**Original Paper**

# Developing and Validating a Predictive Model for Stroke Progression

L.E. Craig    O. Wu    H. Gilmour    M. Barber    P. Langhorne

College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

**Abstract**

**Background:** Progression is believed to be a common and important complication in acute stroke, and has been associated with increased mortality and morbidity. Reliable identification of predictors of early neurological deterioration could potentially benefit routine clinical care. The aim of this study was to identify predictors of early stroke progression using two independent patient cohorts. **Methods:** Two patient cohorts were used for this study – the first cohort formed the training data set, which included consecutive patients admitted to an urban teaching hospital between 2000 and 2002, and the second cohort formed the test data set, which included patients admitted to the same hospital between 2003 and 2004. A standard definition of stroke progression was used. The first cohort (n = 863) was used to develop the model. Variables that were statistically significant (p < 0.1) on univariate analysis were included in the multivariate model. Logistic regression was the technique employed using backward stepwise regression to drop the least significant variables (p > 0.1) in turn. The second cohort (n = 216) was used to test the performance of the model. The performance of the predictive model was assessed in terms of both calibration and discrimination. Multiple imputation methods were used for dealing with the missing values. **Results:** Variables shown to be significant predictors of stroke progression were conscious level, history of coronary heart disease, presence of hyperosmolarity, CT lesion, living alone on admission, Oxfordshire Community Stroke Project classification, presence of pyrexia and smoking status. The model appears to have reasonable discriminative properties [the median receiver-operating characteristic curve value was 0.72 (range 0.72–0.73)] and to fit well with the observed data, which is indicated by the high goodness-of-

Louise Craig

College of Medical, Veterinary and Life Sciences
University of Glasgow
1 Lilybank Gardens, Glasgow G12 8RZ (UK)
Tel. +44 141 330 7172, E-Mail louise.craig@glasgow.ac.uk

fit p value [the median p value from the Hosmer-Lemeshow test was 0.90 (range 0.50–0.92)]. **Conclusion:** The predictive model developed in this study contains variables that can be easily collected in practice therefore increasing its usability in clinical practice. Using this analysis approach, the discrimination and calibration of the predictive model appear sufficiently high to provide accurate predictions. This study also offers some discussion around the validation of predictive models for wider use in clinical practice.

Copyright © 2011 S. Karger AG, Basel

## Introduction

Early neurological deterioration within the first few hours or days of stroke onset has been referred to as stroke progression. It is a common event that has been reported to occur in 20–40% of acute strokes, and is associated with increased mortality and morbidity [1]. Stroke progression can be defined as a patient admitted to hospital with acute stroke whose neurological condition deteriorates over 48–72 h [2]. Stroke progression is often characterised by a fall of ≥2 points on a neurological scale measuring stroke severity, such as the National Institute of Health Stroke Scale. An alternative, validated approach is to define progression as any significant deterioration in conscious level, speech, arm or leg power, or facial weakness [2]. It is accepted that in stroke progression, neurological worsening takes place slowly with an amplification of previous deficits or appearance of new symptoms corresponding to the same vascular territory. Progression has been shown to be a common and important complication in acute stroke, which has also been linked to poor long-term outcome.

Reliable identification of predictors of early neurological deterioration could potentially be of benefit for routine clinical care. Therefore, it is important to understand the mechanism by which progression in stroke occurs and identify the factors that may be able to predict such an occurrence. Yet, the causes of stroke progression are largely unknown and the predictors identified in previous studies vary [3, 4]. A previous case-control study using 873 consecutive acute stroke admissions identified a history of diabetes and raised systolic blood pressure as predictors of stroke progression [4]. However, another prospective study of 868 patients with acute stroke found that the risk of early progression decreased as systolic blood pressure increased [5]. This same study also showed that diabetes was a risk factor for early progression and initial stroke severity for late progression. A smaller study of 152 consecutive patients with first-ever ischaemic strokes showed that high levels of glucose on admission and brain swelling on the first computed tomography (CT) were the predictors of progression [6]. A more recent study of 196 patients with ischaemic stroke identified blood urea nitrogen or creatinine as a predictor of stroke progression [7]. Another study investigating 133 lacunar stroke patients suggested that hypertriglyceridaemia may be a possible predictor of stroke progression [8]. The aim of this study was to identify reliable predictors of early neurological deterioration in stroke.

