

Predicting the outcome of acute stroke: prospective evaluation of five multivariate models and comparison with simple methods

J R F Gladman, D M J Harwood, D H Barer

Abstract

Five multivariate models designed to predict the outcome of stroke were tested prospectively on 102 consecutive stroke patients admitted to a district general hospital. The results were compared with predictions made using two simple clinical variables (the conscious level on admission and the state of urinary continence at four weeks). Of the three models (developed in Belfast, Guy's Hospital and Uppsala) intended for use in the acute stages of stroke the last two were slightly more accurate in their prediction of death (75%) than was the admission conscious level alone (65%), whereas the Belfast model had an accuracy of only 50% in this situation. At a later stage, the state of urinary continence predicted good and poor outcomes with similar accuracy to that of a multivariate model from Edinburgh. A model developed in Bristol performed poorly. When tested prospectively, these multivariate models proved considerably less accurate than when they were first described. Only the Uppsala model showed any advantage over simple clinical methods. This might be of value in defining prognostic strata for clinical studies, but not in the management of individual patients. Simple clinical variables thus offer as much to clinicians as complex multivariate models.

Accurate prediction of the outcome of acute stroke is important for several reasons. A reliable prognosis must be given to relatives and realistic rehabilitation goals must be set. Specialised resources should be allocated in the most efficient way, and stroke patients should be stratified into different prognostic groups for clinical trials.

Many clinical signs and other variables have been found to have prognostic importance, but since most of them are markers of overall stroke severity they tend to be strongly inter-related. Multivariate analysis has therefore been advocated to identify and bring together those factors that make an independent contribution to the prediction of outcome. Several models have been derived¹⁻⁵ which combine variables into a discriminant function or multiple regression formula, but, as previously noted,⁶ the more complex these models become the less likely they are to be widely applicable. One such model,¹ for instance, makes use of a version of a mental test score⁷

which itself has been modified by means of multivariate analysis to maximise its discriminant power. As a result, some of the questions are negatively weighted so that patients who answer them correctly have a poorer predicted outcome than those who do not. These weightings may help the model to fit more closely the data from which it was derived, but they are unlikely to increase the true prognostic value of the overall score.

Multivariate models are also difficult to use in practice. They may require several minutes work with a pocket calculator, incorporating a range of disparate information, some of which may be based on rating scales of dubious reliability. To outweigh these disadvantages these models must therefore be shown to produce substantial improvements in predictive accuracy over simpler univariate methods.

We therefore chose to test prospectively the three prognostic models that had been developed for use in acute stroke in Belfast,¹ Guy's Hospital² and Uppsala³ and the two that were designed in Bristol⁴ and Edinburgh⁵ for use at a later stage. We evaluated the models prospectively in a large group of unselected patients with stroke admitted to a district general hospital and compared their predictions with those based on single clinical variables. As previously suggested,⁶ we used the admission conscious level (ACL) to make predictions in the first few days and the state of urinary continence to predict recovery at a later stage.

Patients and methods

All patients with acute stroke admitted to the medical and geriatric wards of the City Hospital, Nottingham, over a three month period were examined by a single observer (DH). Clinical data were collected on admission, at 7-10 days, at four weeks, and at three months. Deaths and discharges during the follow up period were recorded and the Barthel activities of daily living index^{8,9} and a modified Rankin handicap score^{1,10} were determined in survivors at three months. Information on discharged patients was obtained by telephone or by post.

Prognostic variables and models

Table 1 shows the clinical variables used in the models from Belfast,¹ Guy's,² Uppsala,³ Bristol,⁴ and Edinburgh⁵ and the times when the assessments should be performed. Different variables and different scoring systems are

Department of Health Care of the Elderly, University Hospital, Nottingham
J R F Gladman

Department of Medicine, Ipswich Hospital
D M J Harwood

Department of Geriatric Medicine, Royal Liverpool Hospital
D H Barer

Correspondence to: Dr J R F Gladman, Department of Health Care of the Elderly, University Hospital, Nottingham NG7 2UH

Received 17 April 1991 and in revised form 25 July 1991.

