



University
of Glasgow

Govan, L. and Langhorne, P. and Weir, C.J. (2009) *Categorizing stroke prognosis using different stroke scales*. *Stroke*, 40 (10). pp. 3396-3399.
ISSN 0039-2499

<http://eprints.gla.ac.uk/8056/>

Deposited on: 5 November 2009

Stroke

American Stroke
AssociationSM

A Division of American
Heart Association



JOURNAL OF THE AMERICAN HEART ASSOCIATION

Categorizing Stroke Prognosis Using Different Stroke Scales

Lindsay Govan, Peter Langhorne and Christopher J. Weir

Stroke 2009;40;3396-3399; originally published online Aug 6, 2009;

DOI: 10.1161/STROKEAHA.109.557645

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214
Copyright © 2009 American Heart Association. All rights reserved. Print ISSN: 0039-2499. Online
ISSN: 1524-4628

The online version of this article, along with updated information and services, is
located on the World Wide Web at:

<http://stroke.ahajournals.org/cgi/content/full/40/10/3396>

Subscriptions: Information about subscribing to *Stroke* is online at
<http://stroke.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters
Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax:
410-528-8550. E-mail:
journalpermissions@lww.com

Reprints: Information about reprints can be found online at
<http://www.lww.com/reprints>

Categorizing Stroke Prognosis Using Different Stroke Scales

Lindsay Govan, PhD; Peter Langhorne, PhD; Christopher J. Weir, PhD

Background and Purpose—Stroke severity and dependency are often categorized to allow stratification for randomization or analysis. However, there is uncertainty whether the categorizations used for different stroke scales are equivalent. We investigated the amount of information retained by categorizing severity and dependency, and whether the currently used cut-offs are equivalent across different stroke scales.

Methods—Stroke severity and dependency have been categorized as mild, moderate, or severe. We studied 2 acute stroke unit cohorts, measuring Scandinavian Stroke Scale (SSS), modified Rankin Scale (mRS), Barthel Index (BI), and modified National Institutes of Health Stroke Scale (mNIHSS). Receiver operating characteristic (ROC) curves were examined to determine the ability of full and categorized scales to predict death and dependency. A weighted kappa analysis assessed agreement between the categorized scales.

Results—When scales are categorized, the area under the ROC curve is significantly reduced; however, the differences are small and may not be practically important. BI, mRS, and SSS all have excellent agreement with each other when categorized, whereas mNIHSS has substantial agreement with mRS and BI.

Conclusions—Little predictive information is lost when stroke scales are categorized. There is substantial to almost perfect agreement among categorized scales. Therefore the use and categorization of a variety of stroke severity or dependency scales is acceptable in analyses. (*Stroke*. 2009;40:3396-3399.)

Key Words: stroke severity ■ scales ■ stroke assessment ■ prognosis

Stroke is a heterogeneous condition which can resolve within a few hours or be rapidly fatal. When measuring severity, further variation may be introduced by the use of different assessment scales.

Stroke severity or dependency categories are often used to stratify randomization or analysis. However, there is uncertainty whether the categorizations used for different stroke scales are equivalent. Although discrepancies between severity or dependency classifications in the different stroke scales would not influence estimates of treatment effect, they may bias the assessment of interactions between severity and treatment. We investigated the amount of information retained by categorizing severity or dependency, and whether the currently used cut-offs are equivalent across different stroke scales.

Methods

We studied 2 hospital-based cohorts of consecutive unselected acute stroke admissions. The first cohort of 733 patients, Barber,¹ measured baseline SSS, mRS, and the BI²; death and dependency (defined as equivalent to mRS >2) were measured at one month. In addition to baseline mRS and BI, the second cohort of 412 patients, Sellars,³ recorded the mNIHSS⁴; death and dependency were measured 3 months poststroke.

Table shows an example (from the Stroke Units Trialists' Collaboration⁵) of stroke scale categories for acute stroke severity. The cut-off values were chosen in a similar fashion to the minimum

probability value approach.⁶ Three categories were chosen because a larger number of strata may lead to some strata containing few patients and hence incomplete randomized blocks. Conversely, dichotomizing severity may discard too much information.

ROC curves were used to assess the usefulness of each stroke scale in predicting outcome with comparisons made using nonparametric methods.⁷ Agreement between the categorizations of different scales was assessed using a weighted kappa analysis,⁸ exact agreement being given a weight of 1 and disagreements in adjacent and disparate categories weighted as 0.5 and 0, respectively.⁹

Results

Of 733 patients in the first cohort, 665 (91%) had mRS, BI, and SSS recorded within 3 days of admission. In the second cohort (412 patients), 405 (98%) had mRS, BI and mNIHSS recorded by day 5. In both cohorts, death and dependency were recorded for all patients.

