

Assessment scales in stroke: clinimetric and clinical considerations

Jennifer K Harrison¹
Katherine S McArthur²
Terence J Quinn²

¹Department of Cardiovascular Sciences, University of Leicester, Leicester, UK; ²Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK

Abstract: As stroke care has developed, there has been a need to robustly assess the efficacy of interventions both at the level of the individual stroke survivor and in the context of clinical trials. To describe stroke-survivor recovery meaningfully, more sophisticated measures are required than simple dichotomous end points, such as mortality or stroke recurrence. As stroke is an exemplar disabling long-term condition, measures of function are well suited as outcome assessment. In this review, we will describe functional assessment scales in stroke, concentrating on three of the more commonly used tools: the National Institutes of Health Stroke Scale, the modified Rankin Scale, and the Barthel Index. We will discuss the strengths, limitations, and application of these scales and use the scales to highlight important properties that are relevant to all assessment tools. We will frame much of this discussion in the context of “clinimetric” analysis. As they are increasingly used to inform stroke-survivor assessments, we will also discuss some of the commonly used quality-of-life measures. A recurring theme when considering functional assessment is that no tool suits all situations. Clinicians and researchers should choose their assessment tool based on the question of interest and the evidence base around clinimetric properties.

Keywords: Barthel Index, clinimetrics, clinical trial, disability, methodology, modified Rankin Scale, National Institutes Health Stroke Scale, scales, stroke, outcomes

Why measure functional outcomes in stroke trials?

Large-scale clinical trials have created a robust evidence base to inform much of what is now standard acute-stroke practice.¹⁻³ The classical clinical trial is designed to test efficacy of a particular intervention over a comparator, for example, placebo or “usual care.” To facilitate comparison between the groups requires a standard measure of outcome that is relevant and suited to the clinical question, valid for the population studied, and meaningful to the research team. In those trials that describe interventions designed to impact on quantifiable physiological variables, such as glycemia or blood pressure, choice of end-point assessment is reasonably straightforward. Choice of assessment strategy is more challenging for a chronic, nonprogressive, or variably progressive disorder with potential multisystem effects such as cerebrovascular disease. “Hard” clinical end points such as stroke mortality or stroke recurrence are useful, but do not fully capture the potential devastating effect of a disabling but survivable stroke. As stroke represents the leading global cause of adult disability,⁴ an important consideration for any study of stroke interventions is functional recovery. This is recognized by regulatory authorities, who now recommend a measure of functional recovery/disability as primary or coprimary end point for stroke intervention trials.

Correspondence: Terry Quinn
Department of Academic Geriatric
Medicine, Walton Building, Glasgow
Royal Infirmary, Glasgow G4 0SF, UK
Tel +44 141 211 4976
Fax +44 141 211 4033
Email terry.quinn@glasgow.ac.uk

Although the focus of this review will be functional assessment tools for stroke trials, these instruments also have utility in clinical practice. As functional assessment scales give a numerical value to abstract concepts such as “disability,” they can be used to objectively quantify deficits and track change over time. This can be particularly useful in a rehabilitation setting. In clinical practice, an appreciation of how to describe stroke recovery in terms of common stroke scales allows for development of a common language between professionals caring for stroke survivors that facilitates comparisons of patients and services. Within a single review, it would be impossible to review all stroke-specific and generic scales that may be needed in a stroke survivor’s journey (Figure 1). For the interested reader, we recommend a number of reference works.⁵⁻⁷ We recognize there is also extensive literature on assessment strategies for cognitive function poststroke. We will not review cognitive testing, suffice to say that there are a multitude of tools available with little consistency in choice of assessment.⁸

Which functional measure to use

A large number of stroke-assessment scales are described, with novel scales frequently appearing (and often subsequently disappearing) in the literature. For those who are

new to functional assessment, the large and varied nature of available scales and tools may seem daunting. The World Health Organization’s International Classification of Functioning, Disability and Health (WHO-ICF)⁹ gives a conceptual framework that can aid classification of the scales and help decide on the appropriate measure for a particular purpose.

WHO-ICF describes levels of pathology (in this case, the stroke lesion), impairments (the direct loss of function), activity limitation (formerly called disability), and societal participation (formerly called handicap). The WHO-ICF grades do not exist in isolation; they interact and often create feedback loops. For example, an ischemic stroke (pathology) may cause a hemianopia (impairment); this may lead to poor mobility (activity limitation) and may restrict the stroke survivor from driving (societal participation limitation). These problems may result in a fall with soft-tissue injury (impairment), and fear of falling may cause the stroke survivor to forgo usual hobbies and activities (societal participation limitation) (Figure 2).

Tools that assess stroke at all these levels are available. Measures of pathology (for example, size of infarct on imaging) or impairment (for example the Medical Research Council Motor Assessment Scale) are straightforward to perform and interpret, but give little useful information on how stroke affects the individual. For this reason, impairment scales are

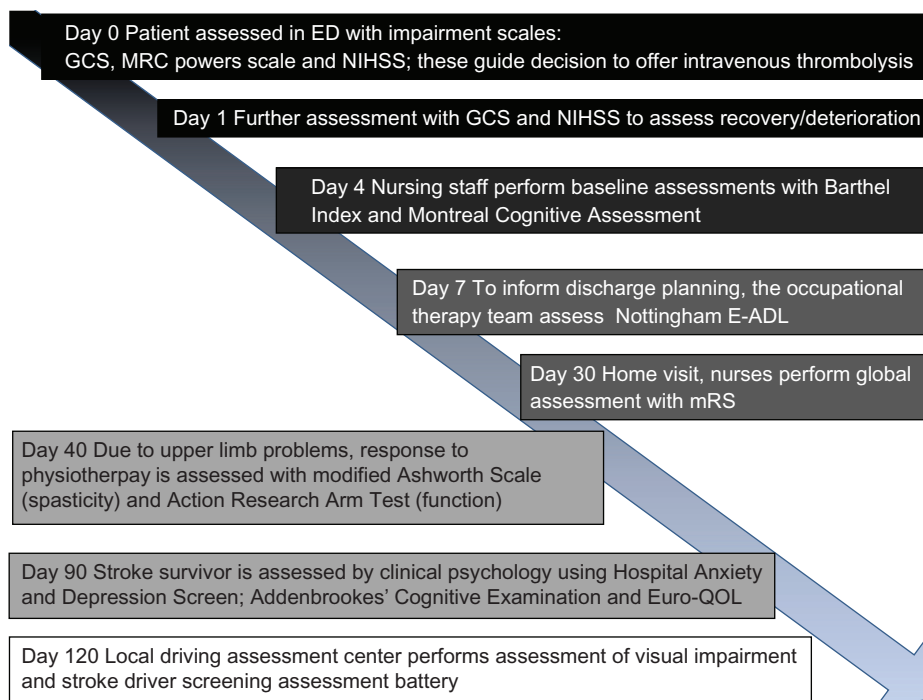


Figure 1 Scales used at various points in the stroke survivor’s journey.