Accepted 1 August 1991

Table 1 Variables used in predictive formulas

Model	Variables used
On admission:	
Belfast model	Level of consciousness Weighted mental test score Visuospatial neglect Arm power Mobility
Guy's model	Non-specific ECG changes (ST segment changes not definitely indicative of ischaemia) Age Loss of consciousness at onset of stroke Conscious level after 24 hours Higher cerebral dysfunction, hemianopia, and hemiplegia Complete paralysis of the worst affected limb
Uppsala model	Uncomplicated hemiparesis Level of consciousness Orientation Dysphasia Conjugate gaze deviation Facial weakness Motor strength of each limb Performance Disability Status Scale Reflexes Sensation
At 1 week:	
Bristol model	Age Hemianopia Arm motor deficit Sitting balance Urinary incontinence
At 4 weeks:	
Edinburgh model	Arm motor function Postural function Proprioception

used in each model. (Details of the scoring systems can be found in the original papers). For example, in the Belfast model, two five-point scales (range 1–5) for leg function (mobility) and arm power are used; in the Guy's model, complete paralysis of the worst affected limb is given a score of –12; in the Uppsala model, the strength in each limb is measured on a three-point scale (0,2,5) and the scores are summed; in the Bristol model, a four-point scale (1–4) measuring arm function is used; and in the Edinburgh model, arm function is assessed using a five-point scale (0–4) and posture/mobility is assessed on a four-point scale (0–3).

In general the worst possible score was assumed if a patient could not be assessed or was unable to perform a particular test. The exception was Albert's test of visuospatial neglect,¹¹ used in the Belfast model, where patients unable to attempt the test were assigned the average score of the group, as recommended in the original paper.¹

As well as using different rating scales, the models combine the information and categorise the predicted outcomes in different ways. The Guy's, Uppsala, and Edinburgh models use formulas based on admission data to produce a numerical score that separates the patients into broad prognostic categories. The Uppsala score defines a group with a high risk of death in hospital. The Guy's model categorises patients on admission into groups with a good, intermediate, or poor chance of recovery at two or six months, and the Edinburgh model, applied at 12 weeks, makes similar predictions of recovery at six months.

The Belfast model produces no overall score but combines the prognostic variables on admission in four separate equations which estimate the probabilities of death, dependency, recovery to independence, and full recovery,

which are the four categories of a modified Rankin handicap scale.^{1–10} The highest probability indicates the category of most likely outcome.

The Bristol model is designed to predict the Barthel activities of daily living score^{8,9} at six months from the clinical findings at about one week. The equation can occasionally produce scores above the maximum for the scale, and we rounded down such scores to 20/20.

The single variable predictors we used were conscious level on admission (where "alert" was used to predict survival and "drowsy or unconscious" to predict death) and urinary continence (where "continent" was used to predict good outcome and "occasionally or persistently incontinent" to predict poor outcome). Both of these variables had been identified in previous studies,^{6,12} based on separate datasets, as being of particular prognostic value.

Analysis

As a first step, rank correlations were computed between (a) each of the model scores and the modified Rankin grade at three months; and (b) the later predictive scores (Bristol and Edinburgh models) and the Barthel index at three months. The accuracy of the model predictions was then tested in more detail by contingency table analysis.

Prediction of death in the acute stages: (1) with original cut off scores. The ability of the Guy's, Uppsala, and Belfast models to predict death correctly at three months was tested and compared with a prediction of death made by the finding of an impaired conscious level on admission. Cut off scores were used as in the original models to indicate a "poor chance of recovery" (Guy's) and a "high likelihood of death" (Uppsala).

Prediction of death in the acute stages: (2) with

adjusted cut off scores. The various models were developed in different patient populations with different overall mortality and therefore are based on different "prior probabilities." One way to correct for this inequality is to adjust the cut off scores so that the models all have the same sensitivity for predicting death, and this was done with the Guy's and Uppsala models. The Belfast model cannot be adjusted in this way as cut off scores are not used.

Prediction of good outcome at later stages. The ability of the Bristol and Edinburgh models to predict correctly a good outcome (defined as a Barthel score at three months of > 10/20) was compared to a prediction based upon the degree of bladder control at 4 weeks. The "good outcome" group defined by the Edinburgh model was taken to indicate a Barthel score of > 10/20, whereas the Bristol model was designed to predict the Barthel score directly.

