

Critical Appraisal and Review of the Rankin Scale and Its Derivatives

Peter Wayne New^{a, b} Rachelle Buchbinder^{b, c}

^aRehabilitation and Aged Services Program, Kingston Center, Southern Health, Melbourne,

^bDepartment of Epidemiology and Preventive Medicine, Monash University, and

^cMonash Department of Clinical Epidemiology at Cabrini Hospital, Malvern, Australia

Key Words

Stroke · Clinimetric · Outcome assessment

Abstract

Background and Purpose: Efforts to reduce stroke burden require accurate assessment of outcomes in order to compare treatments. The Rankin Scale and its derivatives, the Modified Rankin Scale and the Oxford Handicap Scale, taken together, are among the most common outcome measures that have been used in stroke research. The aim of this study was to perform a critical appraisal of the clinimetric properties of these scales. It was also planned to review the use of these scales in a selection of articles to illustrate concerns raised by the critical appraisal. **Summary of Review:** A literature search was performed using electronic databases to locate relevant articles about the reviewed scales. The scales were appraised using a structured format regarding the following properties: purpose, development, presentation, language, method of administration, content validity, face validity, feasibility, construct validity, reliability, responsiveness, and generalizability. There are concerns in each of the appraised areas regarding the clinimetric properties of these scales. **Conclusion:** Further work is needed to improve the clinimetric properties of the reviewed scales to ensure that they are more useful tools in determining the outcome of stroke. Alternatively, a newer global outcome scale with improved clinimetric properties may be a better option for future stroke research.

Copyright © 2006 S. Karger AG, Basel

Introduction

Stroke is an important public health issue. It is the second most common cause of death worldwide, after cardiovascular disorders [1]. It is currently the seventh highest cause of disability, and it is predicted that by the year 2020 it will be the fourth highest cause [2]. Because most stroke patients survive beyond the acute period, it is important to reduce the impact of stroke morbidity. Efforts to reduce stroke burden require accurate assessment of outcomes in order to compare treatments. Morbidity from stroke has many facets. Outcomes reported in stroke trials include impairments, disability, handicap, and quality of life. Each of these may be assessed using a variety of measurement tools.

The Rankin Scale (RS) [3] has been used as an outcome measure of stroke trials and routine care for over four decades. Modifications to the original scale have been made, resulting in the Modified Rankin Scale (MRS) [4] and the Oxford Handicap Scale (OHS) [5].

In order for an outcome measure to be useful in clinical trials, it must be valid for the purpose that it is intended, reliable and responsive. Some clinimetric properties of the RS [6] and the MRS [7–9] have been discussed previously, however, no comprehensive appraisal of the clinimetric properties of the RS and its derivatives has been published. In this study we performed a critical appraisal of the clinimetric properties of the RS, MRS, and OHS pertaining to their use as outcome measures in stroke trials. We also reviewed the use of these scales in a selection of articles to illustrate concerns raised by our critical appraisal.

Table 1. Description of the scales: development, purpose and application

	RS [3]	MRS [4]	OHS [5]
Country of origin	UK (Scotland)	UK	UK
Year	1957	1988	1989
Method of development	NR	Investigators' judgment	Investigators' judgment
Stated purpose	Descriptive	Evaluative	Descriptive
Original setting	Hospital	Hospital	Community
Original population	Acute stroke	Acute stroke	Acute stroke
Domain assessing	Disability	Unclear. Described as handicap but uses impairment and disability wording	Unclear. Described as handicap but uses impairment and 'lifestyle' wording
Key terms defined	no	no	no
Standardized protocol	no	no	no
Original method of administration	NR	NR	NR

RS = Rankin scale; MRS = modified Rankin scale; OHS = Oxford handicap scale; NR = not reported.

Methods

Identification of Relevant Literature

All published data that tested or reported information about one or more of the clinimetric properties of the RS, MRS and OHS were sought. Articles in any language and of any type, including letters and abstracts, were included provided that they referred to patients with a hemorrhagic or ischemic stroke.

The electronic databases Medline (1966–2004), PsycINFO (1967–2004) and CINAHL (1984–2004) were searched using the following terms: stroke, cerebrovascular disorders, outcome assessment, Rankin, Modified Rankin and Oxford Handicap. The reference section of all relevant articles identified and pertinent textbooks were also reviewed for any other useful articles. Medline searches were also performed of the authors of relevant articles for any other potentially appropriate articles.

To illustrate concerns raised from the critical appraisal, we reviewed a convenience sample of studies that used the reviewed scales as an outcome measure following stroke.

Critical Appraisal

One author (P.W.N.) reviewed the development, presentation and administration of scales and critically appraised their clinimetric properties. The following characteristics were considered in the appraisal: purpose, content validity, face validity, feasibility, construct validity, reliability, responsiveness, and generalizability. Accepted principles for appraising the clinimetric properties of health outcome scales were followed [10–12]. A modified version of a structured framework for appraisal was used to facilitate the review process [13]. Modifications were made to account for the nature of these scales covering one domain and being a single item. An item was added to cover responsiveness and the appraisal of generaliz-

ability was expanded. For each item of the critical appraisal, one of five responses was possible: yes, no, partial, don't know, or other.

Results

Search Results

The initial search identified 249 articles. Of these, 12 had relevant information specifically regarding the clinimetric properties of the reviewed scales [3–5, 14–22]. An additional 6 articles with relevant information related to the use of these scales were also identified [6–9, 23, 24].