The distribution of severity or dependency categories was consistent across scales within cohorts and between cohorts: mild (54% to 57%), moderate (22% to 31%), and severe (15% to 21%).

Full Versus Categorized Scales

The ROC curves for both cohorts show moderate or high accuracy in predicting outcome for both the full and categorized scales. Figures 1 and 2 show little difference between the full and categorized scales for the outcome of death,

Received May 12, 2009; final revision received June 9, 2009; accepted June 29, 2009.

From the Robertson Centre for Biostatistics (L.G., C.J.W.) and the Academic Section of Geriatric Medicine (P.L.), University of Glasgow, UK.

Correspondence to Lindsay Govan, Section of Public Health and Health Policy, 1 Lilybank Gardens, University of Glasgow, Glasgow, G12 8RZ, UK. E-mail lindsay@stats.gla.ac.uk

© 2009 American Heart Association, Inc.

Stroke is available at <http://stroke.ahajournals.org>

DOI: 10.1161/STROKEAHA.109.557645

Table. Baseline Stroke Severity Categorization

Scale	Mild	Moderate	Severe
Barthel Index (20 point scale)*	10–20	3–9	0–2
Modified Rankin Scale*	0–3	4	5
Scandinavian Stroke Scale*	43–58	26–42	0–25
Modified NIH Stroke Scale†	0–5	6–14	15–31

*Recorded on day 3; †recorded on day 1–5.

whereas there is slightly lower predictive accuracy of the categorized scales for death or dependency. The predictive accuracy for the full scales is also lower for this outcome.

The areas under the ROC curve for mRS and BI in Figure 1 (1-month outcome) are larger than the corresponding areas in Figure 2 (3-month outcome).

Comparison of Stroke Scales

For the Barber¹ cohort (Figure 1), the majority of the significant differences in areas lie within the death or dependency outcome. For both the full and categorized scales, mRS has a significantly larger area under the curve than SSS and BI, whereas BI has a larger area than SSS, suggesting mRS has the best predictive accuracy at 1 month after stroke.

However, any differences were small with narrow confidence intervals. The differences in predictive accuracy between stroke scales are less apparent for 3-month outcome (Figure 2).

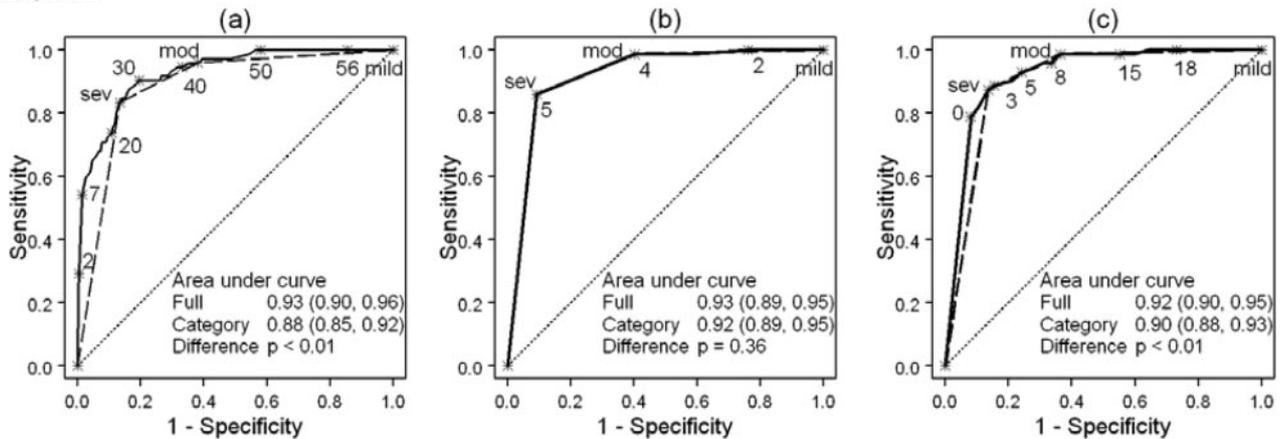
Subjective guidelines for interpreting weighted kappa analysis¹⁰ indicate BI, mRS, and SSS all have excellent agreement with each other when categorized (BI/mRS weighted kappa 0.85, percentage agree 88%; BI/SSS 0.80, 84%; mRS/SSS 0.77, 82%), whereas mNIHSS has substantial agreement with mRS and BI (mNIHSS/mRS 0.66, 73%; mNIHSS/BI 0.66, 74%).

Weighted kappa results across Oxfordshire Community Stroke Project (OCSP) classification categories were broadly consistent with the weighted kappa for the entire dataset: BI/mRS weighted kappa 0.77 to 0.80, percentage agree 73% to 88%; BI/SSS 0.65 to 0.80, 78% to 90%; mRS/SSS 0.62 to 0.84, 76% to 93%; mNIHSS/mRS 0.39 to 0.60, 0.72 to 0.76%; mNIHSS/BI 0.21 to 0.63, 68% to 78%.