Prediction of poor outcome at later stages. The ability of the Edinburgh model to predict correctly a poor outcome (defined as death or prolonged hospital stay) was compared to the likelihood of a poor outcome if a patient was incontinent of urine at four weeks. The Bristol model cannot be compared in this way as it was derived from a population of survivors.

Patients who died after the index assess-

ments were included in the analyses of outcome at later stages and were given a Barthel score of 0/20 for the analysis predicting good outcome.

Results

We studied 102 consecutive patients admitted with acute stroke over a three month period. The mean age was 75 (range 42-91) years, and 22 had suffered one or more strokes in the past.

By the end of the first week, 13 patients (13%) had died and 19 (19%) had been discharged, leaving 70 patients (68%) still in hospital. At four weeks 23 patients (23%) had died, 38 (37%) had been discharged, and 41 (40%) remained in hospital, and by three months the equivalent figures were 33 patients (32%), 55 (54%), and 14 (14%).

Albert's line cancellation test could not be attempted at the first assessment by 45 patients (44%); they were assigned the average test score, which was 28.8 for our group, compared with 28.5 in the original Belfast study.¹ The Edinburgh score could not be calculated in two patients and we were unable to perform a Barthel activities of daily living assessment at three months in one patient.

Table 2 shows the rank correlations between the predictive model scores, the conscious level and urinary continence, and the various measures of recovery. As expected the correlations were all highly significant (p < 0.001).

Table 3 compares the accuracy of the Guy's, Uppsala, and Belfast models and of the admission conscious level (ACL) alone in predicting death within three months. The Belfast model was the most sensitive and least specific in predicting fatal outcome. The explanation can be seen in the right hand column, which shows that this model predicted an overall fatality rate of 78%, more than twice the actual rate. Conversely, the Uppsala model predicted a very low overall death rate and was correspondingly highly specific but insensitive. The overall accuracy and prognostic contribution

Table 2 Rank correlation between predictive model scores and measures of outcome at three months

	Spearman's Rho	
	Modified Rankin grade	Barthel index
Prediction on admission:		
Conscious level	0.48	
Belfast score	0.38	
Guy's score	0.52	
Uppsala score	0.58	
Prediction at later stages:		
Continence (at 4 weeks)	0.52	0.71
Bristol score (at 1 week)	0.60	0.64
Edinburgh score (at 4 weeks)	0.59	0.65

p < 0.001 For all coefficients.

Table 3 Accuracy of models on day 1 in predicting death within three months

Predictive model	Sensitivity (%)	Specificity (%)	Likelihood ratio (95% CI)	Positive predictive value (%)	Accuracy (%)	Predicted mortality (%)
Impaired ACL	76	59	1.9 (1.3 to 2.6)	47	65	52
Belfast model	94	29	1.3 (1.1 to 1.6)	39	50	78
Guy's model	58	83	3.3 (1.8 to 6.0)	61	75	30
Uppsala model	30	96	7.0 (2.1 to 24)	77	75	13

Actual death rate 32%.

Predicted outcome	Actual outcome		
	+	-	
+	a	b	N = Total no of cases a = No of true positives b = No of false positives c = No of false negatives d = No of true negatives
-	c	d	

Sensitivity = $100 \times a / (a + c)$
 Specificity = $100 \times d / (b + d)$
 Positive predictive value = $100 \times a / (a + b)$
 Accuracy = $100 \times (a + d) / N$
 Likelihood ratio = Sensitivity / (100 - Specificity)
 Predicted mortality (or outcome rate) = $(a + b) / N$

(as represented by the likelihood ratio) of the Belfast model was poor.

Table 4 examines the predictions of the Guy's and Uppsala models when their cut off points were adjusted to produce the same sensitivity as achieved by the use of an impaired ACL alone. This greatly reduced the likelihood ratios of both models (which relate the "pre-test and post-test" odds of dying), though the Uppsala model remained a slightly better predictor than the other two.

In a further analysis the cut off scores were adjusted to make the predicted death rate equal to the actual mortality of 32%. Under these conditions the Guy's model gave a likelihood ratio of 3.2 (95% confidence interval 1.8 to 5.6) compared with 4.2 (2.3 to 7.6) for the Uppsala model. The Guy's model was derived in patients aged under 75, but restricting the analysis to this age group in our study did not significantly improve its predictive accuracy.