## Patients and Methods

Consecutive patients admitted to an urban teaching hospital with a diagnosis of stroke between 2000 and 2002, and subsequently between 2003 and 2004, were registered to a research database. Baseline data on stroke characteristics, stroke severity and early neurological deterioration were collected prospectively during both periods. This created two patient cohorts for this study – the first cohort formed the training data set and included 1,029 consecutive patients admitted to hospital, and the second cohort formed the test data set and

EXTRA
Cerebrovascular Diseases

Cerebrovasc Dis Extra 2011;1:105–114

DOI: 10.1159/000334473
Published online: December 3, 2011

© 2011 S. Karger AG, Basel
www.karger.com/cee

107

Craig et al.: Predictive Model for Stroke Progression

included 219 patients admitted to the same hospital. Only patients with complete outcome data were assessed in the analysis – 863 patients in the training data set and 216 patients in the test data set. With respect to acute medical management, patients received aspirin after CT, hypertension was not actively managed and premorbid antihypertensives were usually withheld for at least 72 h. No patients in the study group received thrombolysis.

Stroke progression was defined as a $\geq 2$-point worsening in the Scandinavian Stroke Scale (SSS) in either conscious level, arm, leg or eye movement scores, and/or a $\geq 3$-point SSS worsening in speech score within 72 h of hospital admission [9]. Outcome assessment was conducted 3 days after stroke.

The training data set was used to develop the predictive model. Clinical opinion and the existing literature were used to reduce the number of variables collected at baseline to those considered potentially predictive of stroke progression. Univariate analysis was then conducted between these potentially predictive variables and the outcome variable to determine which should be included in the model [10, 11]. The variables that were statistically significant ($p < 0.1$) on univariate analysis were included in the multivariate model. Logistic regression was the technique employed using backward stepwise regression to drop the least significant variables ($p > 0.1$) in turn.

The performance of the model was investigated using the test data set and assessed in terms of both calibration and discrimination [12]. It is the area under a receiver-operating characteristic (ROC) curve which provides an assessment of how good the model is in discriminating between individuals with and without the outcome [10]. The calibration of the model, which is comparing the observed proportion of events against the predicted probabilities, was tested using the Hosmer-Lemeshow test.

Multiple imputation methods were used for dealing with the missing values in both data sets. It was assumed that the missing data on the potentially prognostic variables were missing at random. This assumes that the probability that a value is missing is dependent on the values of variables that were actually collected and not independent of patient characteristics. Multiple imputation allows for the uncertainty about the missing data by creating 10 different plausible imputed data sets and appropriately combining results obtained from each of them [13].

Two predictive models were developed and validated using observed and imputed data sets. All analyses were carried out in Stata 11.

## Results

Stroke progression was recorded in 197 (23%) and 40 patients (18.5%) in the training and test data sets, respectively. Clinical opinion identified 27 variables as potentially predictive of stroke progression (table 1). Overall, the characteristics of the patients in the two cohorts were similar (table 1). The majority of patients were previously independent, with approximately 50% of the patients with hypertension and one third of patients having had a previous stroke. Co-morbidities such as atrial fibrillation, cardiac failure and diabetes were recorded in a minority of the patients. Patients who had a haemorrhage stroke were not recruited to the test cohort of patients.