Development, Purpose and Administration of Scales

Each scale is reproduced in appendix 1 and a description of the development, purpose and application of the scales is shown in table 1. The RS was developed in Scotland in 1957 and first used to describe recovery in a group of 192 stroke patients at discharge or transfer from hospital [3]. It consists of a single item, with five grades representing no, slight, moderate, moderately severe and severe disability.

In 1988, as part of a study of aspirin in stroke prevention, the RS was modified and renamed the MRS [25]. The modification was not reported in the aspirin study, but was described subsequently [4]. An additional grade

Table 2. Differences in administration of scales

Scale	Time post stroke	Setting	Procedure for assessment
RS	Not specified [3] Acute and 3 months [26]	Acute hospital [3] Acute hospital and outpatients/ community [26]	Direct assessment [3, 26]
MRS	7.5–15.9 days mean [27] 90 days [16, 29, 30] 1 week and 5 months [4] 6 months [17, 32] 39 months mean (range 1–60) [31] Acute hospital discharge and 43.1 months mean (range 1–228) [28]	Acute hospital and outpatients/ community [4, 16, 27, 28] Outpatients [31] Not specified [17, 29, 30, 32]	Direct assessment [17, 27] Direct assessment, mail or telephone [28] Direct assessment, chart review, and family [31] Assessment of patient, family and treating nurses [4] Assessment of patient and treating nurses [16] Direct assessment or telephone [32] Telephone assessment [33] Not specified [29, 30]
OHS	6 months [34] 1st week and 1, 6 and 12 months [35]	Not specified [34]	Semi-structured interview [34] Direct assessment or telephone [35]

was included (grade 0, no symptoms at all), because of reported concerns about a lack of grading comprehensiveness. The wording of the definitions for grades 1 and 2 were also altered because of concerns about ambiguity [5]. The changes were reportedly also made to ‘accommodate language disorders and cognitive defects, to allow comparison between patients with different kinds of neurological deficits and to add a further dimension by referring to previous activities’ [4].

The MRS was revised in 1988, for a study of stroke patients in the community, and renamed the OHS [5, 14]. While still consisting of 6 grades, the grades were renamed as follows: no symptoms, mild symptoms, and minor, moderate, moderately severe and severe handicap (replacing disability). These changes were reportedly made in order to reduce the confusion that arose by calling the MRS a handicap scale while using disability terminology. The definitions were altered to reflect this change, and to include additional important post-stroke handicaps other than mobility. A reference to lifestyle was also introduced.

All three scales have been used in a variety of settings, at different times after acute stroke, and administered using an assortment of techniques [3, 4, 16, 17, 26–35]. The differences in the administration of these scales are illustrated in table 2. The RS, MRS and OHS have been used as both descriptive outcome measures in patients who have had a stroke [3, 26–28, 31, 35–37] and also to evaluate treatment effects in secondary prevention trials of stroke therapy [29, 30, 33, 38–40].

Critical Appraisal of Clinimetric Properties

The critical appraisal of the RS, MRS, and OHS is shown in table 3. Additional information in key areas is given in the following text.

Construct Validity

The results of the MRS construct validity assessments are summarized in table 4. In evaluating the construct validity of the MRS it has been compared with several other scales. These include a measure of impairment, the National Institute of Health Stroke Scale [41], a measure of disability, the Barthel Index (BI) [42, 43], and a global outcome measure for traumatic brain injury, the Glasgow Outcome Scale (GOS) [44, 45]. The best construct validity results were found when the MRS was assessed with the GOS. No studies have been located that assess the construct validity of either the RS or the OHS.

Reliability

The reliability of a scale refers to the extent to which measurements of individuals obtained under different circumstances yield similar results [12]. There are several aspects of reliability. It is not necessary to consider internal consistency of the RS and its derivatives because it is a single item test. It is more appropriate to consider the inter-rater and intra-rater reliability. The results of studies that report the inter-rater reliability of the RS, MRS and OHS are shown in table 5.

Van Swieten [4] assessed the use of the MRS by 34 observers in 2 centers. Patients all had cerebral infarction,

Table 3. Summary of the critical appraisal of the RS, MRS and OHS

	RS [3]	MRS [4]	OHS [5]
<i>Purpose</i>			
1 Is the purpose, population, and setting clearly specified?	yes	yes	yes
<i>Content validity</i>			
1 Is the domain and all specific exclusions from this domain clearly specified?	no	no	no
Are all relevant categories included?	no	no	no
2 Is the breakdown of categories appropriate, considering the purpose?	no	no	no
Are the categories mutually exclusive?	no	no	yes
3 Was an appropriate method of development described?	no	no	no
<i>Face validity</i>			
1 Is the nomenclature used to label the grades satisfactory?	no	no	no
Are the terms used based upon empirical (i.e. directly observable) evidence?	partial	partial	no
2 Are the criteria for determining inclusion into each grade clearly specified?	no	no	no
Do these criteria appear reasonable?	don't know	don't know	don't know
Have the criteria been demonstrated to have validity and/or reliability?	don't know	don't know	don't know
3 Are the definitions of criteria clearly specified?	no	no	no
<i>Feasibility</i>			
1 Is classification simple to understand?	yes	yes	yes
2 Is classification easy to perform?	no	no	no
3 Does it rely on clinical examination alone?	no	no	no
4 Are special skills, tools, and/or training required?	yes	yes	yes
5 How many minutes does it take to perform?	5	5–15	5–15
<i>Construct validity</i>			
1 Does it discriminate between entities that are thought to be different in a way appropriate for the purpose?	don't know	don't know	don't know
2 Does it perform satisfactorily when compared to other scales which measure the same domain?	don't know	yes	don't know
<i>Reliability</i>			
1 Has the inter-observer reliability been assessed?	no	yes	yes
2 Has the intra-observer reliability been assessed?	no	yes	no
<i>Responsiveness</i>			
1 Has the responsiveness to clinically significant change been assessed?	no	yes	no
<i>Generalizability</i>			
1 Has the scale been used in other studies and/or setting?	yes	yes	yes
2 Have clinimetric properties been appraised in the other settings of use?	no	no	no