Table 5 compares the predictive value of urinary continence and of the Bristol score at one week, and of the Edinburgh score at four weeks for "good outcome" (a Barthel index of > 10/20). Continence was a more sensitive predictor than the Edinburgh model, which indicated that only 13% of patients would have a good outcome, a third of the true figure. Moreover, the ability of the Edinburgh score to predict a bad outcome (death or prolonged hospital stay) was little better than that of urinary incontinence alone.

Discussion

If they are to be widely applied predictive models should be derived from an unselected

and representative series of patients. The patients in our study were admitted consecutively to a large district general hospital and should therefore be representative of the generality of patients admitted to hospital with acute stroke, the target population for most studies of outcome prediction. Other studies have been more selective: the Belfast workers excluded patients with previous mental and physical disability;¹ the Guy's study excluded all those over 75 years of age;² the Uppsala model was designed for patients aged under 70 deemed to have had a first cerebral infarction;³ the Edinburgh score was derived from a selected "middle band" of patients participating in an evaluation of a stroke unit,⁴ and the Bristol score could be derived only retrospectively, after deaths had been excluded.⁵

The choice of the study population used when producing a multivariate model will also affect the factors selected by the modelling process and their relative weightings. For example, the Bristol score can predict a Barthel score of less than 10/20 at six months only if the patient has maximum impairment and is aged over 100 years. This model was derived from a selected group of patients who had a generally favourable outcome, and this largely explains its poor performance in our more heterogeneous study population.

Our findings have confirmed previous suggestions⁶ that complex models are likely to be much less accurate when used prospectively to make real predictions than when used retrospectively to fit previously collected data. The Belfast model was 83% accurate in "predicting" death or survival in the original report,¹ but its predictions were only 50% accurate in our series. Similarly, Wade *et al* claimed to

Table 4 Accuracy of models in predicting death within three months when adjusted to the same sensitivity as impaired conscious level

Predictive model	Sensitivity (%)	Specificity (%)	Likelihood ratio (95% CI)	Positive predictive value (%)	Accuracy (%)	Predicted mortality (%)
Impaired ACL	76	59	1.9 (1.3 to 2.6)	47	65	52
Guy's model	76	62	2.0 (1.4 to 2.9)	49	67	50
Uppsala model	76	70	2.5 (1.7 to 3.7)	54	72	45

Actual death rate 32%.

Table 5 Comparison of ability of later predictive models and state of urinary continence to predict patients with good outcome (Barthel index > 10/20) or a poor outcome (death or prolonged hospital stay)

Predictive model	Sensitivity (%)	Specificity (%)	Likelihood ratio*	Positive predictive value (%)	Accuracy (%)	Predicted outcome rate (%)
<i>Barthel index > 10/20 at 3 months</i>						
Urinary continence at 4 weeks	92	88	7.9	79	89	37
Bristol score at 1 week	100	0	1.0	38	38	100
Edinburgh score at 4 weeks	33	96	8.7	80	76	13
<i>Death or prolonged hospital stay (> 3 months)</i>						
Urinary incontinence at 4 weeks	68	47	1.3	63	59	62
Edinburgh score at 4 weeks	55	65	1.5	67	59	46

38% of those alive at 1 week had Barthel score > 10/20 at 3 months, 32% of those alive at 4 weeks had Barthel score > 10/20 at 3 months, 56% of those alive at 4 weeks were dead or still in hospital at 3 months.

*Confidence intervals unreliable.

“predict” the Barthel score at six months to within two points in 72% of cases,⁴ whereas we found the Bristol model could correctly predict a Barthel score above or below 10 in only 38% of cases. Admittedly we measured outcome at three months, as opposed to six months in the Belfast and Bristol studies, but this by no means explains the large discrepancy in the results.

The Belfast model was inordinately pessimistic in our survey, predicting death in 78% of cases whereas the actual death rate was only 32%. Reducing the model's high sensitivity for fatal outcome might have improved its specificity and discriminatory power, but its complex design makes this difficult to do.