Notes: Note how scale domains move from impairment to activity and participation as the subject progresses. Note also the various agencies that may use scales in their assessment.

Abbreviations: ED, emergency department; GCS, Glasgow Coma Scale; NIHSS, National Institutes of Health Stroke Scale; E-ADL, extended activities of daily living; mRS, modified Rankin Scale; QOL, quality of life; MRC, Medical Research Council.

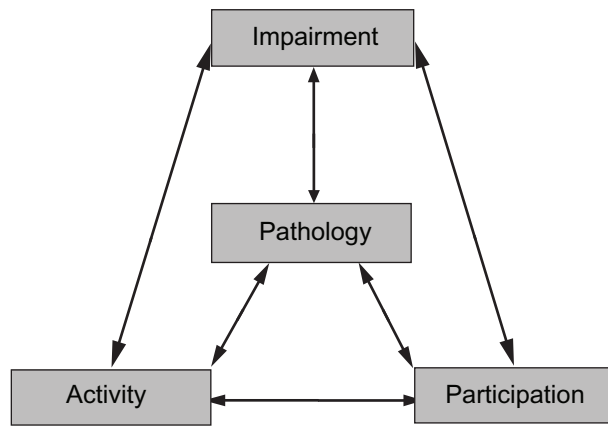


Figure 2 World Health Organization international classification.

often used in early phase trials. For phase III studies, activity measures or measures of participation are usually preferred. Although not part of the WHO-ICF, a further concept of quality of life (QOL) is also described, and tools exist for its measurement. Measures of QOL give a far more detailed assessment, but as a result can be more burdensome to the patient and are often more difficult to interpret (Figure 3).

Clinimetric properties of scales

Clinimetrics is the study of properties of clinical assessment tools;¹⁰ the term is derived from the theory of psychometrics.¹¹ Classical test theory describes important properties such as

validity and reliability.¹² Other important factors for clinical scales are acceptability, both to patient and to assessor, and responsiveness to change. Although in psychometrics, classical test measures are increasingly being superseded by contemporary theories of “item response,” the measures of validity, reliability, and responsiveness remain important for understanding clinical scales, and we will discuss them in turn.

The clinimetric property of validity seeks to assess whether a scale measures the concept it purports to measure. Adequate validity is essential for a stroke scale to have clinical utility, as a functional assessment tool that does not measure function is meaningless. Validity can be assessed in various complementary ways.⁵⁻⁷ There is no “gold standard” for poststroke function, so assessment of criterion validity, where a scale is compared to a reference standard, is not possible. However, concurrent validity can be applied to a stroke scale by comparing it with another measure that purports to measure a similar construct; for example, comparing a novel impairment scale with an established scale. Face validity is an assessment of whether a priori the scale should measure the concept of interest, usually assessed by experts in the field. Content validity asks whether the various items of a scale can adequately describe the concept of interest. Prognostic or predictive validity for a stroke scale may be examined by, for example, studying if an impairment scale is associated with longer-term stroke outcomes.

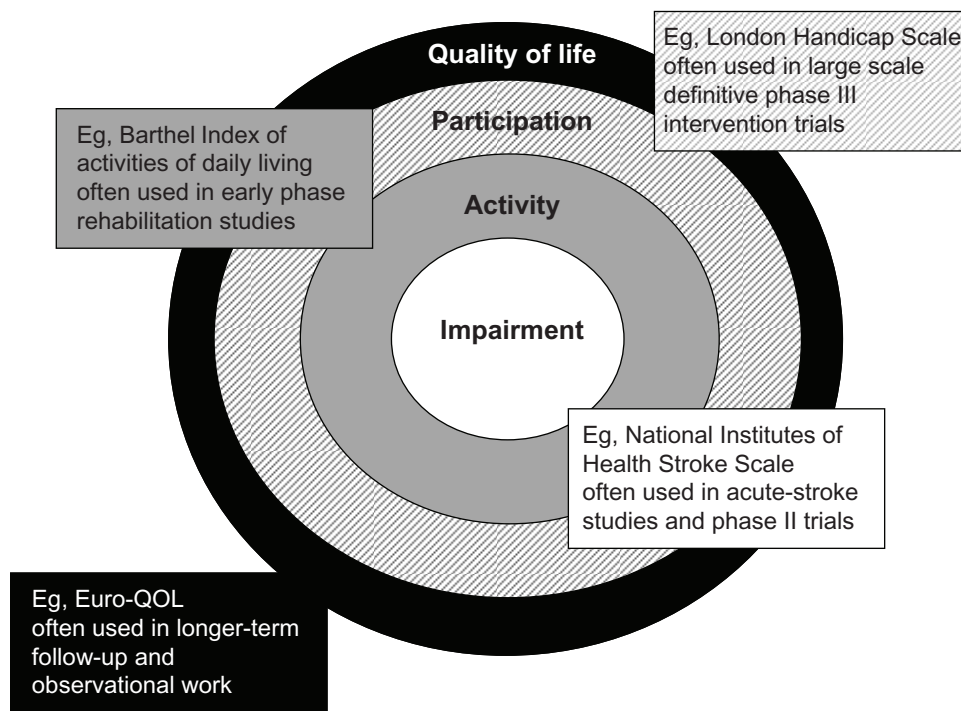


Figure 3 Examples of commonly used stroke scales at differing levels of function. Abbreviation: QOL, quality of life.

Reliability is a measure of consistency in scoring. For stroke scales, important reliability measures include the reproducibility of repeat scoring by the same observer (intra-observer reliability or test–retest reliability) and between scorers (interobserver variability). Whether all items within a scale measure the same construct is a further measure of reliability, usually termed internal consistency. In contemporary stroke trials, where many thousands of stroke survivors may be assessed by hundreds of international research teams, reliability of assessment is clearly paramount. Whereas validity of a scale is inherent, reliability of assessment may be modified. Various methods to improve consistency of assessment are employed in large-scale trials, including training in use of scales, certification exams, and use of standardized protocols. While validity is relative, reliability can be objectively described. There is no consensus on the optimal method to measure reliability, although kappa statistics are frequently used in the biomedical literature to assess agreement. Kappa statistics quantify agreement above that which would be expected by chance. A kappa of 0 would imply no agreement other than that expected by chance, and perfect agreement is scored as 1.0. Traditionally, a kappa greater than 0.6 is taken as sufficient agreement to justify use of a scale. Various forms of statistical “weighting” of kappa values can be used to give a measure of the degree of difference between raters.^{13,14} Increasingly, more sophisticated analyses, such as that of Bland–Altman, are being used to assess reliability.¹⁵

Responsiveness can be thought of as the ability to detect meaningful change over time. Meaningful change is clearly a subjective term, and will vary with the context in which the scale is used. The issue of responsiveness and the ability to detect small but meaningful change is especially important for a condition with high incidence and prevalence, such as stroke. If a scale does not pick up change in function, treatment effects that are modest for the individual but potentially important at a population level could be missed.

