and between patients with differing degrees of disease severity within a given diagnosis; and 2) are able to detect clinically important changes in patients’ health status over time. Although some authors consider this latter issue of responsiveness to be distinct from validity,74 others argue that it is simply another aspect (albeit an important one) of instrument validity.75

Table 1 provides a relatively detailed survey of each of the 10 QL measures, including 1) number of items, average time to complete, and self-completion versus observer rating; 2) QL domains assessed and scoring; and 3) psychometric data (i.e., reliability and validity). In the following paragraphs some global remarks will be made regarding each of these measures.

The COOP scale comprises nine simple charts that are visually appealing because of the use of pictures. The scale was developed to routinely measure patients’ perceived health and physical, emotional, and social functioning in a busy office practice. Studies have demonstrated its potential for generating new information on patients and for influencing communication and management in general practice and outpatient settings. Although the COOP scale is suitable in clinical situations, the use of one item to measure each life domain may limit its analytic value in stroke outcome studies.

The McMaster Health Index Questionnaire (MHIQ) has 59 items measuring some physical symptoms (hearing, sight, headache), mobility, disabilities, quantitative aspects of social relations, social activities, and emotions (e.g., feelings of hopelessness and loss of control). Some characteristics of the MHIQ, such as the use of item skip patterns in the case of inapplicable questions and the use of different response categories, may be confusing for less-educated patients or patients with cognitive deficits. Not all reliability studies in terms of homogeneity and test–retest stability are convincing.

In comparison to the MHIQ, the Nottingham Health Profile (NHP) is somewhat shorter and can be administered more quickly. The NHP consists of two parts. Part one primarily addresses serious health problems and particularly focuses on levels of disability, sleep disturbances, and subjective feelings of energy, pain, emotional distress, and qualitative aspects of social relations. The seven items of part two refer to the effects of health on seven life areas (e.g., social life, vocational function). The use of such a series of single items, in combination with the dichotomous response categories, restricts the range of possible scores and thus may be of limited analytic value. The subscales are in need of further validation.

The Sickness Impact Profile (SIP) is a well-evaluated 136-item measure organized into 12 subscales. The subscales assessing communication, cognitive alertness, emotional behavior, and social functioning may be particularly relevant for use in stroke outcome studies. The physical dimension contains items measuring a broad range of ADL, mobility, and complex physical activities. Floor effects and especially ceiling effects, which are common to most specific disability scales, are therefore less likely. (Disability scales often address
only a narrow range of basic daily functions. As a result of this restricted range, health improvement or deterioration may not be detected because these changes can occur beyond the end points of the scale.) Because of its length the SIP is primarily suitable for cross-sectional studies. Since reliability data indicate poor reproducibility, it is recommended not to analyze item responses but to use aggregated category scores for description and analysis. The SIP defines QL only in behavioral terms, while other measures also incorporate aspects of subjective health. When using such a purely behavioral measure it may be advisable to add questions on feelings of well-being, health, happiness, and life satisfaction.

The Medical Outcomes Study 20-Item Short-Form Health Survey (MOS-20) is intended for use in large-scale outcome studies. The MOS-20 emphasizes mental health, health perception, and physical functioning. The mental health scale includes clearly defined items on distress and mood. In comparison with the other QL instruments, much attention is paid to general health perceptions and subjective evaluation of health status. The physical functioning scale primarily assesses complex physical activities. Only one item refers to ADL. Since limitations in self-care are important in stroke, this may be problematic in stroke outcome studies. When using the MOS-20 it is therefore worth considering the addition of a specific disability scale, such as the Barthel ADL Index.37 Future studies should establish the usefulness of the MOS-20 in stroke populations. Evaluations of a revised, 36-item version of the MOS instrument are currently underway.78

The Frenchay Activities Index (FAI) has been developed specifically for use with stroke patients. The FAI measures lifestyle in terms of more complex physical activities and social functioning. The index is easy to complete and can be administered in a few minutes. The item weights take potential sex differences in the scores into account. If one is primarily interested in measuring outcomes with an emphasis on disabilities, the FAI may be quite appropriate. However, additional studies of the reliability and validity of this instrument are necessary. The Karnofsky Performance Status Scale (KPSS) is a rating scale scored by an observer or interviewer. The KPSS was originally designed as an outcome measure in cancer research but has also been used in other chronically ill populations, including stroke.62 The ratings are narrowly based on physical disabilities and need for care. The measure incorporates death as a relevant end point. This is an important measurement feature, particularly in longitudinal stroke studies. If death is omitted, an increasing rate of mortality may be paradoxically related to a higher level of QL in the group of survivors.71 The scale lacks a standardized observational procedure, although some suggestions in this regard have been made.60

The QL-Index was developed to assess QL in cancer or other chronic diseases and encompasses the major life domains. Although this five-item scale, like the COOP Charts, is less precise, it may be useful in stroke research when one lacks resources for a more comprehensive patient-based evaluation or when the assessment of QL is not the main focus of the study. The reproducibility and the validation of the ratings with patients’ self-reports need further attention. When using the QL-Index it is advisable to standardize the interview or observational procedure.