Data on age, gender, conscious level, living alone, Oxfordshire Community Stroke Project (OCSP) classification, previous independence, type of CT lesion and the side of the lesion were complete for all patients. Of the 27 prognostic variables, 18 variables had incomplete data. Data on arm power, atrial fibrillation, abnormal physiological features, diabetes, history of coronary heart disease, hypoxia, hypertension, leg power, maximum systolic blood pressure, minimum systolic blood pressure, pyrexia, hyperglycaemia, previous stroke, previous transient ischaemic attack, smoking status, alcohol intake, osmolarity and verbal score were missing in

Craig et al.: Predictive Model for Stroke Progression

**Table 1.** Patient demographics and clinical variables

| | Training data set (n = 863) | Test data set (n = 216) |
|---|---|---|
| Age (mean ± SD), years | 69.0 ± 13.3 | 69.9 ± 13.5 |
| Missing | – | |
| Systolic blood pressure (mean ± SD), mm Hg | | |
| Min. | 128.2 ± 23.4 | not available |
| Missing | 12 (1.4%) | – |
| Max.[1] | 174 ± 32.5 | 152.1 ± 29.4 |
| Missing | 14 (1.6%) | – |
| Time from onset (median, IQR), h | 2 (1–8) | 6 (2–12)[2] |
| Gender | | |
| Male | 434 (50.3%) | 97 (45.0%) |
| Female | 429 (49.7%) | 119 (55.1%) |
| Missing | – | – |
| Lives alone | | |
| Yes | 284 (32.9%) | 73 (33.8%) |
| No | 579 (67.1%) | 143 (66.2%) |
| Missing | – | – |
| Previous independence | | |
| Yes | 695 (80.5%) | 178 (82.4%) |
| No | 168 (19.5%) | 38 (17.8%) |
| Missing | – | – |
| Verbal score | | |
| <6 SSS score | 154 (17.8%) | 62 (28.7%) |
| 6–12 SSS score | 709 (82.2%) | 154 (71.3%) |
| Arm power | | |
| 0–2 SSS score | 218 (25.3%) | 66 (30.6%) |
| 3–5 SSS score | 644 (74.7%) | 150 (69.4%) |
| Missing | 1 (0.1%) | – |
| Leg power | | |
| 0–2 SSS score | 309 (35.8%) | 56 (25.9%) |
| 3–5 SSS score | 554 (64.2%) | 160 (74.1%) |
| Conscious level | | |
| Reduced | 103 (11.9%) | 29 (13.4%) |
| Normal | 760 (88.1%) | 187 (86.6%) |
| Hypertension | | |
| Yes | 417 (48.3%) | 107 (49.5%) |
| No | 428 (49.6%) | 109 (50.5%) |
| Missing | 18 (2.1%) | – |
| Atrial fibrillation | | |
| Yes | 95 (11.0%) | 22 (10.2%) |
| No | 744 (86.2%) | 194 (89.8%) |
| Missing | 24 (2.8%) | – |
| Cardiac failure | | |
| Yes | 286 (33.1%) | 76 (35.2%) |
| No | 562 (65.1%) | 140 (64.8%) |
| Missing | 15 (1.7%) | – |
| Diabetes | | |
| Yes | 149 (17.3%) | 32 (14.8%) |
| No | 705 (81.7%) | 184 (85.2%) |
| Missing | 9 (1.0%) | – |

