# Deriving Modified Rankin Scores From Medical Case-Records

Terence J. Quinn, Gautamananda Ray, Sari Atula, Matthew R. Walters, Jesse Dawson
and Kennedy R. Lees

The online version of this article, along with updated information and services, is
located on the World Wide Web at:
http://stroke.ahajournals.org/content/39/12/3421

# Deriving Modified Rankin Scores From Medical Case-Records

Terence J. Quinn, MRCP; Gautamananda Ray, MRCP; Sari Atula, MD; Matthew R. Walters, MD; Jesse Dawson, MRCP; Kennedy R. Lees, MD

***Background and Purpose***—Modified Rankin score (mRS) is traditionally graded using a face-to-face or telephone interview. Certain stroke assessment scales can be derived from a review of a patient's case-record alone. We hypothesized that mRS could be successfully derived from the narrative within patient case-records.

***Methods***—Sequential patients attending our cerebrovascular outpatient clinic were included. Two independent, blinded clinicians, trained in mRS, assessed case-records to derive mRS. They scored "certainty" of their grading on a 5-point Likert scale. Agreement between derived and traditional face-to-face mRS was calculated using attribute agreement analysis.

***Results***—Fifty patients with a range of disabilities were included. Case-record appraisers were poor at deriving mRS ($k=0.34$ against standard). Derived mRS grades showed poor agreement between observers ($k=0.33$). There was no relationship between certainty of derived mRS and proportion of correct grades ($P=0.727$).

***Conclusion***—Accurate mRS cannot be derived from standard hospital records. Direct mRS interview is still required for clinical trials. (***Stroke***. 2008;39:3421-3423.)

**Key Words:** clinical trials ■ drug trials ■ methodology ■ outcomes ■ randomized, controlled trials ■ scales ■ therapy ■ treatment

The modified Rankin score (mRS) is the preferred disability outcome scale in acute stroke trials.[1] Traditionally, mRS grading has been based on face-to-face or telephone interviews. Such an approach is possible for a prospective trial but does not allow for retrospective disability grading. Previous observational studies have attempted to derive mRS using information contained in patient case-records.[2,3] The clinometric properties of such an approach have not been described.

Assessment of functional capacity is an important element of stroke clinic review. As each mRS grade describes a broad range of disability, reasonable estimation from narrative case-record information should be possible. Several stroke assessment scales in common use can be successfully derived from routinely collected data. The National Institutes of Health Stroke Scale,[4] the Canadian Neurological Scale,[5] and the Scandinavian Stroke Scale[6] score have all been derived with acceptable validity and reliability.

We hypothesized that mRS could be derived accurately and reliably from information recorded at outpatient follow-up.

## Methods

All discharges from our urban teaching hospital acute stroke unit are allocated 3-month outpatient hospital clinic follow-up. A sequential series of these outpatients consented to participate in a study of video-based mRS assessment. From this trial population, we further selected patients for inclusion in the mRS derivation study using an online random sampling process (www.random.org). The study had full ethical approval with patients or proxies providing written consent.

Patients attending the clinic for their routine consultation were first seen and managed according to normal practice. Doctors leading the outpatient consultation were not aware that their case-record notes would be used for retrospective analysis. We provide no guidance on documentation during clinic review.

Two independent stroke physicians derived mRS grades from the case-records, blinded to mRS grades. They were given access to complete case-records with no external editing unless explicit reference was made to the mRS. In addition to mRS grading, researchers documented degree of confidence in their assessment using a 5-point Likert scale that ranged from 0="not at all confident" to 5="extremely confident." All clinicians involved were fully trained in mRS assessment.[7]

For each outpatient, full mRS assessment was video-recorded immediately after the routine consultation. These interviews were performed according to the recommendations of the mRS training program by certified raters;[7] half of the interviews followed a structured format.[8] Four stroke physicians and 3 research nurses later reviewed these video recordings and independently assigned mRS grades with final "correct" mRS decided by majority scoring. (Figure).

We calculated agreement between correct mRS and derived mRS and interobserver variability using attribute agreement analysis. Accuracy of mRS grading was described by calculating mean and SD of actual mRS for derived mRS grades. We

*Stroke* is available at http://stroke.ahajournals.org     DOI: 10.1161/STROKEAHA.108.519306
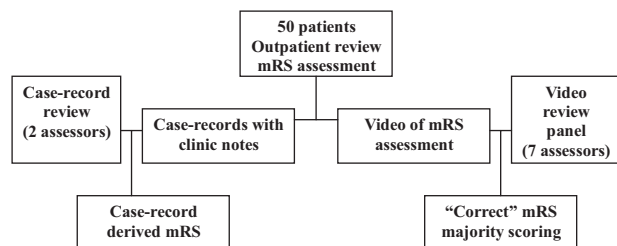
**Figure.** Schematic diagram of evaluation process for case-record-derived mRS versus "correct" mRS.

performed all statistical analysis using Minitab software (version 13.1; Minitab Inc).

## Results

Fifty patients were selected with a median age of 78 years (range, 30 to 92 years); median mRS was 2. The group comprised a variety of stroke subtypes (7 total anterior circulation syndrome; 16 partial anterior cerebral syndrome; 4 posterior circulation syndrome; 17 lacunar syndrome; 6 unclassified). Patients were reviewed at a median of 16 weeks (range, 2 to 56 weeks) from index stroke event. One patient withdrew consent after interview and was not included in the final analysis. To ensure there was no recall bias, we excluded 4 patients in which one or both of the case-record reviewers had been involved in their care.