The Euroqol is a utility scale that is still under development. The scale comprises five items to classify patients into various health profiles. The ordinal response scales are somewhat imbalanced, and the scale lacks the possibility of scoring the level of disability of patients in wheelchairs. Currently, the utilities of only some of the possible health profiles are assessed.

The Quality of Well-being Scale (QWBS) is probably the most widely used utility scale. The utilities of health states and symptoms of the target patient group, measured respectively on three functional scales and a symptom/problem complex, have been assessed in open populations and in various (nonstroke) patient groups. The functional scales focus on disabilities and social functioning. The problem/symptom complex reflects the physical and, to some extent, the emotional dimension of QL and includes relevant stroke outcomes (e.g., cognitive dysfunctions, communication disorders, emotional distress, paralysis). Similar to the KPSS, the QWBS incorporates death as an end point. Although the scale requires extensive interviewer training, it has been used in a wide range of settings. Additional studies of reliability, however, are recommended.

To summarize, with the exception of the QL-Index, the MOS-20, and, on the level of aggregate scores, the SIP, the QL measures reviewed are in need of further evaluation of their reliability and validity. Furthermore, with the exception of the QL-Index, none of the scales reviewed has been satisfactorily evaluated for its responsiveness to change in patients’ clinical status over time. Although evidence of an instrument’s responsiveness may be relatively unimportant in cross-sectional descriptive studies, it is critical in evaluating the appropriateness of a given QL measure for use in prospective clinical trials in stroke.

Conclusions and Recommendations

In this article we have reviewed nine generic QL measures and one stroke-specific instrument. Although it was not possible to incorporate all the literature pertaining to these instruments into our review, we are nevertheless confident that the studies reviewed provide a representative picture of the strengths and weaknesses of the QL measures.

A well-considered choice of suitable instruments should be based not only on the psychometric properties of the measures but also on the specific research questions, the relevance to the objectives of the study, the available resources, and the specific characteristics of the stroke patients under study. In selecting a suitable QL instrument some trade-off between the level of detail provided and feasibility seems to be inevitable. There are measures, such as the SIP and the MHIQ, that assess a wide range of QL domains in depth. However, in selecting a lengthy instrument one also has to consider its feasibility in terms of the available resources as well as patient and staff burden. The shorter instruments, such as the QL-Index, the COOP Charts, and the MOS-20, provide less detail but may be more feasible for use in prospective clinical trials. If the objective of the study is to assess general perceptions of health and well-being, measures such as the SIP and the FAI, which focus on the behavioral performance of the
<table>
<thead>
<tr>
<th>Scale</th>
<th>Length/time (min)/rater</th>
<th>Domains/scoring</th>
<th>Reliability</th>
<th>Clinical validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>COOP Charts(^{39,40})</td>
<td>9 items/2–3/patient or interviewer</td>
<td>Physical, emotional, social, role, global health, health change, pain, overall QL, social support/3 scale scores</td>
<td>NR</td>
<td>±</td>
</tr>
<tr>
<td>McMaster Health Index Questionnaire(^{41–45})</td>
<td>59 items/20/patient</td>
<td>Functional, emotional, social/3 summed subscale scores; standardized index value (item weights)</td>
<td>±</td>
<td>±</td>
</tr>
<tr>
<td>Nottingham Health Profile(^{46–50})</td>
<td>45 items/5–15/patient</td>
<td>Part 1: Physical, emotional, social, pain, energy, sleep/6 summed dimension scores (item weights) Part 2: Effects on 7 life areas/scoring not clear</td>
<td>... (Part 1)+ (Part 2)±</td>
<td>NR − +</td>
</tr>
<tr>
<td>Sickness Impact Profile(^{51–56})</td>
<td>136 items/20–30/patient</td>
<td>Ambulation, mobility, body care, social, emotional, communication, alertness, sleep, eating, home management, recreation, employment/12 summed category scores; 2 summed dimension scores; 1 summed total score (item weights)</td>
<td>+ (Aggregate)+ (Item level)±</td>
<td>NR ± +</td>
</tr>
<tr>
<td>Medical Outcomes Study(^{57–59})</td>
<td>20 items/3–4/patient</td>
<td>Physical, emotional, social, pain, role, global health/6 summed scale scores (item weights)</td>
<td>±</td>
<td>±</td>
</tr>
<tr>
<td>Frenchay Activities Index(^{77,72})</td>
<td>15 items/5/patient or caregiver</td>
<td>Domestic chores, outdoor activities, leisure and work/3 subscale scores; 1 summed total score (item weights)</td>
<td>... ... ±</td>
<td>± ±</td>
</tr>
<tr>
<td>Karnofsky Performance Status Scale(^{55,60–63})</td>
<td>11-point rating scale/1/observer or interviewer</td>
<td>Physical disability, need of care/1 score</td>
<td>NR</td>
<td>± ... +</td>
</tr>
<tr>
<td>Quality of Life Index(^{51,63,64})</td>
<td>5 items/1–2/observer or interviewer</td>
<td>Physical, emotional, social/1 summed total score (subscale analysis possible)</td>
<td>+</td>
<td>... + + +</td>
</tr>
<tr>
<td>Euroqol(^{65,66})</td>
<td>6 items (with 2 or 3 ordinal steps)/patients: &lt;5; valuation of health states: 20/patient; health states are valued by general public</td>
<td>Utility scale: Mobility, self-care, daily activities, family and leisure activities, pain and discomfort, mood/1 weighted summarized score (health profiles of patients in terms of steps on the 6 domains; health profiles are valued by public preferences)</td>
<td>NR</td>
<td>± ... ± ...</td>
</tr>
<tr>
<td>Quality of Well-being Scale(^{61,67,71})</td>
<td>3 functional scales with 3 (or 5?) ordinal steps; 27 (no consistent reports) symptoms/patients: 12–17; general public: 7–13/patient or caregiver; health states are valued by general public or patient populations</td>
<td>Utility scale: Mobility, physical activity, social activity, symptom/problem complex/1 weighted summarized score (patients are classified on the functional scales and the symptom/problem scale; health states are valued by public or patient preferences)</td>
<td>... + ... ± ...</td>
<td></td>
</tr>
</tbody>
</table>