The ideal scale would be easy and quick to administer, acceptable to patients and researchers, valid for its chosen purpose, reliable, and responsive to meaningful clinical change. There is no ideal stroke measure that fulfills all these criteria (nor is there ever likely to be). Although some guidance on stroke assessment for trials is emerging, debate continues as to the relative strengths and limitations of differing assessment strategies, and there is no consensus as to the optimal outcome measure(s) for use.

The stroke literature describes a variety of instruments, generic and specific to stroke, for functional assessment

of recovery. A recent analysis of tools used in stroke trials suggests substantial heterogeneity in choice of assessment measure and in method of application.¹⁶ Use of bespoke, nonvalidated assessments is still seen, although less commonly than previously. Certain assessments are used more frequently than others and are increasingly recommended by specialist societies.¹⁷ For the non-stroke specialist, a basic knowledge of the more prevalent stroke scales will allow for improved understanding and critical analysis of stroke studies. We will describe three common stroke assessments: the National Institutes of Health Stroke Scale (NIHSS), the modified Rankin Scale (mRS) and the Barthel Index (BI). We will also discuss some of the commonly used QOL scales. For each scale, we will discuss history, development, and application, and use the scales to further discuss the importance of clinimetric properties.

The scientific study of assessment scales, particularly stroke assessment, is rapidly expanding and it would be impossible to comprehensively cover all areas. In this review, we will not describe the optimal analysis of functional scales for stroke trials. Debate continues as to the relative merits of various statistical techniques, including dichotomization and use of the complete range of a scale, with the differing approaches having vocal proponents.^{18,19} Equally, we will not consider the literature on outcome assessment in animal models of stroke.^{20,21}

National Institutes of Health Stroke Scale

The NIHSS is a 15-item scale that standardizes and quantifies the basic neurological examination, paying particular attention to those aspects most pertinent to stroke. The NIHSS provides an ordinal, nonlinear measure of acute stroke-related impairments by assigning numerical values to various aspects of neurological function.²² The scale incorporates assessment of language, motor function, sensory loss, consciousness, visual fields, extraocular movements, coordination, neglect, and speech.²² It is scored from 0 (no impairment) to a maximum of 42. Scores of 21 or greater are usually described as “severe.” A standardized approach to assessment, starting with fundamental assessments such as level of consciousness, is recommended, and guidance is given on how to score where the stroke survivor is not able to respond to commands.

The NIHSS was developed in the early 1980s as a research tool to allow consistent reporting of neurological deficits in acute-stroke studies, particularly the early trials of thrombolysis and putative neuroprotectants.²² The NIHSS was

developed through a robust consensus approach, taking the most informative measures from existent stroke-examination scales (Toronto Stroke Scale, Oxbury Initial Severity Scale, and Cincinnati Stroke Scale) and creating a composite scale that was further reviewed by a panel of stroke researchers and amended (further items were added to ensure the assessment was as comprehensive as possible). The resulting scale was piloted and refined in a controlled trial of naloxone in acute stroke. It has since been used as primary or coprimary end point in landmark trials of thrombolytic agents and is commonly used in clinical acute-stroke practice.

Using a factor-analysis process, utility of individual components of the NIHSS has been assessed.²³ This work has formed the basis for development of the modified NIHSS or mNIHSS, which removed components deemed unreliable.²⁴ The resulting 11-point mNIHSS has been prospectively assessed, and improved reliability is described.²⁵ As the standard NIHSS is already fairly quick to perform and reliable, it is debatable whether a shorter scale is needed, and at present the mNIHSS is not frequently used in trials or practice. Further amendments to the NIHSS have been described to facilitate use of the scale in prehospital settings.²⁶ A pediatric NIHSS (pedNIHSS) is also described.²⁷

The NIHSS has many advantages as a stroke outcome-assessment tool. It is relatively straightforward and takes around 6 minutes to perform, with no need for additional equipment. In the acute-stroke environment, the NIHSS is well suited to serial measures of impairment. It has been suggested that a change in the NIHSS of more than 2 points represents clinically relevant early improvement or deterioration.²⁸

NIHSS scores are reliable across observers, and this has been demonstrated both in cohorts of neurology-trained and non-neurologist raters. The availability of a reliable method for neurological exam that is suitable for nonspecialists is a particular strength of the NIHSS. Reliability and validity has also been demonstrated for remote NIHSS assessment via telemedicine.²⁹ The interobserver reliability of the NIHSS is further improved by the various training materials available. Training resources now exist, such as DVDs and online educational aids, as well as pocket-sized NIHSS summary scales. Practitioners can undergo certification to demonstrate their proficiency in assessment and interpretation of the NIHSS.³⁰

Content validity of the NIHSS has been demonstrated, although high internal consistency suggests that certain items of the NIHSS may be redundant. The NIHSS has predictive validity, as initial score is a robust predictor of in-hospital

complication and outcome at 3 months.^{31,32} Correlations with objective measures of stroke severity, such as size of infarct on imaging, provide further evidence of NIHSS validity.^{23,33,34} Compared with BI and mRS, the NIHSS is the more sensitive outcome score, requiring potentially smaller sample sizes to detect relevant therapeutic effects.^{23,35} The NIHSS is responsive to change and can measure impairment throughout the expected range of stroke severity.³⁶

A criticism of the NIHSS relates to its validity in certain nondominant-hemisphere stroke syndromes. It is well recognized that an individual can score 0 on the NIHSS, despite having evidence of ischemic stroke, particularly in the posterior circulation territory.³⁷ Examination of the component subscales of the NIHSS reveals a focus on limb and speech impairments and relatively little attention to, for example, cranial nerve lesions. Similarly, when the NIHSS is used to predict dependent living, lower scores are seen in posterior circulation events compared to anterior circulation.³⁸ There are radiological correlates, when quantifying extent of cerebral damage for a specified NIHSS score, the median volume of right-hemisphere strokes is larger than the volume of left-hemisphere strokes, suggesting nondominant strokes are required to be more severe to reach the same grading on the NIHSS.³⁹ As an impairment scale, the NIHSS can give only limited information on how stroke has affected the individual stroke survivor. For example, an NIHSS score of 1 is considered an “excellent” outcome from stroke; a hemianopia that precludes driving and may necessitate loss of employment would score NIHSS 1, but for the individual this may not seem an “excellent” result.