but 86 were seen within the first week of their stroke and 14 outpatients were seen within 5 months of stroke onset. In each center the observers were randomly allocated into 50 pairs. The scales were explained to the observers in a training session. The patients were assessed by direct questioning, and nursing staff or relatives were also 'sometimes questioned'. The two assessments were conducted within 6 h to avoid disagreement caused by a change in the patients' condition. Agreement was best for grades 0 and 5, probably because disagreement is only possible in one direction, and was

worst for grades 2, 3 and 4. There were no significant differences between centers.

Wolfe [16] evaluated the MRS inter-rater reliability in patients at least 3 months post-stroke, either at home or in hospital (breakdown not specified) by 3 observers who had 6 months previous experience with using the scale. The patients were questioned directly and the nursing staff caring for them were also questioned 'when necessary'. The time between assessments was not mentioned, but none of the patients received rehabilitation in the interval between assessments. Analysis of variance revealed

Table 4. MRS construct validity

Author	Scale	Comparison	n	Outcome assessment
Tilley et al. [15]	MRS (binary outcome)	NIHSS [41] (impairment scale) BI [42, 43] GOS [44, 45]	333	agreement 86%, $\Phi = 0.67$ agreement 87%, $\Phi = 0.76$ agreement 94%, $\Phi = 0.88$
Wolfe et al. [16]	MRS	BI	50	$\kappa = 0.72$ (95% CI 0.61–0.82) $\kappa_w = 0.91$ (95% CI 0.87–0.95)
De Haan et al. [17]	MRS	5 impairment scales	87	Pearson's $r^2 = 36.5\%$ (mean range 0.56 to –0.71)

NIHSS = National Institute of Health Stroke Scale; BI = Barthel Index; GOS = Glasgow Outcome Scale.

Table 5. Inter-rater reliability of RS, MRS and OHS

Author	Scale	n	Inter-rater reliability agreement
Jaillard [18]	RS	70	$\kappa_w = 0.79$. Exact agreement, κ and 95% CI all NR
Van Swieten et al. [4]	MRS	100	Overall exact agreement 65%, $\kappa = 0.56$, $\kappa_w = 0.91$ Outpatients exact agreement 87%, $\kappa = 0.82$, $\kappa_w = 0.91$ Inpatients exact agreement 61%, $\kappa = 0.51$, $\kappa_w = 0.89$
Wolfe et al. [16]	MRS	50	Exact agreement 80%, $\kappa_w = 0.90$ (95% CI = 0.84–0.97)
Wilson et al. [19]	MRS	58	Before structured interview Overall agreement 57%, $\kappa = 0.44$, $\kappa_w = 0.78$ (95% CI 0.53–1.0) After structured interview Overall agreement 78%, $\kappa = 0.70$, $\kappa_w = 0.93$ (95% CI 0.67–1.0)
Newcommon et al. [21]	MRS	34	Direct observation of patients, $\kappa = 0.72$ (95% CI = 0.55–0.89) Telephone assessment, $\kappa = 0.30$ (95% CI = 0.13–0.47) and $\kappa = 0.38$, (95% CI = 0.21–0.55)
Bamford et al. [5]	OHS	NR	$\kappa = 0.42$, $\kappa_w = 0.72$

NR = Not reported; CI = confidence interval.

that there was strong evidence of a systematic difference between the raters ($F_{2,48} = 6.02$, $p = 0.005$), with raters 1 and 3 estimating the grade 0.42 and 0.33 points (standard error 0.13) higher than rater 2.

Wilson [19] developed a structured interview to improve the inter-rater reliability of the MRS. A group of stroke survivors who were between 6 and 24 months post-stroke were rated on the MRS. Assessments were done by 2 observers before training in a structured interview technique and then again after 8 weeks. The inter-rater reliability improved significantly after training in the structured interview.

Recently, Newcommon et al. [21] compared direct observation of the MRS with a structured interview technique for telephone assessment. The inter-rater reliability was assessed on the two techniques. Patients were 3–6 months post-stroke and the telephone assessment was conducted 5 days after the direct assessment. Not unexpectedly, the direct observation method was much more reliable.

The intra-rater reliability has only been reported for the MRS [16]. A sample of 14 patients was assessed twice by the same observer within a 2-week period at least 3 months after the stroke. This time interval was chosen to minimize the effect of patient progression, improve-

Table 6. Differences in dichotomizing and trichotomizing the RS and MRS

Scale	Cut-off and descriptors for subgroups	Justification given
RS	0–2 = nondisabling stroke, 3–5 = severe disability [39, 40]	no
MRS	0–1 = favorable [33, 38]	yes [15]
	0–4 = functionally at least partially independent, 5 = dependent [37]	no
	0–2 = nondisabling stroke, 3–5 = disabling stroke [31], independence, dependence [48]	no
	0–3 = minor stroke, 4–5 = major stroke [28], severe disability [27]	no
	0–1, 2–3, 4–5 [30, 33, 49]	no

ment or observer recall. There was exact agreement in 86% of observations, with a κ_w of 0.95 (95% CI = 0.88–1.00).

Responsiveness

For scales to be useful in evaluating the outcome of interventions, they must be responsive to a clinically significant change in a patient's condition. For a study with independent group design in which only test results obtained after the intervention are available, an instrument must be able to differentiate between subjects and within subjects over time. If both pre- and postintervention results are available, then the instrument only needs to detect within subject change over time [46].