Both reviewers were confident in their grading (mean confidence 3.2 out of 5; reviewer 1=2.6 out of 5; reviewer 2=3.8 out of 5). There was no relationship between certainty of derived mRS and proportion of correct grades ($P=0.727$). Derived mRS showed poor agreement with correct grade (overall $k=0.34$; appraiser 1 $k=0.35$; appraiser 2 $k=0.31$) and between observers ($k=0.33$; Table 1). Agreement was greatest at extremes of mRS. Case-record reviewers tended to underscore disability (Table 2).

## Discussion

We have shown that mRS derived from patient case-records has unacceptable accuracy and reliability for use in clinical research. This contrasts with other commonly used stroke scales, in which quantitative outcome data have been reliably described using qualitative case-record information.

Scales that have been successfully derived from case-records measure physical impairment only. Transforming bedside neurological examination into a quantitative scale is

**Table 1. Agreement With 'Correct' mRS and Agreement Between Observers for Case-Record-Derived mRS**

| mRS | Agreement With Standard, $k$ | Interobserver Agreement, $k$ |
|---|---|---|
| 0 | 0.52 | 0.49 |
| 1 | 0.24 | 0.06 |
| 2 | 0.27 | 0.15 |
| 3 | 0.34 | 0.48 |
| 4 | 0.28 | 1.00 |
| 5 | N/A | N/A |
| Total | 0.33 | 0.34 |

N/A indicates not applicable (no patients with mRS 5 level of disability were included).

**Table 2. Accuracy (Mean and 95% CI) for Derived mRS Versus 'Correct' mRS**

| 'Correct' mRS | Combined Derived mRS | Reviewer 1 Derived mRS | Reviewer 2 Derived mRS |
|---|---|---|---|
| 0 (n=8) | 0.31 (−0.11–0.73) | 0.5 (−0.39–1.39) | 0.13 (−0.17–0.42) |
| 1 (n=7) | 1.00 (0.55–1.45) | 1.29 (0.41–2.17) | 0.71 (0.26–1.17) |
| 2 (n=17) | 1.29 (0.99–1.60) | 1.59 (1.11–2.07) | 1.00 (0.64–1.36) |
| 3 (n=9) | 2.00 (1.20–2.80) | 2.43 (1.14–3.72) | 1.67 (0.45–2.88) |
| 4 (n=5) | 3.22 (2.71–3.73) | 3.25 (2.45–4.04) | 3.20 (2.16–4.24) |
| 5 (n=0) | N/A | N/A | N/A |

N/A indicates not applicable (no patients with mRS 5 level of disability were included).

straightforward if comprehensive physical examination is recorded. As a global disability scale, mRS review requires measures of physical, cognitive, emotional, and functional status. Such data may not always be recorded during a busy outpatient assessment.

Although case-record reviewers were poor at deriving mRS, it is interesting that they felt able to derive a score for every patient and were confident in their grading for the majority. This may explain why previous trialists have been happy to use abstracted mRS without first testing validity or reliability.

Our results should be generalizable to other centers involved in stroke trials. Patients were reviewed at approximately 3 months after the event, the period when mRS is traditionally assessed. Review was performed by practicing stroke physicians trained in the use of the mRS. Specific proformas are not used for data capture and consulting doctors can document as much information as they wish. Our clinic staff comprises internal medicine physicians with a stroke interest. It is possible that in the context of a specific rehabilitation, or privately funded service, consultations may be longer with more emphasis on disability.

The poor reliability inherent in standard mRS assessment is well recognized.[7] To ensure that our "correct" mRS grading was suitably robust, we used multiple independent raters with final mRS chosen by the majority. To put our results into context, although reliability of derived mRS was poor ($k=0.34$), traditional face-to-face interview mRS has also demonstrated poor reliability ($k=0.25$).[8] This suggests that improved reliability of mRS assessment is needed; however, any novel methodologies should have clinometric properties studied before clinical trial use.

Accurate mRS cannot be derived from standard hospital records. Even among cases in which the appraiser was "certain," the proportion of correct grades was no better than chance. Deriving mRS should not be encouraged; a directed interview remains essential.

## Acknowledgments

## Source of Funding

## Disclosures

## References

1. Sulter G, Steen C, Keyser JD. Use of the Barthel Index and modified Rankin Scale in acute stroke trials. *Stroke*. 1999;30:1538–1541.
2. Frey JL, Jahnke HK, Goslar PW, Partovi S, Flaster MS. tPA by telephone: extending the benefits of a comprehensive stroke centre. *Neurology*. 2005; 64:154–156.
3. Hylek EM, Go AS, Chang Y, Jensvold NG, Henault LE, Selby JV, Singer DE. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. *N Engl J Med*. 2003;349:1019–1026.
4. Kasner SE, Chalela JA, Luciano JM, Cucchiara BL, Raps EC, McGarvey ML, Conroy MB, Localio AR. Reliability and validity of estimating the NIH Stroke Scale score from medical records. *Stroke*. 1999;30: 1534–1537.
5. Stavem K, Lossius M, Ronning OM. Reliability and validity of the Canadian Neurological Scale in retrospective assessment of initial stroke severity. *Cerebrovasc Dis*. 2003;16:286–291.
6. Barber M, Fail M, Shield M, Stott DJ, Langhorne P. Validity and reliability of estimating the Scandinavian Stroke Scale score from medical records. *Cerebrovasc Dis*. 2004;17:224–227.
7. Quinn TJ, Lees KR, Hardemark HG, Dawson J, Walters MR. Initial experience of a digital training resource for modified Rankin Scale assessment in clinical trials. *Stroke*. 2007;38:2257–2261.
8. Wilson JT, Hareendran A, Grant M, Baird T, Schulz UG, Muir KW, Bone I. 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.