Barthel Index

Adapted from the Maryland Disability Index, the BI authors – Florence I Mahoney and Dorothea W Barthel – intended their scale for use as “a simple index of independence, useful in scoring improvement in rehabilitation.”⁴⁰ First described in the 1950s and published in 1965, the BI was developed to assist in discharge planning from long-term care wards. With time, the BI has been adopted by other disciplines and is a recommended assessment in older adult care.⁴¹ The BI is the most commonly used functional measure in stroke-rehabilitation settings and the second most commonly used functional outcome measure across stroke trials.^{16,42} Many scales have been described that take the name “Barthel Index”. Some authors have sought to modify or adapt the BI from the original; these include reducing the number of items,⁴³ extending it with the addition of cognitive and social domains,⁴⁴ and attempts to further subdivide

the outcomes to include different degrees of assistance.⁴⁵ However, each of these requires independent validation, as it is known that even comparatively minor adaptations alter the validity of a tool and accuracy of responses.⁴⁶ For consistency, it is recommended that a single BI measure is used; the scale as described by Wade and Collin⁴⁷ has been used in many trials.

The BI assesses ten functional tasks of daily living (activities of daily living – ADL), scoring the individual depending on independence in each task. Scores range from 0 and 100, with a higher score indicating greater independence (Table 1). The BI is usually summed to give a total score. While this can be useful for statistical analysis, it is more informative in practice to present the scores for the individual domains. An unresolved issue for trials is how to define a “good” BI outcome, with significant heterogeneity within the published literature⁴⁸ and attempts to subcategorize, based on total score.^{49,50} A popular interpretation of BI scores is that subjects with BI > 80 are generally independent and should be able to return home; while subjects with BI < 40 are very dependent.⁵¹ Other interpretations of favorable and unfavorable BI outcomes have been described: statistical modeling looking at differing BI scores as a trial end point suggested a score of 95/100 was the optimal descriptor of an excellent outcome and that 75/100 was the best cut point for defining a poor outcome.⁵²

Validity of BI is well described. The scale is recognized as a valid prognostic tool following stroke, in particular as predictor of recovery, level of care required,⁵³ and duration of rehabilitation required following stroke.⁵⁴ BI scores correlate with other stroke-assessment scales,⁵⁵ including other more detailed ADL scales.⁵⁶ Interobserver reliability is usually quoted as a strength of the BI, and reliability has been demonstrated in nonstroke populations.⁵⁷ Systematic review of

reliability of BI in stroke also suggests reasonable reliability, although few multicenter reliability studies are available.⁵⁸

The original BI is not without its limitations. As a scale of primarily physical function, it does not reflect the burden on the individual of communication and cognitive deficits that can result from a stroke event.⁵⁹ For clinical trials, the BI lacks a result to represent stroke mortality, and this can complicate analysis of results. However, the major limitation of the BI for clinical trial use is its responsiveness to change. Although in certain stroke-care settings, the BI is as sensitive to change as other scales,⁶⁰ a score must be able to represent changes throughout the entire spectrum of potential functional outcomes. It is in this regard that the “floor” and “ceiling” effects of the BI become apparent.⁵⁰ “Floor and ceiling” describes the phenomenon by which the score does not change from minimum or maximum despite clinical change.⁶¹ For example, a stroke patient in a neurointensive care setting can make significant gains but still score a total 0 on the BI; conversely, a patient who is discharged from hospital and independent may still have substantial functional problems but will score 100 on the BI. Given this limitation, the BI may be best suited to stroke survivors requiring inpatient rehabilitation, while other scales may be needed to assess functional change in those with more major or minor stroke symptoms.⁶²

The BI can be considered as a measure of “basic” ADL (self-care and mobility). Scales have been developed to encapsulate performance in more complex tasks. These are variously described as “instrumental” or “extended” activities of daily living (E-ADL) measures.⁶³ The term “instrumental ADL” was first used in Lawton and Brody’s work, and a Lawton I-ADL scale is described.⁶⁴ A validated measure that has been used with stroke survivors is the Nottingham Extended ADL Scale, which asks participants to reflect their

Table 1 The Barthel Index of activities of daily living

Domain assessed	Score			
	0	5	10	15
Feeding	Unable	Requires assistance	Independent	
Bathing	Dependent	Independent		
Grooming	Needs help	Independent		
Dressing	Dependent	Needs some help	Independent	
Bowels	Incontinent	Occasional accident	Continent	
Bladder	Incontinent or catheterized	Occasional accident	Continent	
Toilet use	Dependent	Needs some help	Independent	
Transfers (bed to chair and back)	Unable	Major help	Minor help	Independent
Mobility (on level surface)	Immobile	Wheelchair independent > 50 yards	Walks with help of one	Independent
Stairs	Unable	Needs help	Independent	

Note: The wording and scoring presented is that of Wade and Collin.⁴⁷

actual activities over the preceding weeks, rather than simply what they have the capability to do.^{65,66} The Nottingham Extended ADL Scale compares favorably to the BI, and is less susceptible to the ceiling effects described.⁶⁷

Modified Rankin Scale

The mRS is a 6-point, ordinal hierarchical scale that describes “global disability” with a focus on mobility (Table 2). The original Rankin Scale was developed by the Scottish physician John Rankin to describe the positive outcomes he was achieving in his prototypic stroke unit.⁶⁸ Although not originally intended as an assessment for clinical trials, a slightly modified version of Rankin’s eponymous scale was used as end point in the first multicenter stroke trial (the UK TIA study).⁶⁹ Since this time, the mRS has grown in popularity and is now the most commonly used functional measure in stroke trials, and has been the primary or coprimary outcome in most recent large-scale stroke trials.¹⁶ A further variation of the mRS, the Oxford Handicap Scale, has been described but is not commonly used by trialists. In contemporary stroke studies, the mRS is often used both as a measure of premorbid ability to assist in selection of patients and as final outcome measure.

The mRS has many potential strengths, and it is acceptable to patient and assessor, with nonstandardized interviews taking around 5 minutes to complete.⁷⁰ Concurrent validity is demonstrated by strong correlation with measures of stroke pathology (for example, infarct volumes) and agreement with other stroke scales.^{71,72} The six potential scores on the mRS (0–5) describe a full range of stroke outcomes, with a score of 6 usually added to denote death. With a limited number of scores, the mRS may be less responsive to change than

some other scales; however, a single-point change on the mRS will always be clinically relevant.