No studies have been located evaluating the responsiveness of the RS or OHS, or what the minimally important change in grade scores for each of these scales should be. The responsiveness of the MRS has been recently evaluated in a group of 95 acute stroke patients in rehabilitation who were a mean of 9 days post stroke [22]. The responsiveness of the MRS was compared with that of other scales, including the BI and the Functional Independence Measure (FIM) [47]. The MRS detected change in 55 patients, the BI detected change in 71 patients and the FIM detected change in 91 patients. When compared to the FIM, the receiver operating characteristics analysis showed that the MRS (C-statistic $C = 0.59$) was much less sensitive to change compared with the BI (C-statistic $C = 0.82$), indicating a corresponding lower sensitivity and specificity for the MRS.

In evaluative trials, grading results have only been obtained on patients after a stroke treatment had been given. No account has been taken of patients' premorbid functioning, although this should be addressed by randomization. In two descriptive studies, attempts have been made at assessing premorbid functioning using the RS [26] and MRS [23].

When using the RS or MRS, some authors have advocated dichotomizing or trichotomizing the results. That is, using various grade scores to categorize patients into 2 or 3 subgroups, respectively. Dichotomization has been justified by arguing that this allows the formation of a global test based on combined measures where no single outcome would provide sufficient evidence of efficacy and where a drug with potentially serious side effects was used [15]. Table 6 illustrates the different cut-off scores and the descriptors used for the resultant subgroups in a selection of studies [15, 27, 28, 30, 31, 33, 37–40, 48, 49].

Referencing Errors

In preparing this review, errors and inconsistencies have been found in some of the papers regarding the referencing of scales [16, 17, 26–28, 31, 34, 39, 40, 50, 51] (table 7). These errors make interpretation and comparisons involving these studies difficult.

Discussion

Development, Purpose and Administration of Scales

It is an accepted aspect of clinimetrics that the requirements for maximizing one of the functions of discrimination, prediction or evaluation may impede optimizing the properties of the other functions. Validating an outcome measure for one of the above purposes does not automatically ensure that it can be used for the others [52]. This issue has not been addressed adequately by those who have developed and used the reviewed scales.

None of the reasons given for the modification of the RS to the MRS appear justifiable. There is no apparent advantage of the MRS in accommodating language or cognitive defects any better than the RS, as no reference to these functions is made in either scale. The RS does

Table 7. Referencing concerns detected in critiqued papers

Paper(s)	Scale(s) reportedly used	Scale referenced
39, 50	RS	MRS
16, 17, 27, 28, 31	RS and MRS	MRS
26	Initially MRS. Subsequently RS	RS
34	RS then MRS and subsequently RS Table reproduced OHS	RS, OHS
40	RS	Unclear. Refers to another paper [34]
51	RS	Reproduced scale does not correspond to RS, MRS or OHS
29	MRS, then RS	RS

include a reference to ‘previous activities’ in its grading criteria, but with a different emphasis to the MRS. In the subsequent modification of the MRS to the OHS, the reference to ‘lifestyle’ in the OHS is also problematic. This term has not been shown to equate with handicap, and could be influenced by impairment, disability or handicap factors.

None of the scales have associated written guidelines on when, where or how they should be administered. This has resulted in a wide range of practices among users of these scales that could influence the grading result. The recent publication of a structured interview for the MRS has been shown to significantly improve the inter-rater reliability [19]. This should help to improve the quality of results from clinical trials using this measure in the future. It would be ideal if stroke trialists who have had practical experience with the use of the RS and OHS established working parties to develop, validate, and publish formal structured interviews for use with these scales.

It is difficult to accurately and fully assess the disability or handicap of patients while they are still in hospital. This applies especially to acute hospitals, and to a lesser extent in rehabilitation hospitals. It is easier to assess self-care abilities in hospital than domestic or community activities. Patients would need to be in their home environment before they would have an opportunity of attempting all their previous activities and roles in a relevant manner. This raises a concern about the appropriateness of assessing handicap while in hospital. This issue was acknowledged by the developers of the OHS [5], but it has still not been adequately addressed.

In administering these scales no guidance is given regarding whether patients can have rehabilitation to improve their performance. Providing aids, equipment or

modifying an environment are part of the rehabilitation process and can improve patients’ ability to compensate for some of the consequences of stroke. It is unclear whether the term ‘without assistance’, which is used in the RS and the MRS, allows for aids or modifications, or whether it refers only to assistance from another person. Likewise, it is not specified whether it is permissible to use aids, equipment or modifications to achieve a higher grade in the OHS.

Stroke survivors with persisting deficits sometimes have somebody else perform tasks that they can do, but which require excessive time or effort to complete. None of the scales specify whether the patients should be assessed regarding what they actually do perform for themselves, or what they can do, but choose not to for practical reasons.

It is generally accepted that most recovery after stroke takes place over the first 3 months and can then continue more slowly for at least 6–12 months after the stroke [53]. Assessing patients before 3 months may not give a true reflection of outcome. Some of the critiqued papers have used these scales earlier than 3 months after stroke [27, 31]. A previous review of stroke outcome measures has recommended that assessments are performed at 6 months [9].

Until recently, none of the methods of assigning a patient their grade on the reviewed scales that does not involve direct observation has been assessed for reliability. Some papers on stroke outcome have graded patients using telephone assessments. There are numerous concerns regarding these papers that are discussed in detail below. A recent letter by Newcommon et al. [21] highlighted the poor inter-rater reliability using the MRS that occurs when a telephone assessment is performed on stroke patients. It has also been reported that moderate and severe

disability may go undetected by telephone assessment [54].