Discussion

The SSS, mRS, BI and mNIHSS all have moderate to high predictive accuracy. When the scales are categorized the reduction in area under the ROC curve, although statistically significant, is small and may be unimportant practically: little predictive information is lost.

Death



Death or dependency

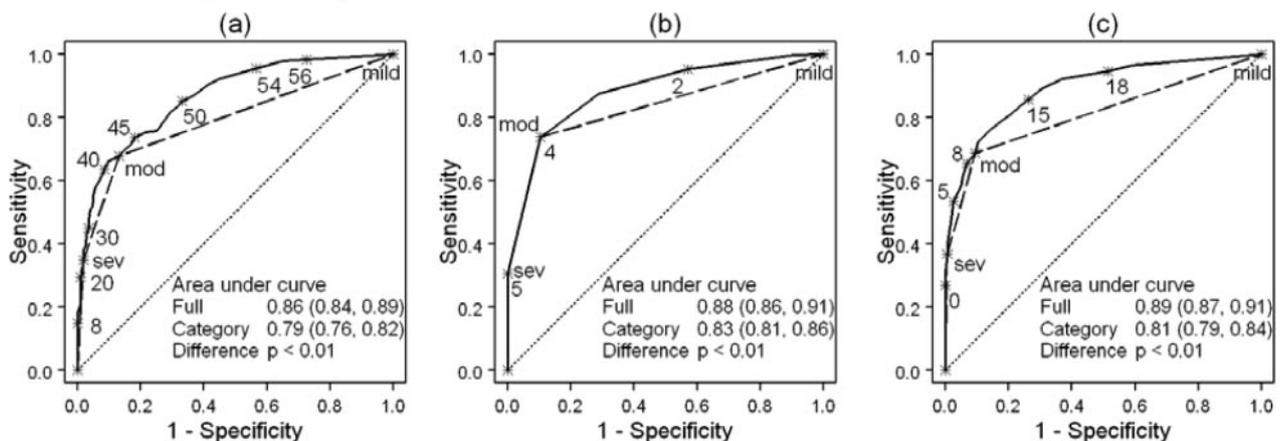
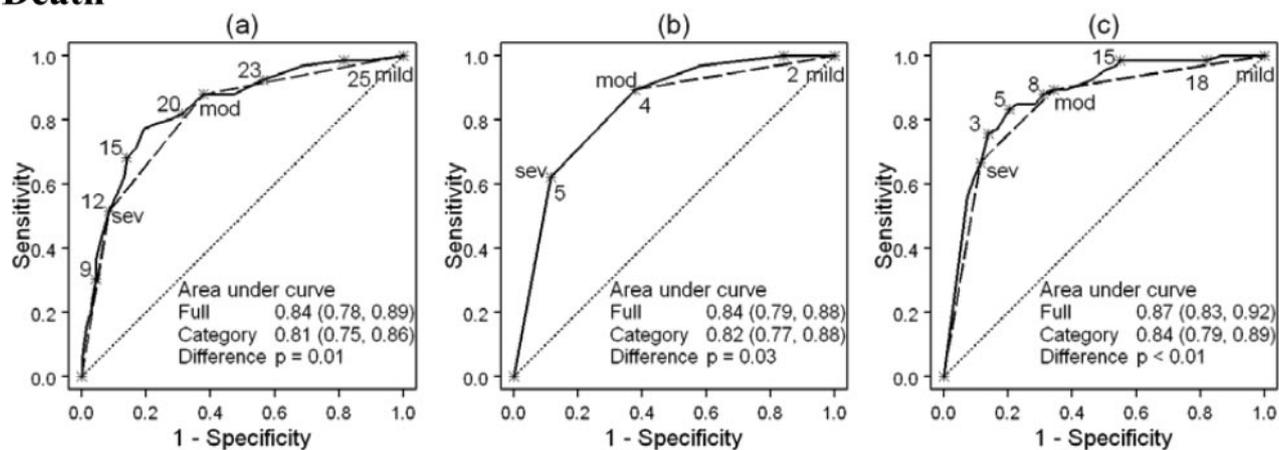


Figure 1. ROC curves of (a) SSS, (b) mRS, and (c) BI in Barber¹ cohort for the outcomes of death and death or dependency. The solid and dashed lines represent the full and categorized scales, respectively. Sensitivity and specificity are calculated at each level: less or equal to the level for SSS and BI and greater or equal to the level for mRS. The dotted line represents “no information.”