The principle limitation of the mRS is its reliability, with the potential for substantial interobserver variability. A study describing the interobserver variability of the mRS is available; indeed, in the first clinical studies that used the mRS as end point, the trialists described interobserver variability for a third of subjects interviewed by paired assessors.⁷³ A systematic review and meta-analysis of studies describing interobserver variability of the mRS reports pooled reliability across ten published studies ($n = 587$ patients) of $\kappa = 0.46$.⁷⁴ Those studies that assessed mRS reliability with multiple raters and centers (ie, similar to a contemporary clinical trial) revealed a worryingly low agreement of $\kappa = 0.21$.⁷⁵ This level of inconsistency will impact on the validity of the trial results and conclusions. The statistical “noise” created by the interobserver variability will increase the possibility of a type II error, ie, a beneficial treatment effect is missed. It has been postulated that problems with mRS reliability may have partly explained a series of unexpected neutral results in large-scale neuroprotectant studies. There are published examples of nonstroke studies whose results were fundamentally altered when statistical analysis accounted for observer variation.⁷⁶

Recognizing the problems of reliability in standard mRS assessments, trialists have explored various interventions to improve consistency in scoring. Usual mRS interviews are unstructured, and researchers vary considerably in their length of interview and number of questions asked. More structured approaches to assessment have been described, from a comprehensive scripted interview⁷⁵ to use of anchoring questions that require a yes/no answer.⁷⁷ The groups that developed these assessments describe substantial improvements in reliability. However, improvements have not been seen when the structured interviews have been tested by independent centers.⁷⁸ Training in use of the mRS can also offer potential to improve consistency. As with the NIHSS and BI, an online training resource is available with an accompanying certification exam.⁷⁹ A further trial modification that may improve reliability is to record mRS interviews and have a remote consensus grading by experienced stroke trialists. This approach is currently being utilized by a number of multicenter studies.

Two other modifications to mRS assessment are commonly used and deserve some discussion: using proxies to substitute for stroke survivors in the mRS interview and calculating a “prestroke” mRS. Stroke survivors often have physical, language, or cognitive impairments that may

Table 2 The modified Rankin Scale (mRS)

Grade 0	No symptoms at all
Grade 1	No significant disability despite symptoms; able to carry out all usual duties and activities
Grade 2	Slight disability; unable to carry out all previous activities, but able to look after own affairs without assistance
Grade 3	Moderate disability; requiring some help, but able to walk without assistance
Grade 4	Moderately severe disability; unable to walk without assistance, unable to attend to needs without assistance
Grade 5	Severe disability; bedridden, incontinent, and requiring constant nursing care and attention
Grade 6	Dead

Notes: The grade of mRS 6 was added for clinical trial purposes and was not part of the original scale.

complicate a standard face-to-face interview. In this situation, an informant who knows the patient often supplements the interview or substitutes it completely. While this approach makes intuitive sense, we should not assume validity, and clinimetric analysis is still required. A recent systematic review of proxy stroke scales (the mRS was not included) suggested that the properties of certain proxy-based assessments may differ from equivalent standard assessments.⁸⁰ A study of proxy mRS described suboptimal reliability and validity, and recommended that direct mRS interview with the patient should be the preferred assessment if possible.⁸¹ In stroke trials, traditional statistical analyses assess numbers achieving a “good” functional outcome. To improve trial power, subjects with disability prior to the stroke event are often excluded. Assessment of a “prestroke mRS” has been used in many landmark stroke trials, where prestroke mRS > 2 is used as the exclusion criterion during participant selection. The wording of the mRS grades is not suited to such prestroke assessment, and it is perhaps unsurprising that when formally assessed, prestroke mRS had only moderate reliability and validity.⁸²

Quality of life

In view of improving longer-term survival and functional outcomes following stroke,⁸³ it could be argued that assessments against participation or QOL will become increasingly important.⁸⁴ Certainly evaluations of health-related QOL in stroke survivors can provide a rich description of the multifaceted effects of a stroke, providing insights above those recorded with traditional impairment and activity measures.⁸⁵ Measuring health-related QOL in stroke presents particular challenges. Important predictors and components of QOL following stroke will vary at different periods following the event.⁸⁶ Thus, we must balance having a suitably comprehensive assessment that is sensitive to the nuances of QOL against the time and burden required for this assessment. QOL is very much dependent on the individual’s experience of their condition. This poses a particular problem where the stroke survivor has difficulty communicating. Carer/family-based assessments of the patient’s QOL are often biased, with the proxies reporting poorer outcomes than the subject.⁸⁷

Various QOL scales have been proposed, some generic and some specific to stroke/brain injury. QOL scales should be subject to the same rigor of clinimetric assessment as any other scale. It is evident from the published literature that for QOL there is a propensity to generate new scales rather than validating existing ones.^{88,89} It has been argued that QOL can be assessed by asking just two questions assessing

Table 3 Domains assessed in three commonly used quality-of-life scales

Scale		
SF-36	Euro-Qol	SS-QOL
Domains		
Physical functioning	Mobility	Mobility
Physical role limitations	Self-care	Self-care
Emotional role limitations	Anxiety/depression	Mood
Emotional well-being	Usual activities	Work/productivity
Social functioning	Pain/discomfort	Social roles
Bodily pain		Family roles
General health		Language
Vitality		Personality
		Energy
		Thinking
		Upper extremity function
		Vision

Note: Differences in content and scope for three scales that purport to measure the same construct.

Abbreviations: SF-36, Short Form 36; SS-QOL, stroke-specific quality of life; QOL, quality of life.

“dependency” and “problems.”⁹⁰ An alternative approach is to apply existing health-related QOL scales or to use disease-specific scores. There are strengths and limitations to each approach.

The Short Form 36 (SF-36) is a generic scale intended for patient completion that assesses eight domains of health-related QOL derived from the Medical Outcomes Study (Table 3).⁹¹ Although the SF-36 is validated for stroke patients,⁹² noncompletion bias and marked floor and ceiling effects may limit its utility.^{93,94} The generic QOL scale, Euro-Qol, was developed based on the findings of an international postal survey.⁹⁵ The self-completion questionnaire requires assessment across five domains complemented by a visual analog scale (Table 3).⁹⁶ EuroQol has been validated in stroke populations.⁹⁷ However, noncompletion bias is recognized: in one study, only 61% of stroke survivors could complete the scale without external assistance.⁹⁷ The stroke-specific QOL scale was developed based on interviews with stroke survivors.⁹⁸ It is based on twelve domains (Table 3).⁹³ The scale is validated in stroke populations, and values for “minimal detectable change” and “clinically important difference”⁹⁹ are established. A modification for those with poststroke aphasia is also described.¹⁰⁰