Cerebrovasc Dis Extra 2011;1:105–114

DOI: 10.1159/000334473
Published online: December 3, 2011

© 2011 S. Karger AG, Basel
www.karger.com/cee

Craig et al.: Predictive Model for Stroke Progression

109

**Table 1** (continued)

| | Training data set (n = 863) | Test data set (n = 216) |
|---|---|---|
| Previous stroke | | |
| Yes | 265 (30.7%) | 48 (22.2%) |
| No | 584 (67.7%) | 168 (77.8%) |
| Missing | 14 (1.6%) | – |
| Previous TIA | | |
| Yes | 122 (14.1%) | 29 (13.4%) |
| No | 702 (81.3%) | 186 (86.1%) |
| Missing | 39 (4.5%) | 1 (0.46%) |
| Smoking status | | |
| No | 241 (27.9%) | 71 (32.9%) |
| Current | 299 (34.7%) | 78 (36.1%) |
| Ex-smoker | 159 (18.4%) | 51 (23.6%) |
| Missing data | 164 (19.0%) | 16 (7.4%) |
| Alcohol use | | |
| None | 257 (29.8%) | 99 (45.8%) |
| 1–20 units | 210 (24.3%) | 44 (20.4%) |
| >20 units | 108 (12.5%) | 34 (15.7%) |
| Missing data | 288 (33.4%) | 39 (18.1%) |
| Side of lesion | | |
| Not classifiable | 51 (5.91%) | – |
| Left | 389 (45.1%) | 117 (54.2%) |
| Right | 394 (45.7%) | 99 (45.8%) |
| Posterior | 29 (3.4%) | – |
| Missing data | – | – |
| CT lesion | | |
| No scan | 104 (12.1%) | – |
| No appropriate lesion visible | 193 (22.4%) | 92 (42.6%) |
| Infarction | 479 (55.5%) | 124 (57.4%) |
| PICH | 65 (7.5%) | – |
| Other | 22 (2.6%) | – |
| Missing data | – | – |
| OCSP | | |
| Not classifiable | 69 (8.0%) | – |
| Total ACS | 203 (23.5%) | 63 (29.2%) |
| Partial ACS | 300 (34.8%) | 92 (42.6%) |
| LACS | 233 (27.0%) | 46 (21.3%) |
| POCS | 58 (6.7%) | 15 (6.9%) |
| Missing data | – | – |
| Pyrexia | | |
| Yes | 226 (26.2%) | 67 (31.0%) |
| No | 627 (72.7%) | 148 (68.5%) |
| Missing data | 10 (1.2%) | 1 (0.5%) |
| Hyperglycaemia | | |
| Yes | 134 (15.0%) | 41 (19.0%) |
| No | 634 (73.5%) | 167 (77.3%) |
| Missing data | 95 (11.0%) | 8 (3.7%) |
| Hyperosmolarity | | |
| Yes | 218 (25.3%) | 51 (23.6%) |
| No | 536 (62.1%) | 129 (59.7%) |
| Missing data | 109 (12.6%) | 36 (16.7%) |

Cerebrovasc Dis Extra 2011;1:105–114

DOI: 10.1159/000334473
Published online: December 3, 2011

© 2011 S. Karger AG, Basel
www.karger.com/cee

110

Craig et al.: Predictive Model for Stroke Progression

**Table 1** (continued)

|  | Training data set (n = 863) | Test data set (n = 216) |
|---|---|---|
| Hypoxia |  |  |
|     Yes | 154 (17.8%) | 30 (13.9%) |
|     No | 683 (79.1%) | 185 (85.7%) |
|     Missing data | 26 (3.0%) | 1 (0.5%) |
| Bad physiological feature |  |  |
|     None recorded | 385 (44.6%) | 77 (35.7%) |
|     ≥1 feature | 475 (55.0%) | 119 (55.1%) |
|     Missing data | 3 (0.4%) | 20 (9.3%) |

Entries are n (%), unless stated otherwise. LACS = Lacunar syndrome; POCS = posterior circulation syndrome; PICH = primary intracerebral haemorrhage; TIA = transient ischaemic attack.

[1] For the test data set, only systolic blood pressure on admission is available.

[2] Estimate from the literature [18].

0.1–33.4% of the patients. The variables with the highest level of missing data were alcohol use (33.4%), smoking status (19.0%), hyperglycaemia (11.0%) and hyperosmolarity (12.6%; table 1).