Death



Death or dependency

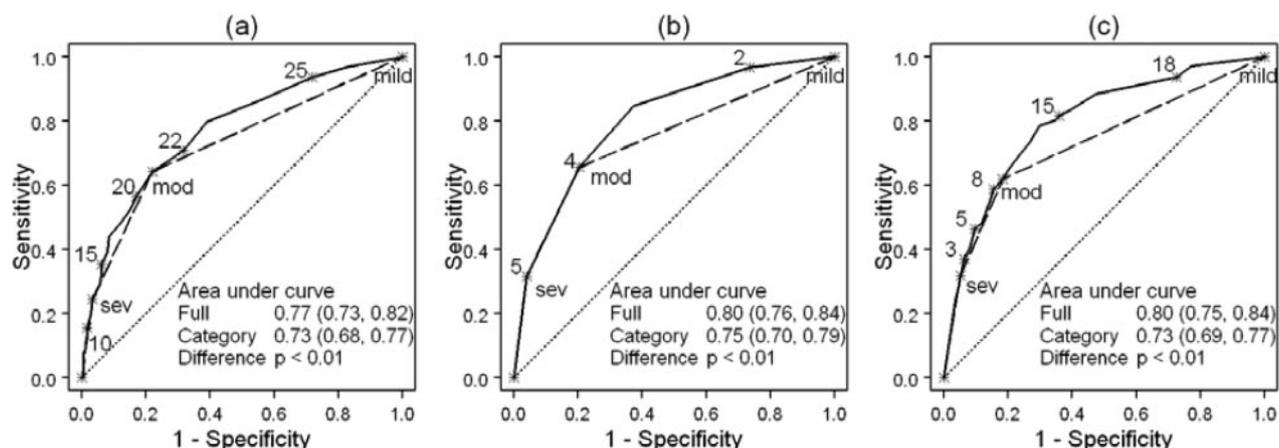


Figure 2. ROC curves of (a) mNIHSS, (b) mRS, and (c) BI in Sellars³ cohort for the outcomes of death and death or dependency as in Figure 1.

When comparing the prognostic accuracy of stroke scales at 1-month follow-up, the mRS predicted death or dependency better than the SSS or the BI. Dependency was measured using the mRS, therefore baseline mRS would be expected to be a better predictor. Nevertheless, the differences between scales either when full or categorized were small and of little practical importance. Weighted kappa analysis of the categorized scales showed there was substantial to almost perfect agreement among scales, and the results are broadly consistent within OCSF classifications. The mNIHSS is poor at detecting the symptoms and signs of posterior circulation syndrome. However, with the exception of this category the levels of mNIHSS agreement were consistently substantially higher (mNIHSS/mRS 0.51 to 0.60; mNIHSS/BI 0.48 to 0.63).

The 2 cohorts contained several hundred unselected patients with a broad range of case mix. There were few missing data, limiting the possibility of bias. Each cohort had several severity measurements allowing us to test 4 commonly used stroke scales for equivalence between full and categorized versions. The different follow-up times for the cohorts showed that the results are not restricted to one duration of follow-up. However, it would be of interest to examine whether these results could be replicated in other cohort studies.

This study indicates that categorization of a stroke scale does not substantially reduce its predictive ability. Scales stratified in this way are broadly equivalent. Although prognostic accuracy for longer follow-up is lower, it is not further reduced by categorization of severity. These findings emphasize that stratifying randomization in acute stroke clinical trials by severity can be a pragmatic approach, retaining much of the prognostic information contained in the corresponding full assessment scale.

Sources of Funding

L.G. is supported by a Chest Heart and Stroke Scotland research studentship.

Disclosures

None.

References

- Barber M, Wright F, Stott DJ, Langhorne P. Predictors of early neurological deterioration after ischaemic stroke: a case-control study. *Gerontology*. 2004; 50:102-109.
- De Hann R, Horn J, Limburg M, Van Der Meulen J, Bossuyt P. A comparison of five stroke scales with measures of disability, handicap and quality of life. *Stroke*. 1993;24:1178-1181.

3. Sellars C, Bowie L, Bagg J, Sweeney MP, Miller H, Tilston J, Langhorne P, Stott DJ. Risk factors for chest infection in acute stroke: a prospective cohort study. *Stroke*. 2007;38:2284–2291.
4. 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;32:1310–1316.
5. Stroke Unit Trialists' Collaboration. Organised inpatient (stroke unit) care for stroke. *Cochrane Database of Systematic Reviews*. 2007.
6. Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of using "optimal" cut points in the evaluation of prognostic factors. *J Natl Cancer Inst*. 1994;86:829–835.
7. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837–845.
8. Cohen J. Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*. 1968;70:213–220.
9. Cicchetti DV, Allison T. A new procedure for assessing reliability of scoring EEG sleep recordings. *Am J EEG Technol*. 1971;11:101–109.
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159–174.