Conclusion

Many assessment tools, spanning various functional domains, are available to clinicians and researchers working with stroke survivors. We have given a flavor of the marked heterogeneity in use of assessment scales. Lack of consistency

in outcome assessment has hindered comparative research and meta-analysis, and so we would recommend that future researchers use a common set of outcome assessments. No perfect stroke-assessment scale exists, and in this review we have deliberately avoided suggestions that one scale is better than another. We have focused on the three most commonly used stroke scales (mRS, BI, NIHSS) as exemplars. These scales have been validated, are familiar to many, and have proven utility, with each suited to differing assessment scenarios. Thus, in the absence of a “perfect” assessment, we would recommend continuing use of the three core assessment scales: the mRS as an outcome if the study is describing global disability, the NIHSS for studies looking at neurological impairment, and the BI for studies looking at basic ADL. Trialists and clinicians can supplement these core assessments with specific tools suited to the clinical scenario/research question. Increasing awareness of the importance of clinimetric properties has highlighted deficiencies and potential limitations with stroke functional assessment. Clinicians and researchers should always select their assessment tool(s) based on the question of interest and the evidence base around clinimetric properties. Where, as is often the case, the research around clinimetric properties of a scale is sparse, we would encourage researchers to design and conduct their own clinimetric studies.

Disclosure

The authors report no conflicts of interest in this work.

References

- Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke. *Cochrane Database Syst Rev*. 2007;4:CD000197.
- Wardlaw JM, Murray V, Berge E, del Zoppo GJ. Thrombolysis for acute ischaemic stroke. *Cochrane Database Syst Rev*. 2009;4:CD000213.
- Sandercock PAG, Counsell C, Kamal AK. Anticoagulants for acute ischaemic stroke. *Cochrane Database Syst Rev*. 2008;4:CD000024.
- Bath PMW, Lees KR. ABC of arterial and venous disease. Acute stroke. *BMJ*. 2000;320:920–923.
- The Internet Stroke Centre. Stroke assessment scales overview. 1995. Available from: <http://www.strokecenter.org/trials/scales/scales-overview.htm>. Accessed December 7, 2012.
- Wade DT. *Measurement in Neurological Rehabilitation*. Oxford: Oxford University Press; 1992.
- Kane RL. *Assessing Older Persons: Measures, Meaning and Practical Applications*. New York: Oxford University Press; 2000.
- Lees R, Fearon P, Harrison JK, Broomfield NM, Quinn TJ. Cognitive and mood assessment in stroke research: focused review of contemporary studies. *Stroke*. 2012;43:1678–1680.
- World Health Organization. *Towards a Common Language for Functioning, Disability and Health*. Geneva: WHO; 2002. Available from: <http://www.who.int/classifications/icf/training/icfbeginnersguide.pdf>. Accessed December 7, 2012.
- Feinstein AR. An additional basic science for clinical medicine: IV. The development of clinimetrics. *Ann Intern Med*. 1983;99:843–848.
- Asplund K. Clinimetrics in stroke research. *Stroke*. 1987;18:528–530.
- Fava GA, Tomba E, Sonino N. Clinimetrics: the science of clinical measurements. *Int J Clin Pract*. 2012;66:11–15.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.
- Kraemer HC, Bloch DA. Kappa coefficients in epidemiology: an appraisal of a reappraisal. *J Clin Epidemiol*. 1988;41:59–68.
- Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res*. 1999;8:135–160.
- Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials. *Int J Stroke*. 2009;4:200–205.
- Turner-Stokes L. Measurement of outcome in rehabilitation: the British Society of Rehabilitation Medicine “basket” of measures. 2000. Available from: <http://www.bsrm.co.uk/ClinicalGuidance/OutcomeMeasuresB3.pdf>. Accessed December 7, 2012.
- Optimising Analysis of Stroke Trials (OAST) Collaboration, Bath PM, Gray LJ, Collier T, Pocock S, Carpenter J. Can we improve the statistical analysis of stroke trials? Statistical reanalysis of functional outcome in stroke trials. The optimising analysis of stroke trials (OAST) collaboration. *Stroke*. 2007;38:1911–1915.
- Saver JL, Gornbein J. Treatment effects for which shift or binary analyses are advantageous in acute stroke trials. *Neurology*. 2009;72:1310–1315.
- Popa-Wagner A, Stöcker K, Balseanu AT, et al. Effects of granulocyte-colony stimulating factor after stroke in aged rats. *Stroke*. 2010;41:1027–1031.
- Schaar KL, Breneman MM, Savitz SI. Functional assessments in the rodent stroke model. *Exp Transl Stroke Med*. 2010;2:13.
- Brott T, Adams HP, Olinger CP, et al. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke*. 1989;20:864–870.
- Lyden P, Lu M, Jackson C, et al. Underlying structure of the National Institutes of Health Stroke Scale: results of a factor analysis: NINDS tPA Stroke Trial Investigators. *Stroke*. 1999;30:2347–2354.
- Lyden PD, Lu M, Levine SR, Brott TG, Broderick J. A modified National Institutes of Health Stroke Scale for use in stroke clinical trials: preliminary reliability and validity. *Stroke*. 2001;32:1310–1317.
- Meyer BC, Henmen TM, Jackson CM, Lyden PD. Modified National Institutes of Health Stroke Scale for use in stroke clinical trials: prospective reliability and validity. *Stroke*. 2002;33:1261–1266.
- Tirschwell DL, Longstreth WT, Becker KJ, et al. Shortening the NIHSS for use in the prehospital setting. *Stroke*. 2002;33:2801–2806.
- Ichord RN, Bastian R, Abraham L, et al. Interrater reliability of the Pediatric National Institutes of Health Stroke Scale (PedNIHSS) in a multicenter study. *Stroke*. 2011;42:613–617.
- Adams HP, Davis PH, Leira EC, et al. Baseline NIH stroke scale strongly predicts outcome after stroke: a report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST). *Neurology*. 1999;53:126–131.
- Meyer BC, Raman R, Chacon MR, Jensen M, Werner JD. Reliability of site-independent telemedicine when assessed by telemedicine-naïve stroke practitioners. *J Stroke Cerebrovasc Dis*. 2008;17:181–186.
- Lyden P, Raman R, Liu L, Emr M, Warren M, Marler J. National Institutes of Health Stroke Scale certification is reliable across multiple venues. *Stroke*. 2009;40:2507–2511.
- Boone M, Chillon JM, Garcia PY, et al. NIHSS and acute complications after anterior and posterior circulation strokes. *Ther Clin Risk Manag*. 2012;8:87–93.
- Johnston KC, Wagner DP. Relationship between 3-month National Institutes of Health Stroke Scale score and dependence in ischemic stroke patients. *Neuroepidemiology*. 2006;27:96–100.
- Muir KW, Weir CJ, Murray GD, Povey C, Lees KR. Comparison of neurological scales and scoring systems for acute stroke prognosis. *Stroke*. 1996;27:1817–1820.
- Adams HP Jr, Bendixen BH, Leira E, et al. Antithrombotic treatment of ischemic stroke among patients with occlusion or severe stenosis of the internal carotid artery: a report of the Trial of Org 10172 in Acute Stroke Treatment (TOAST). *Neurology*. 1999;53:122–125.