On univariate analysis, age, alcohol intake, living alone, previous independence, verbal score, arm power, leg power, conscious level, history of coronary heart disease, smoking status, CT lesion, OCSP classification, pyrexia, hyperosmolarity, one abnormal physiological sign, maximum and minimum systolic blood pressure, and side of lesion were significantly associated with stroke progression. The results of logistic regression of the imputed data set are shown in table 2. Conscious level, history of coronary heart disease, presence of hyperosmolarity, living alone on admission, OCSP classification and presence of pyrexia were associated with a significant increase in the odds of stroke progression. In particular, patients who experienced a total anterior circulation syndrome (ACS) were twice as likely to have a progression of stroke as those that had a partial ACS. In contrast, patients who had no lesion on their CT scan had significantly reduced odds of stroke progression in comparison to those with an infarction on CT scan; a similar effect was found in patients who were current or ex-smokers in comparison to those that were non-smokers.

A similar predictive model was developed, based only on the observed data in the training data set (table 2). Both models shared common independent predictors, with the exception of lesion on CT and living alone; these were not shown to be predictive in the observed data model.

When testing the performance of the predictive model based on the 10 test data sets generated by missing imputation, the median ROC value was 0.72 (range 0.72–0.73). This implies that the model generated is moderately accurate and better than chance alone in predicting stroke progression. The median p value for the Hosmer-Lemeshow goodness-of-fit test was 0.90 (range 0.50–0.92), indicating a good fit between the model and the observed data. The observed number of patients with and without stroke progression is compared with the number expected in table 3. The performance of the predictive model was also tested using the training data set. The median ROC value 0.73 (range 0.72–0.74) was similar to that evaluated in the test data set. However, the median p value of 0.38 (range 0.06–0.70) for the Hosmer-Lemeshow goodness-of-fit test was lower indicating a poorer fit between the model and the observed data.

Cerebrovasc Dis Extra 2011;1:105–114

DOI: 10.1159/000334473
Published online: December 3, 2011

© 2011 S. Karger AG, Basel
www.karger.com/cee

111

Craig et al.: Predictive Model for Stroke Progression

**Table 2.** Predictive model for stroke progression

| Predictors | Model with imputed data (n = 863) | | Model with observed data (n = 603) | |
|---|---|---|---|---|
| | OR (95% CI) | p value | OR (95% CI) | p value |
| Conscious level | | | | |
| Reduced | 1.00 | | 1.00 | |
| Normal | 1.70 (0.95–3.00) | 0.07 | 3.41 (1.51–7.72) | <0.01 |
| Coronary heart disease | | | | |
| No | 1.00 | | 1.00 | |
| Yes | 1.44 (1.01–2.10) | 0.05 | 1.53 (0.99–2.36) | <0.05 |
| Hyperosmolarity | | | | |
| No | 1.00 | | 1.00 | |
| Yes | 1.55 (1.03–2.32) | 0.03 | 1.93 (1.24–3.02) | <0.01 |
| Lesion on CT scan | | | | |
| Infarction | 1.00 | (overall <0.01) | – | |
| No scan | 0.72 (0.41–1.24) | 0.23 | – | |
| No lesion visible | 0.40 (0.23–0.70) | <0.01 | – | |
| PICH | 1.60 (0.90–2.90) | 0.11 | – | |
| Other | 0.85 (0.30–2.70) | 0.79 | – | |
| Living alone | | | | |
| Yes | 1.00 | | – | |
| No | 1.54 (1.05–2.25) | 0.03 | – | |
| OCSP classification | | | | |
| Partial ACS | 1.00 | (overall <0.01) | 1.00 | (overall <0.01) |
| Not classifiable | 0.44 (0.18–1.10) | 0.08 | 0.16 (0.37–0.69) | <0.01 |
| Total ACS | 2.30 (1.44–3.70) | <0.01 | 2.56 (1.46–4.49) | <0.01 |
| LACS | 0.81 (0.51–1.30) | 0.37 | 0.58 (0.33–0.99) | 0.05 |
| POCS | 0.74 (0.34–1.63) | 0.45 | 0.67 (0.26–1.70) | 0.40 |
| Pyrexia | | | | |
| No | 1.00 | | 1.00 | |
| Yes | 1.73 (1.17–2.55) | 0.01 | 2.10 (1.29–3.30) | <0.01 |
| Smoker | | | | |
| Not | 1.00 | (overall 0.02) | 1.00 | (overall <0.01) |
| Current | 0.63 (0.41–0.95) | 0.03 | 0.59 (0.40–0.94) | 0.03 |
| Ex-smoker | 0.56 (0.33–0.94) | 0.03 | 0.46 (0.25–0.83) | 0.01 |

Overall = Median p value; LACS = lacunar syndrome; POCS = posterior circulation syndrome; PICH = primary intracerebral haemorrhage.