One paper reports that ‘several studies have validated telephone assessment of the outcome of stroke’ [33]. Two of the references given to support this claim [54, 55], however, do not relate to the RS, MRS or the OHS, but to other scales, including the BI. Validating telephone assessment of the BI does not validate this method of administration for other scales. Another reference on validating telephone assessment reported that it evaluated the RS for intermodality agreement between telephone and direct observation [51]. The abstract reports to have used the MRS, but the text only mentions the RS. The authors reproduce the scale that they used, but this has different wording and numbering from both the RS and the MRS. The intermodality evaluation was done with a different person performing the assessments and it is not reported whether a structured questionnaire or interview was used to ensure consistency. An abstract reported telephone assessment using the RS in 70 patients, or carers, 6 months after stroke, by a neurologist and a general practitioner [18]. No confidence intervals, kappa or exact agreement, however, were given. Also, it is not mentioned whether the interview was structured.

Finally, there are reported to be differences in patients’ functioning depending on who conducts the assessment. Patients’ self-assessment of their disability tends to be less than that of their carers [56]; carers tended to report lower patient abilities than the assessments made by occupational therapists [57], and doctors’ rating of patient abilities from clinical interviews tend to be higher than nurses’ ratings from patient observation [58]. Consistency of administration technique is therefore important, but lacking, in the use of the reviewed scales.

Content Validity

There is a focus in both the RS and the MRS on mobility and continence. There is no consideration of other important and common stroke-related disabilities, such as communication and cognition. These omissions are significant and limit the content validity of these scales.

There are concerns about the terminology used in all reviewed scales. The RS and MRS, which refer to ‘disability’, do not specifically define what is meant by ‘usual duties and activities’. This could refer to activities and roles in either the disability or handicap domains. The MRS and OHS use terminology which refers to both impairment (‘symptoms’) and disability terms, while reporting to measure handicap [4, 5]. The OHS reference to ‘lifestyle’ does not justify its claim to be measuring

handicap. Lifestyle has not been shown to equate with handicap and can be influenced by both disability and handicap factors. There are no references to handicap-related roles in any of the grades using the definitions which were in use at the time the MRS and OHS were developed [59]. As discussed by others, these problems create uncertainty regarding exactly what domains of impairment, disability and handicap these scales are measuring [60–62].

It has been suggested that the OHS may measure ‘overall disablement’, rather than handicap [61]. It is probably more appropriate to consider these scales, particularly the OHS, as a global outcome measure, rather than measuring a specific domain.

Face Validity

Many of the grades in all scales require some degree of subjective judgement. No concise definitions or guidelines have been produced to ensure consistency and reduce subjectivity in grading patients.

The issue of whether these scales measure disability or handicap raises an important face validity consideration. Both domains are acknowledged as being multidimensional constructs [59]. It is unresolved and somewhat controversial whether it is possible to adequately measure a multidimensional construct with a single item outcome measurement scale [6].

The reviewed scales have been used in numerous multi-center [29, 30, 39, 40] and local trials [17, 26–28, 31, 34, 35] in countries with languages other than English as their official language. No reports have been located that describe the translation of these scales into languages other than English, and then confirm the adequacy of the translated scales’ clinimetric properties in these other languages. This raises the possibility of reduced reliability and validity when these scales are used in non-English-speaking countries [12].

Finally, the OHS contains a reference to ‘himself’ in grade 2. A gender neutral term would seem more appropriate.

Construct Validity

Construct validity is assessed by comparing the reviewed scale to other scales that measure the same domain. Disability and handicap are multidimensional concepts and there are numerous scales that have been developed to measure them.

The use of the BI for assessing the construct validity of the MRS raises some concerns. It is a disability scale that focuses on self-care, continence and mobility. It does

not assess communication, cognition or domestic and community activities that are often affected by stroke. As a minimum, a scale that assesses construct validity in stroke outcomes should include self-care, continence, mobility, communication and cognition.

The results of Tilley et al. [15] suggest that the MRS is more closely related to the GOS than it is to disability, and that it is also related to impairment. These results raise further concern about the construct validity of the MRS.

An empirical comparison of BI with MRS was derived by assigning the most common MRS grade to patients with a given BI score [16], but no supporting justification was given for this approach. The kappa values lend support to the assertion that the MRS is closer to a disability scale than a handicap scale.

When the MRS was compared with 5 impairment scales [17], as would be expected, there was not a strong correlation. The study claimed that this reinforces the hierarchical nature of the World Health Organization's classification of impairment, disability and handicap [26]. Nothing, however, can be concluded about what the MRS might measure.

Reliability

The studies of the MRS inter-rater reliability by Van Swieten et al. [4] and Wolfe et al. [16] did not include a homogeneous study sample because they consisted of patients both in hospital and at home. This could have artificially increased the reliability of the results obtained [12]. Although the results reported for the intra-rater reliability of the MRS appear reasonable, the sample size was quite small.

The study of inter-rater reliability of the OHS was based on written case histories only [5]. No evidence was presented to justify this method of assessing reliability. If the case histories were of heterogeneous patients, then this could have improved the values obtained. No mention was made of the numbers in the study sample or the exact agreement.

Responsiveness

There has been no assessment of the responsiveness to change of either the RS or the OHS, and the performance of the MRS is relatively poor. There is no information about what the minimally important change in grades should be for any of the scales. These issues raise the possibility that clinical trials of stroke interventions using these outcome measures may fail to detect a clinically significant difference between stroke treatments. That is,

a type II error could occur because the outcome measure lacks sensitivity. There is also a serious concern about the use of dichotomization of outcome using the reviewed scales because this approach to reporting would be even less responsive to change.