35. Young FB, Weir CJ, Lees KR; GAIN International Trial Steering Committee and Investigators. Comparison of the National Institutes of Health Stroke Scale with disability outcome measures in acute stroke trials. *Stroke*. 2005;36:2187–2192.
36. Lindsell CJ, Alwell K, Moomaw CJ, et al. Validity of a retrospective National Institutes of Health Stroke Scale scoring methodology in patients with severe stroke. *J Stroke Cerebrovasc Dis*. 2005;14(6):281–283.
37. Martin-Schild S, Albright KC, Tanksley J, et al. Zero in the NIHSS does not equal the absence of stroke. *Ann Emerg Med*. 2011;57: 42–45.
38. Sato S, Toyoda K, Uehara T, et al. Baseline NIH Stroke Scale Score predicting outcome in anterior and posterior circulation strokes. *Neurology*. 2008;70:2371–2377.
39. Woo D, Broderick JP, Kothari RU, et al. Does the National Institutes of Health Stroke Scale favor left hemisphere strokes? *Stroke*. 1999;30:2355–2359.
40. Mahoney FI, Barthel D. Functional evaluation: the Barthel Index. *Md State Med J*. 1965;14:56–61.
41. Royal College of Physicians England. *Standardised Assessment Scales for Elderly People. Report of the Joint Workshops of the Research Unit of the Royal College of Physicians of London and the British Geriatric Society*. London: Royal College of Physicians; 1992.
42. Sangha H, Lipson D, Foley N, et al. A comparison of the Barthel Index and the Functional Independence Measure as outcome measures in stroke rehabilitation: patterns of disability scale usage in clinical trials. *Int J Rehabil Res*. 2005;28:135–139.
43. Hobart JC, Thompson AJ. The five item Barthel index. *J Neurol Neurosurg Psychiatry*. 2001;71:225–230.
44. Prosiegel M, Bottger S, Schenk T. Der Erwertiertr Barthel Index (EBI)-eine neue Skala zur Erfassung von Fahigkeitsstorungen bei neurologischen patienten. [The extended Barthel Index a new scale for assessment of functional ability in neurological patients]. *Neurol Rehabil*. 1996;1:7–13.
45. Shah S, Vanclay F, Cooper B. Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol*. 1989;42:703–709.
46. Picavet HS, van den Bos GA. Comparing survey data on functional disability: the impact of some methodological differences. *J Epidemiol Community Health*. 1996;50:86–93.
47. Wade DT, Collin C. The Barthel ADL Index: a standard measure of physical disability. *Int Disabil Stud*. 1988;10:64–67.
48. Sulter G, Steen C, De Keyser J. Use of the Barthel Index and Modified Rankin Scale in acute stroke trials. *Stroke*. 1999;30:1538–1541.
49. Nakao S, Takata S, Uemura H, et al. Relationship between Barthel Index scores during the acute phase of rehabilitation and subsequent ADL in stroke patients. *J Med Invest*. 2010;57:81–88.
50. Balu S. Differences in psychometric properties, cut-off scores, and outcomes between the Barthel index and modified Rankin Scale in pharmacotherapy-based stroke trials: systematic literature review. *Curr Med Res Opin*. 2009;25:1329–1341.
51. Sinoff G, Ore L. The Barthel activities of daily living index: self-reporting versus actual performance in the old-old (> or = 75 years). *J Am Geriatr Soc*. 1997;45:832–836.
52. Uyttenboogaart M, Stewart RE, Vroomen PC, De Keyser J, Luijckx GJ. Optimizing cutoff scores for the Barthel index and the modified Rankin scale for defining outcome in acute stroke trials. *Stroke*. 2005;36:1984–1987.
53. Huybrechts KF, Caro JJ. The Barthel Index and modified Rankin Scale as prognostic tools for long term outcomes after stroke: a qualitative review of the literature. *Curr Med Res Opin*. 2007;23:1627–1636.
54. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. *Arch Phys Med Rehabil*. 2000;81 Suppl 2: S21–S29.
55. Granger CV, Albrecht GL, Hamilton BB. Outcome of comprehensive medical rehabilitation: measurement by PULSES Profile and Barthel Index. *Arch Phys Med Rehabil*. 1979;60:145–154.
56. Hobart JC, Lamping DL, Freeman JA, et al. Evidence-based measurement: which disability scale for neurologic rehabilitation? *Neurology*. 2001;57:639–644.
57. Sainsbury A, Gudrun S, Bansal A, Young JB. Reliability of the Barthel Index when used with older people. *Age Ageing*. 2005;34:228–232.
58. Duffy L, Gajree S, Langhorne P, Stott DJ, Quinn TJ. Reliability of Barthel Index in stroke – systematic review and meta-analysis. *Stroke*. 2012. In press.
59. Novak S, Johnson J, Greenwood R. Barthel revisited: making guidelines work. *Clin Rehabil*. 1996;10:128–134.
60. Dromerick AW, Edwards DF, Diringner MN. Sensitivity to changes in disability after stroke: comparison of four scales useful in clinical trials. *J Rehabil Res Dev*. 2003;40:1–8.
61. Schepers VP, Ketelaar M, Visser-Meily JM, Dekker J, Lindeman E. Responsiveness of functional health status measures frequently used in stroke research. *Disabil Rehabil*. 2006;28:1035–1040.
62. Weimar C, Kurth T, Kraywinkel K, et al. Assessment of Functioning and disability after ischemic stroke. *Stroke*. 2002;33:2053–2059.
63. Chong DK. Measurement of instrumental activities of daily living in stroke. *Stroke*. 1995;26:1119–1122.
64. Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9: 179–186.
65. Nouri FM, Lincoln NB. An extended ADL scale for use with stroke patients. *Clin Rehabil*. 1987;1:301–305.
66. Gladman JRF, Lincoln NB, Adams SA. Use of the extended ADL scale in stroke patients. *Age Ageing*. 1993;22:419–424.
67. Sarker SJ, Rudd AG, Douiri A, Wolfe CD. Comparison of 2 extended activities of daily living scales with the Barthel Index and predictors of their outcomes: cohort study within the South London Stroke Register (SLSR). *Stroke*. 2012;43:1362–1369.
68. Quinn TJ, Dawson J, Walters M. Dr John Rankin; his life, legacy and the 50th anniversary of the Rankin Stroke Scale. *Scot Med J*. 2008;53:44–47.
69. Farrell B, Godwin J, Richards S, Warlow C. The United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: final results. *J Neurol Neurosurg Psychiatry*. 1991;54:1044–1054.
70. Quinn TJ, McArthur K, Dawson J, Walters MR, Lees KR. Reliability of structured modified Rankin Scale assessment. *Stroke*. 2010;41:602–603.
71. Schiemanck SK, Post MWM, Kwakkel G, et al. Ischaemic lesion volume correlates with long term functional outcome and quality of life of middle cerebral artery stroke survivors. *Restor Neurol Neurosci*. 2005;23:257–263.
72. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin Scale: implications for stroke clinical trials: a literature review and synthesis. *Stroke*. 2007;38:1091–1096.
73. Van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJA, van Gijn J. Inter-observer agreement for the assessment of handicap in stroke patients. *Stroke*. 1988;19:604–607.
74. Quinn TJ, Dawson J, Walters MR, Lees KR. Reliability of the modified Rankin Scale: a systematic review. *Stroke*. 2009;40:3393–3395.
75. Wilson JTL, Hareendran A, Hendry A, Potter J, Bone I, Muir KW. Reliability of the modified Rankin Scale across multiple raters: benefits of a structured interview. *Stroke*. 2005;36:777–781.
76. Jaffar S, Leach A, Smith PG, Cutts F, Greenwood B. Effects of misclassification of causes of death on the power of a trial to assess the efficacy of a pneumococcal conjugate vaccine in the Gambia. *Int J Epidemiol*. 2003;32:430–436.
77. Saver JL, Filip B, Yanes A, et al. Improving the reliability of stroke disability grading in clinical trials and clinical practice: the Rankin Focused Assessment. *Stroke*. 2010;5:992–995.
78. Quinn TJ, Dawson J, Walters MR, Lees KR. Exploring the reliability of the modified Rankin Scale. *Stroke*. 2009;40:762–766.
79. Quinn TJ, Dawson J, Lees KR, Hardemark HG, Walters MR. Initial experience of a digital training resource for modified Rankin Scale assessment in clinical trials. *Stroke*. 2007;38:2257–2261.
80. Oczkowski C, O'Donnell M. Reliability of proxy respondents for patients with stroke: a systematic review. *J Stroke Cerebrovasc Dis*. 2010;5:410–416.