## Discussion

This analysis suggests that conscious level, history of coronary heart disease, the lesion shown on CT scan, OCSP classification, living alone on admission, presence of hyperosmolarity, presence of pyrexia and smoking status are independent predictors of stroke progression. Some of the variables identified in this study have also been found to be predictive of stroke progression in other studies. However, some discrepancies do exist. For example, in this study, patients who experienced a total ACS were twice as likely to have a progression of stroke than those who experienced a partial ACS, while a previous study showed that neurological progression was observed twice as often in patients with posterior circulation infarction than in those with anterior circulation infarction [14]. Similarly, although reduced level of consciousness and a history of coronary heart disease have been associated with increased risk of stroke progression in a few other studies [2, 15], in this study it was those pa-

**Table 3.** Predicted probabilities, observed and expected numbers (medians) by decile[1]

| Decile | Predicted probabilities | Observed progression | Expected progression | Observed no progression | Expected no progression | Total |
|---|---|---|---|---|---|---|
| 1 | 0.05 | 2.0 | 1.1 | 23.0 | 23.9 | 25.0 |
| 2 | 0.07 | 1.0 | 1.3 | 18.0 | 17.8 | 19.0 |
| 3 | 0.11 | 1.0 | 1.9 | 20.0 | 19.2 | 21.0 |
| 4 | 0.13 | 2.0 | 2.6 | 20.0 | 19.4 | 22.0 |
| 5 | 0.16 | 3.0 | 3.0 | 18.0 | 18.0 | 21.0 |
| 6 | 0.20 | 4.0 | 4.0 | 17.5 | 18.0 | 22.0 |
| 7 | 0.24 | 4.0 | 4.9 | 18.0 | 17.1 | 22.0 |
| 8 | 0.28 | 6.5 | 5.5 | 14.0 | 15.5 | 21.0 |
| 9 | 0.36 | 5.5 | 6.9 | 16.5 | 15.1 | 22.0 |
| 10 | 0.53 | 11.0 | 8.9 | 10.0 | 12.0 | 21.0 |

[1] This goodness-of-fit table shows the observed proportion of events against the predicted probability for 10 groups of equal size. For a perfectly fitted model, the observed proportions of events and predicted probability should create a 45° line, i.e. the slope is 1. This table is accompanied by the Hosmer-Lemeshow test giving a p value of 0.90, indicating that there is minimal difference between the observed and predicted event probabilities.

tients that had a normal conscious level that appeared to have an increased odds of stroke progression. The way in which these variables were measured clinically and how they were put into the predictive model could account for these differences. Finding the variables not living alone and having a normal level of consciousness being predictive of stroke progression is considered unexpected. It may be that not living alone is associated with a higher level of frailty. This may also be the result of over-fitting (including too many variables in) the model resulting in the inclusion of apparently important predictors which are not actually independent predictors [16]. Stroke care may have changed since the study was conducted between 2000 and 2004. Therefore, it would be interesting to assess the performance of the model in a more recent cohort of patients. However, differences in performance may be difficult to attribute to changes in stroke care.