The use of dichotomization and a selective cut-off to make the difference between a positive and negative trial has been discussed in detail previously [8, 9, 24]. It has been shown that when using an MRS score of 0 or 1 to define a 'favorable outcome', that 15% of patients reach this endpoint. However, 42% of patients achieve a 'favorable outcome' when an MRS score of 0, 1 or 2 is used [24]. It has been recommended that a 'favorable outcome' using the MRS be defined as an improvement in grade, rather than using a dichotomization process [24].

An alternative to using improvement in grade as a positive outcome is to use prognostic dichotomization. Different criteria are set for a positive outcome based on stroke severity [20]. Using this approach, patients with a milder stroke have the dichotomization for a positive response set at a higher level of function, e.g. MRS 0 or 1. As the initial severity of stroke increases, the cut-off for a positive outcome would be set at a lower level. For example, in patients with a severe stroke, an MRS of 0–4. This process would also require an agreed approach to classifying patients into prognostic groups.

There are significant differences in the burden of care and community supports that would be required for a stroke survivor to live independently, or supported in the community, in moving from a 5 to a 4, from a 4 to a 3 or a 3 to a 2, on any of the reviewed scales. It could be argued that improving the outcomes of the worst-affected group of patients would be more desirable from a health-economics and social perspective. The ability to detect these differences would be negated by dichotomizing or trichotomizing outcomes.

Acute stroke intervention trials generally aim to increase neurological recovery and improve function. It is unrealistic to expect a complete and full recovery in all patients. Therefore, it is important to consider in outcome assessment improvements that are of significant functional benefit, but less than a complete recovery.

Finally, there is a concern about the use of dichotomization in one of the reviewed studies where the dichotomization was based on a comparison of self-reported independence or dependence in activity of daily living tasks with the MRS grade [32]. The assessment was performed either in person or by telephone. The activities included were not specifically defined and there was a potential for

reduced reliability by having self-reporting of dependence and different methods of grading patients.

Strokes are more common in the elderly, and most deaths from stroke occur in those over 65 years of age [63]. Older patients who survive stroke may have had a pre-existing disability or handicap related to other conditions. The scales discussed do not allow for this possibility to be factored into the grading process and could affect the responsiveness of these scales to detect a significant change in level of functioning.

Because multiple variables can influence stroke recovery, it is necessary to characterize patients both on dimensions that are prognostic for outcomes chosen, and for those that might affect response to treatment. Even though randomization should theoretically account for differences between groups in previous functioning that can influence stroke outcome, this may not always be the case. Randomization may not adequately ensure that the treatment and control groups have the same mix of premorbid function in their subject because various biases can cause differences [64]. Consideration should therefore be given to stratifying patients on the basis of previous functioning. The recommendation has been made previously that allowance needs to be made in stroke trials for pre-stroke level of function because this factor has been shown to influence stroke recovery [23].

Conclusion

This review has examined the clinimetric properties of the RS, MRS, and the OHS, and identified concerns about the language, method of administration, validity, feasibility, reliability and responsiveness of all the scales. Because of these concerns, the appropriateness of using these scales as an outcome measure in their current format is questionable.

It may be possible to build on the concepts behind these scales and improve on their weaknesses. The concerns raised in this review could be addressed by formal clinimetric studies that implement changes to improve the properties of these scales. These changes would need to be validated as part of this refinement process.

Alternatively, a more appropriate global outcome scale that can be used as a descriptive or evaluative measure of stroke could be developed. A new scale should be developed in a way that utilizes accepted principles of clinimetrics from the outset. This development would need to be done in a way that is more thorough than has been with the reviewed scales. A new scale should ide-

ally consider the World Health Organization's International Classification of Function in the development process [65].

Acknowledgement

The reproduction of the Rankin Scale [3], the Modified Rankin Scale [4] and the Oxford Handicap Scale [5] was kindly granted with permission of the publishers.

Appendix 1

Grade Description

Rankin Scale [3]

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

Modified Rankin Scale [4]

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

Oxford Handicap Scale [5]

- 0 No symptoms
- 1 Minor symptoms that do not interfere with lifestyle
- 2 Minor handicap, symptoms that lead to some restriction in lifestyle but do not interfere with the patient's capacity to look after himself
- 3 Moderate handicap, symptoms that significantly restrict lifestyle and prevent totally independent existence
- 4 Moderately severe handicap, symptoms that clearly prevent independent existence though not needing constant attention
- 5 Severe handicap, totally dependent patient requiring constant attention night and day

Modifications from RS to MRS highlighted in italic.