+, Substantial indications: reliability coefficients >0.80 for the homogeneity (H), test–retest (TR), and interobserver (IO) agreement (if a scale contains few items a coefficient >0.60 is also sufficient). When IO is assessed by κ, a coefficient >0.60 is considered as substantial.\(^{76}\) The responsiveness (RS) of the scale is demonstrated by the ability to detect health changes in patients over time. The scale is able to differentiate (DIF) patient groups. Convergent (CON) validity is demonstrated by significant correlations between the scale scores and instruments measuring the same or closely associated quality of life (QL) domains. Discriminant (DIS) validity is shown by low correlations between the scale scores and instruments measuring different QL dimensions.
Table 1. Continued

<table>
<thead>
<tr>
<th>CON</th>
<th>DIS</th>
<th>Further support‡</th>
<th>Remarks†</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
<td>Positive relation with number of symptoms and chronic diseases.</td>
<td>IO reliabilities show wide fluctuations ($r=0.50–0.98$). Charts are sensitive for effects of several diseases, but further studies are recommended.</td>
</tr>
<tr>
<td>+</td>
<td>.</td>
<td>.</td>
<td>Not all reliability studies are convincing (in population with rheumatoid disease: social scale, $H=0.48$, TR=0.51; physical scale, $H=0.53$). Scores on physical scale differentiate age groups and patient groups. Scale is claimed to be responsive to change, but supportive data are reported only for functional scale. Cut-off point for social scale is suggested (sensitivity=0.72, specificity=0.77).</td>
</tr>
<tr>
<td>±</td>
<td>.</td>
<td>Relation with global health, work absenteeism, place of residence, length of hospital stay. Sex differences in score profiles.</td>
<td>In part two TR&lt;0.60 in 2 social subscales (evaluated in osteoarthritis population). Average score in stroke population did not change despite improvements in physical ability. Items focus on serious health problems, which probably limits the responsiveness of the scale among less ill patients. Only convergence data for emotion scale.</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>Factor analytic support for construct validity. Relations with clinical measures and scores across demographic strata.</td>
<td>Reliability of scores is high but poor on item level ($r=0.50$). Measures relatively gross changes on group level, is probably more responsive to deterioration than to improvement, but is not sensitive enough for subtle individual changes.</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>Health perception is related to several sociodemographic variables.</td>
<td>Floor effect among severely ill patients may limit responsiveness to health deterioration. Two overlapping items on role functions.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>Factor analytic support for the domains.</td>
<td>Poststroke scores in comparison with prestroke levels of function showed health deterioration on the various subscales. Change scores were based on retrospective data. Scale scores discriminated male but not female patients according to severity of their stroke.</td>
</tr>
<tr>
<td>+</td>
<td>.</td>
<td>Related to health perception, number of symptoms, and pain. Scores predict duration of survival in lung cancer.</td>
<td>Two studies in noncancer populations failed to meet $\kappa$ standards and showed substantial disagreement between rating of physicians and patients. Lacks standardized observational procedure.</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>.</td>
<td>There is no standardized observational procedure. Scoring categories are very broadly defined. Scale is not suitable for relatively healthy patients. A self-administration version is available.</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>Can value health states in the community by postal surveys (although a number of inappropriate responses have to be taken into account). Results of 3 studies show close relations between absolute values and ranking orders of health states. On group level there is logical consistency of valuation.</td>
<td>Scale is in development. Translation of measurements into quality-adjusted life years has not been operationalized. Fourteen of the theoretically 216 health states have been valued by the general public. In 1 study (only) 2 health states were presented twice, with high TR correlation.</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>Well-being scores are logically related to number of chronic conditions, symptoms, physician contacts, dysfunctional status, age, subjective well-being, and employment status. Public preferences for various health states do not change over time and show no systematic differences between demographic groups, providers, students, and administrators.</td>
<td>Five symptoms lack empirical weights. Scale is well evaluated on validity, but further analysis of IO reliability and homogeneity is recommended. Data do not consistently support responsiveness of the scale to change. Scale allows calculation of quality-adjusted life years.</td>
</tr>
</tbody>
</table>