The Guy's and Uppsala models were less sensitive but more specific in their prediction of death than the conscious level alone and seemed to have greater overall accuracy. When the cutoff scores were adjusted to equalise the sensitivity, however, it became clear how little additional prognostic information these models contained. The Guy's score performed no better than ACL alone, whereas the Uppsala model produced only a modest increase in likelihood ratio. This might be of some use in stratifying patients for research studies, but the models' predictions are probably not accurate enough to be used as a basis for management decisions on individual patients. There would also seem to be little advantage in using these complex models when giving an approximate prognosis to relatives.

For predicting functional outcome at a later stage, urinary continence proved as accurate on its own as the Edinburgh score. This confirms that urinary continence is a simple yet useful predictor of outcome,^{12,13} and shows the lack of additional prognostic information contained in the Edinburgh model.

This study has shown some of the limitations of multivariate linear models as predictive instruments in stroke. Even if their accuracy could be improved, they are unlikely to be able to cope with the diversity of patients found in everyday practice or be of much use in their management. Additional prognostic information may be obtained from CT scanning, though one study indicated that it did not add much to predictions based on clinical data alone.¹⁴ An alternative approach, in which a quasi-anatomical classification is derived from the clinical signs, has recently been proposed¹⁵ and initial results suggest that this may also be

an efficient method of prognostic stratification. Further work along the lines presented here will be required to validate the classification prospectively in terms both of clinical-CT comparisons and of outcome predictions. We believe that greater flexibility and accuracy could be achieved by an algorithm, which could use different prognostic variables in different situations and could incorporate information on the rate of change of neurological and functional status. Such an algorithm would be based on a Bayesian model which could be continuously updated by new data, rather than the “closed” multiple regression models examined here. It could also be designed to guide clinical decisions, rather than merely make predictions of outcome of little practical value. Much work is required to develop this approach, but in the meantime simple clinical variables offer as much to the clinician as complex multivariate models in the prediction of outcome in stroke.

Dr JRF Gladman was a Chest, Heart and Stroke Association Research Fellow, and Dr DJ Harwood was supported by a grant from Sandoz Pharmaceuticals. We would like to thank the late Professor JRA Mitchell and Professor SB Ebrahim for helpful comments on the preparation of this manuscript, and all the physicians and geriatricians at the City Hospital, Nottingham, for allowing us to study patients under their care.

- 1 Fullerton KJ, MacKenzie G, Stout RW. Prognostic indices in stroke. *QJ Med* 1988;66:147-62.
- 2 Allen CMC. Predicting the outcome of acute stroke: a prognostic score. *J Neurol Neurosurg Psychiatry* 1984;47:475-80.
- 3 Frithz G, Werner I. Studies on cerebrovascular strokes. II. Clinical findings and short-term prognosis in a stroke material. *Acta Med Scand* 1976;199:133-40.
- 4 Wade DT, Skilbeck CE, Langton-Hewer R. Predicting Barthel ADL score at 6 months after an acute stroke. *Arch Phys Med Rehabil* 1983;64:24-8.
- 5 Prescott RJ, Garraway WM, Akhtar AJ. Predicting functional outcome following acute stroke using a standard clinical examination. *Stroke* 1982;13:641-7.
- 6 Barer DH, Mitchell JRA. Predicting the outcome of acute stroke: do multivariate models help? *QJ Med* 1989;70:27-39.
- 7 Hodkinson HM. Evaluation of a mental test score for assessment of mental impairment in the elderly. *Age Ageing* 1972;1:233-8.
- 8 Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Maryland State Med J* 1965;14:61-5.
- 9 Collin C, Wade DT, Davies S, Horne V. The Barthel ADL index: a reliability study. *Int Disabil Stud* 1988;10:61-3.
- 10 Rankin J. Cerebral vascular accidents in patients over the age of 60: II. Prognosis. *Scot Med J* 1957;2:200-15.
- 11 Albert M. A simple test of visual neglect. *Neurology* 1973;23:658-64.
- 12 Barer DH. Continence after stroke: goal of therapy? *Age Ageing* 1989;18:183-91.
- 13 Wade DT, Langton Hower RL. Outlook after acute stroke: urinary incontinence and loss of consciousness compared in 532 patients. *QJ Med* 1985;56:601-8.
- 14 Allen CMC. Predicting outcome after stroke: role of computerised tomography. *Lancet* 1984;ii:464-5.
- 15 Bamford JM. The classification and natural history of acute cerebrovascular disease [dissertation]. Manchester: University of Manchester, 1986.