References

- 1 Murray CJL, Lopez AD: Mortality by cause for eight regions of the world: global burden of disease study. *Lancet* 1997;349:1269–1276.
- 2 Murray CJ, Lopez AD: *The Global Burden of Disease: a comprehensive assessment of mortality and disability from disease, injuries and risk factors in 1990 and projected to 2020*. Cambridge, Harvard School of Public Health on behalf of the World Health Organization and the World Bank, 1996.
- 3 Rankin J: Cerebral vascular accidents in people over the age of 60. II. Prognosis. *Scot Med J* 1957;2:200–215.
- 4 Van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJA, Van Gijn J: Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* 1988;19:604–607.
- 5 Bamford JM, Sandercock AG, Warlow CP, Slattery J: Interobserver agreement for the assessment of handicap in stroke patients (letter). *Stroke* 1989;20:828.
- 6 Lyden PD, Lau GT: A critical appraisal of stroke evaluation and rating scales. *Stroke* 1991;22:1345–1352.
- 7 D'Olhaberriague L, Litvan I, Mitsias P, Mansbach HH: A reappraisal of reliability and validity studies in stroke. *Stroke* 1996;27:2331–2336.
- 8 Sulter G, Steen C, De Keyser J: Use of Barthel index and modified rankin scale in acute stroke trials. *Stroke* 1999;30:1538–1541.
- 9 Duncan PW, Jorgensen HS, Wade DT: Outcome measures in acute stroke trials. A systematic review and some recommendations to improve practice. *Stroke* 2000;31:1429–1438.
- 10 Bombardier C, Tugwell P: Methodological considerations in functional assessment. *J Rheumatol* 1987;14(suppl 15):6–10.
- 11 Feinstein AR: *Clinimetrics*. Yale University Press, 1987.
- 12 Streiner D, Norman G: *Health Measurement Scales: A Practical Guide to Their Development and Use*, ed. 3. New York, Oxford University Press, 2003.
- 13 Buchbinder R, Goel V, Bombardier C, Hogg-Johnson S: Classification systems of soft tissue disorders of the neck and upper limb: Do they satisfy methodological guidelines? *J Clin Epidemiol* 1996;49:141–149.
- 14 Bamford JM, Vessey M, Fowler G, Molyneux A, Hughes T, Burn J, et al: A prospective study of acute cerebrovascular disease in the community: The Oxfordshire Community Stroke Project 1981–1986. 1. Methodology, demography and incident cases of first-ever stroke. *J Neurol Neurosurg Psychiatry* 1988;51:1373–1380.
- 15 Tilley BC, Marler J, Geller NL, Lu M, Legler J, Brott T, et al: Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke t-PA stroke trial. *Stroke* 1996;27:2136–2142.
- 16 Wolfe CDA, Taub NA, Woodrow EJ, Burney PGJ: Assessment of scales of disability and handicap for stroke patients. *Stroke* 1991;22:1242–1244.
- 17 De Haan R, Horn J, Limburg M, Van Der Meulen J, Bossuyt P: A comparison of five stroke scales with measurement of disability, handicap and quality of life. *Stroke* 1993;24:1178–1181.
- 18 Jaillard AS on behalf of the MAST-E group: Value of phone interview in stroke outcome assessment (abstract). *Cerebrovasc Dis* 1995;5:269.
- 19 Wilson JTL, Hareendran A, Grant M, Baird T, Schulz UGR, Muir KW, et al: Improving the assessment of outcomes in stroke. Use of a structured interview to assign grades on the Modified Rankin Scale. *Stroke* 2002;33:2243–2246.
- 20 Berge E, Barer D: Could stroke trials be missing important treatment effects? *Cerebrovasc Dis* 2002;13:73–75.
- 21 Newcommon NJ, Green TL, Haley E, Cooke T, Hill MD: Improving the Assessment of Outcomes in Stroke: Use of a Structured Interview to Assign Grades on the Modified Rankin Scale (letter). *Stroke* 2003;34:377–378.
- 22 Dromerick AW, Edwards DF, Diringer MN: Sensitivity to changes in disability after stroke: A comparison of four scales useful in clinical trials. *J Rehabil Res Dev* 2003;40:1–8.
- 23 McNaughton H, Weatherall M, Taylor W, McPherson K: Factors influencing rate of Barthel Index change in hospital following stroke. *Clin Rehabil* 2001;15:422–427.
- 24 Lai SM, Duncan PW: Stroke recovery profile and the Modified Rankin Scale. *Neuroepidemiology* 2001;20:26–30.
- 25 UK-TIA Study Group: The UK-TIA aspirin trial: Interim results. *Br Med J* 1988;296:316–320.
- 26 Di Carlo A, Lamassa M, Pracucci G, Basile AM, Trefoloni G, Vanni P, et al: Stroke in the very old. Clinical presentation and determinants of 3-month functional outcome: a European perspective. *Stroke* 1999;30:2313–2319.
- 27 Pinto AN, Melo TP, Lourenço ME, Leandro MJ, Brázio A, Carvalho L, et al: Can a clinical classification of stroke predict complications and treatments during hospitalization? *Cerebrovasc Dis* 1998;8:204–209.
- 28 Ferro JM, Crespo M: Prognosis after transient ischaemic attack and ischaemic stroke in young adults. *Stroke* 1994;25:1611–1616.
- 29 Hacke W, Kaste M, Fieschi C, Toni D, Lesaffre E, Von Kummer R, et al: Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric stroke: the European Cooperative Acute Stroke Study (ECASS). *JAMA* 1995;274:1017–1025.
- 30 Steiner T, Bluhmki E, Kaste M, Toni D, Trouillas P, Von Kummer R, et al: The ECASS 3-hour cohort. Secondary analysis of ECASS data by time stratification. *Cerebrovasc Dis* 1998;8:198–203.
- 31 Salgado AV, Ferro JM, Gouveia Oliveira A: Long-term prognosis of first-ever lacunar strokes. a hospital-based study. *Stroke* 1996;27:661–666.
- 32 Kay R, Wong KS, Perez G, Woo J: Dichotomizing stroke outcomes based on self-reported dependency. *Neurology* 1997;49:1694–1696.
- 33 Kwiatkowski TG, Libman RB, Frankel M, Tilley BC, Morgenstern LB, Lu M, et al: Effects of tissue plasminogen activator for acute ischemic stroke at one year. *N Engl J Med* 1999;340:1781–1787.
- 34 De Haan R, Limburg M, Bossuyt P, Van Der Meulen J, Aaronson N: The clinical meaning of Rankin 'handicap' grades after stroke. *Stroke* 1995;26:2027–2030.
- 35 Samuelsson M, Söderfeldt B, Olsson GB: Functional outcome in patients with lacunar infarction. *Stroke* 1996;27:842–846.
- 36 Adams GF: Grading of the functional recovery from strokes. *Age Ageing* 1975;4:137–141.
- 37 Paciaroni M, Arnold P, Van Melle G, Bogousslavsky J: Severe disability at hospital discharge in ischaemic stroke survivors. *Eur Neurol* 2000;43:30–34.
- 38 The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group: Tissue plasminogen activator for acute ischaemic stroke. *N Engl J Med* 1995;333:1581–1587.
- 39 The Multicenter Acute Stroke Trial Europe Study Group: Thrombolytic therapy with streptokinase in acute ischaemic stroke. *N Engl J Med* 1996;335:145–150.
- 40 North American Symptomatic Carotid Endarterectomy Trial Collaborators: Benefit of carotid endarterectomy in patients with symptomatic moderate or severe stenosis. *N Engl J Med* 1998;339:1415–1425.
- 41 Brott T, Adams P, Olinger CP: Measurement of acute cerebral infarction: a clinical examination scale. *Stroke* 1989;20:864–870.
- 42 Mahoney FI, Barthel BW: Functional evaluation: the Barthel index. *Md Med J* 1965;14:61–65.
- 43 Wade DT, Collin C: The Barthel ADL index: a standard measure of physical disability? *Int Disabil Studies* 1988;10:64–67.
- 44 Jennett B, Bond M: Assessment of outcome after severe brain damage. a practical scale. *Lancet* 1975;i:480–484.
- 45 Jennett B, Snoek J, Bond MR, Brooks N: Disability after severe head injury: observations on the use of the Glasgow Outcome Scale. *J Neurol Neurosurg Psychiatry* 1981;44:285–293.
- 46 Guyatt GH, Walter S, Norman G: Measuring change over time: assessing the usefulness of evaluative instruments. *J Chron Dis* 1987;40:171–178.
- 47 Guide for the Uniform Data Set for Medical Rehabilitation (including the FIM instrument), version 5.1. Buffalo, NY, State University of New York at Buffalo, 1997.