81. McArthur K, Beagan MLC, Degnan A, et al. Properties of proxy-derived modified Rankin Scale assessment. *Int J Stroke*. Epub February 15, 2012.
82. Fearon P, McArthur KS, Garrity K, et al. Pre-stroke modified Rankin Stroke Scale has moderate inter-observer reliability and validity in an acute stroke settings. *Stroke*. 2012;43:84–88.
83. Langhorne P, Wright F, Stott D. Improved survival after stroke. *Cerebrovasc Dis*. 2007;23:320–321.
84. Carod-Artal FJ, Egido JA. Quality of life after stroke: the importance of a good recovery. *Cerebrovascular Dis*. 2009;27 Suppl 1:204–214.
85. Patel MD, Tilling K, Lawrence E, Rudd AG, Wolfe CDA, McKeivitt C. Relationships between long-term stroke disability, handicap and health-related quality of life. *Age Ageing*. 2006;35:273–279.
86. Patel MD, McKeivitt C, Lawrence E, Rudd AG, Wolfe DA. Clinical determinants of long-term quality of life after stroke. *Age Ageing*. 2007;36:316–322.
87. Williams LS, Bakas T, Brizendine E, et al. How valid are family proxy assessments of stroke patients' health-related quality of life? *Stroke*. 2006;37:2081–2085.
88. De Haan R, Aaronson N, Limburg M, Hewer RL, van Crevel H. Measuring quality of life in stroke. *Stroke*. 1993;24:320–327.
89. Tengs TO, Yu M, Luistro E. Health-related quality of life after stroke a comprehensive review. *Stroke*. 2001;32:964–972.
90. Dorman P, Dennis M, Sandercock P. Are the modified "simple questions" a valid and reliable measure of health related quality of life after stroke? United Kingdom Collaborators in the International Stroke Trial. *J Neurol Neurosurg Psychiatry*. 2000;69:487–493.
91. Ware JE. SF-36 health survey update. *Spine (Phila Pa 1976)*. 2000;25:3130–3139.
92. Anderson C, Laubscher S, Burns R. Validation of the Short Form 36 (SF-36) health survey questionnaire among stroke patients. *Stroke*. 1996;27:1812–1816.
93. O'Mahony PG, Rodgers H, Thomson RG, Dobson R, James OFW. Is the SF-36 suitable for assessing health status of older stroke patients? *Age Ageing*. 1998;27:19–22.
94. Hobart JC, Williams LS, Moran K, Thompson AJ. Quality of life measurement after stroke: uses and abuses of the SF-36. *Stroke*. 2002;33:1348–1356.
95. [No authors listed.] EuroQol – a new facility for the measurement of health-related quality of life. The EuroQol Group. *Health Policy*. 1990;16:199–208.
96. Brooks R. EuroQol: the current state of play. *Health Policy*. 1996;37:53–72.
97. Dorman PJ, Waddell F, Slattery J, Dennis M, Sandercock P. Is the EuroQol a valid measure of health-related quality of life after stroke? *Stroke*. 1997;28:1876–1882.
98. Williams LS, Weinberger M, Harris LE, Clark DO, Biller J. Development of a stroke-specific quality of life scale. *Stroke*. 1999;30:1362–1369.
99. Lin KC, Fu T, Wu CY, Hsieh CJ. Assessing the stroke-specific quality of life for outcome measurement in stroke rehabilitation: minimal detectable change and clinically important difference. *Health Qual Life Outcomes*. 2011;9:5.
100. Hilari K, Byng S. Measuring quality of life in people with aphasia: the Stroke Specific Quality of Life Scale. *Int J Lang Commun Disord*. 2001;36 Suppl:86–91.

Clinical Interventions in Aging

Publish your work in this journal

Clinical Interventions in Aging is an international, peer-reviewed journal focusing on evidence-based reports on the value or lack thereof of treatments intended to prevent or delay the onset of maladaptive correlates of aging in human beings. This journal is indexed on PubMed Central, MedLine, the American Chemical Society's 'Chemical Abstracts Ser-

Submit your manuscript here: <http://www.dovepress.com/clinical-interventions-in-aging-journal>

vice' (CAS), Scopus and the Elsevier Bibliographic databases. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Dovepress