±, Moderate indications: data of study reviewed are not consistent, results are summarily presented, or studies suffer from methodological flaws; -. no clear indications: data do not support the criteria; NR, not relevant in view of the characteristic of the scale; . . . , no data available.

*For clarity and brevity the table is presented without original data of the studies.

†Additional and explanatory remarks are only made in case the studies reviewed do not (consistently) meet the psychometric criteria.

‡Since there are a confusing number of other names for validity (e.g., biological, concurrent, statistical) and many researchers do not use the same classification system of validity types, we have summarized additional evidence of validity in neutral terms so that the reader can make his own judgment of what type of validity is supported.
patient, may be less suitable than scales such as the MOS-20 and the NHP. In some cases, especially in longitudinal studies, it may be desirable to incorporate survival data in the measurement and analysis strategy.

Stroke patients are typically elderly and have a range of comorbid conditions.79 Considering the frailty and advanced age of this population, QL measures should be evaluated in terms of their simplicity and their capacity to differentiate the effects of age from those of illness. A thorough analysis of item content is therefore also a prerequisite for the selection of a suitable scale. For example, in the functional domain it is sometimes difficult to differentiate physical disabilities resulting from stroke from dysfunctions related to the normal aging process. A possible contamination effect can also occur in subscales measuring psychological distress or depression. Items directed to the physical manifestations of emotions can also reflect somatic outcomes of stroke, and symptoms such as insomnia, anorexia, and anergia may not only be indicators of an affective state but can also be related to the aging process. Isolation or reduced social activities can be present as a result of physical disability without the presence of depression.

To overcome the conceptual and methodological problems surrounding the assessment of QL in stroke, we propose the following directions for future discussion and study: 1) the conceptualization of QL and its relation to the ICIDH should be further clarified; 2) emphasis should be placed on further psychometric evaluation of existing generic QL measures rather than on generating new instruments; 3) in view of the importance of incorporating QL measures in intervention studies, supplementary data should be generated on the responsiveness of the instruments to within-patient changes over time; 4) stroke outcome studies are also in need of a practical disease-specific QL scale that focuses on the specific problems of stroke patients. This instrument must be able to differentiate the effects of stroke from those of aging and to avoid contamination effects of somatic disturbances on measures of emotional distress; 5) for any QL instrument used, data on a nonstroke control sample with an age distribution similar to that of the stroke group under investigation should be collected; 6) with regard to the utility approach to QL assessment, important methodological issues concerning the methods for generating utilities and convergence with classic QL measures still need to be resolved.

Acknowledgment

The authors wish to acknowledge Patrick Bossuyt of the Department of Clinical Epidemiology and Biostatistics of the Academic Medical Center, University of Amsterdam for his helpful advice in preparing this manuscript.

References

4. Guyatt GH, Jaeschke R: Measurements in clinical trials: Choosing the appropriate approach, in Spilker B (ed): Quality of Life Assess-
tion, 1980
26. Hall JA, Epstein AM, McNeil BJ: Multidimensionality of health status in an elderly population: Construct validity of a measure-
ment battery. Med Care 1989;27:S168–S177

Measuring quality of life in stroke.
R de Haan, N Aaronson, M Limburg, R L Hewer and H van Crevel

Stroke. 1993;24:320-327
doi: 10.1161/01.STR.24.2.320

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/24/2/320