- 48 Buchan AM, Barber PA, Newcommon N, Karbalai HG, Demchuk AM, Hoyte M, et al: Effectiveness of t-PA in acute ischemic stroke. Outcomes relates to appropriateness. *Neurology* 2000;54:679–684.
- 49 Hacke W, Bluhmki E, Steiner T, Tatlisumak T, Mahangne M-H, Sacchetti M-L, et al: Dichotomized efficacy end points and global end-point analysis applied to the ECASS intention-to-treat data set: post hoc analysis of ECASS 1. *Stroke* 1998;29:2073–2075.
- 50 Broderick JP, Lu M, Kothari R, Levine SR, Lyden PD, Haley EC, et al: Finding the most powerful measures of the effectiveness of tissue plasminogen activator in the NINDS tPA stroke trial. *Stroke* 2000;31:2335–2341.
- 51 Candelise L, Pinaridi G, Aritzu E, Musicco M: Telephone interview for stroke outcome assessment. *Cerebrovasc Dis* 1994;4:341–343.
- 52 Kirshner B, Guyatt G: A methodological framework for assessing health indices. *J Chron Dis* 1985;38:27–36.
- 53 Gresham GE, Duncan PW, Stason WB, Adams HP Jr, Adelman AM, Alexander DN, et al: Post-Stroke Rehabilitation. Clinical Practice Guideline, No 16, ed 1. Rockville, US Department of Health and Human Services, Agency for Health Care Policy and Research. AHCPR publ No 95-0662, 1995.
- 54 Korner-Bitensky N, Wood-Dauphinee S, Siemietycki J, Shapiro S, Becker R: Health-related information postdischarge: telephone versus face-to-face interviewing. *Arch Phys Med Rehabil* 1994;75:1287–1296.
- 55 Shinar D, Gross CR, Bronstein KS, Licata-Gehr EE, Eden DT, Cabrera AR, et al: Reliability of the activities of daily living scale and its use in telephone interview. *Arch Phys Med Rehabil* 1987;68:723–728.
- 56 Knapp P, Hewison J: Disagreement in patient and carer assessment of functional abilities after stroke. *Stroke* 1999;30:934–938.
- 57 Wyller TB, Sveen U, Bautz-Holter E: The Barthel ADL index one year after stroke: comparison between relatives' and occupational therapists' scores. *Age Ageing* 1995;24:389–401.
- 58 Ranhoff AH, Laake K: The Barthel ADL index scoring by the physician from patient interview is not reliable. *Age Ageing* 1993;22:171–174.
- 59 World Health Organization: International Classification of Impairments, Disabilities and Handicaps: A Manual of Classification Relating to the Consequences of Disease. Geneva, World Health Organization, 1980.
- 60 Bloch RF: Interobserver agreement for the assessment of handicap in stroke patients (letter). *Stroke* 1988;19:1448.
- 61 Burn JPS: Reliability of the Modified Rankin Scale (letter). *Stroke* 1992;23:438.
- 62 Wade DT: Measurement in Neurological Rehabilitation. Oxford, Oxford University Press, 1992.
- 63 Bonita R: Epidemiology of stroke. *Lancet* 1992;339:342–347.
- 64 Jadad A: Randomised Controlled Trials. London, BMJ Books, 1998.
- 65 World Health Organization: ICIDH-2: International Classification of Functioning, Disability and Health. Final draft. Geneva, WHO, 2001.