In terms of calibration, the model showed good agreement between the predicted probabilities of stroke progression and those actually observed. Evaluating the performance of predictive models in a new cohort of patients as opposed to the original cohort used to develop the model is an important feature when testing the validity of predictive models. In this study, the calibration properties of the model were different when evaluated in the test data set as when evaluated in the training data set. One reason for this may be that there were differences although a comparison of the demographics of the patients in the two cohorts suggests that they are similar. The test cohort contained only patients who had an ischaemic stroke and the care pathway between these two cohorts did differ. During the recruitment of the first cohort of patients, the hospital implemented a new acute stroke unit which was in place for the second cohort. This may have resulted in a different distribution of outcomes in this patient population, which may have influenced the calibration of the model. The model appeared to be able to discriminate fairly accurately between those patients likely to progress and those who were not when using the test data set. A very similar ROC value was derived when the model was tested in the training set. However, it is important to extend beyond the internal validation approach used in this study and investigate the model in not only a separate patient cohort, but a cohort that is from a different popula-

Cerebrovasc Dis Extra 2011;1:105–114

DOI: 10.1159/000334473
Published online: December 3, 2011

© 2011 S. Karger AG, Basel
www.karger.com/cee

113

Craig et al.: Predictive Model for Stroke Progression

tion. This would challenge the external validity and generalisability of the model as opposed to assessing internal validation, which may result in the performance of the model being overly optimistic. This reinforces the inherent difficulty with predictive modelling – how feasible is it to have one model apply to all patient subgroups across different models of care and countries [17]? Indeed, a few studies attempted to investigate how prognostic variables may vary across subgroups of patients or by the different causes of stroke. One study identified that different factors were associated with neurological worsening in different causes of stroke [15].

Predictive models for patients with acute stroke can be useful in informing patient management and a number of predictive models in stroke research exist. A common limitation in this area is the management of data sets with data on potential prognostic factors missing. A standard approach to manage this is to conduct a complete case analysis whereby patients that have missing data are excluded from the analysis. This could lead to the exclusion of a defined subset of patients, for example unconscious patients, where it has not been possible to ascertain their smoking habits. The alternative approach, multiple imputation, as used in this study, has not only increased the statistical power of the analysis but also eliminated the bias associated with excluding patients in a complete case analysis. The increased power in the imputed data set resulted in two additional factors found to be significant. Also, on comparing the models based on imputed and observed training data, the confidence intervals of the odds ratios are narrower in model based on imputed data, especially for conscious level and OCSP classification. This is to be expected since these variables did not have any missing data. By using multiple imputation, a technique not yet considered standard in managing missing data, this analysis minimises the bias associated with the more commonly used complete case approach. This study also highlights some of the recognised methodological challenges in prognostic research, such as validity and credibility of models in clinical practice.

## Conclusion

The predictive model developed in this study contains variables that can be easily collected in practice, therefore increasing its clinical applicability. This analysis shows that conscious level, a history of coronary heart disease, presence of hyperosmolarity, the lesion shown on CT scan, living alone on admission, OCSP classification, presence of pyrexia and smoking status are independent predictors of stroke progression. Using this analysis approach, the discrimination and calibration of the predictive model appear sufficiently high to provide accurate predictions.

## Acknowledgments

## Disclosure Statement

The authors have no conflicts of interest.

## References

1  Dávalos A, Castillo J: Progressing stroke; in Fisher M, Bogousslavsky J (eds): Current Review of Cerebrovascular Disease. Philadelphia, Current Medicine, 2001, pp 169–181.
2  Birschel P, Ellul J, Barer D: Progressing stroke: towards an internationally agreed definition. Cerebrovasc Dis 2004;17:242–252.
3  Matsumoto N, Kimura K, Yokota C, Yonemura K, Wada K, Uchino M, et al: Early neurological deterioration represents recurrent attack in acute small non-lacunar stroke. J Neurol Sci 2004;217:151–155.
4  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.
5  Jorgensen HS, Nakayama H, Olsen TS, Raaschou HO: Effect of blood pressure and diabetes on stroke in progression. Lancet 1994;344:156–159.
6  Toni D, Fiorelli M, Gentile M, Bastianello S, Sacchetti ML, Argentino C, et al: Progressing neurological deficit secondary to acute ischemic stroke. A study on predictability, pathogenesis, and prognosis. Arch Neurol 1995;52:670–675.
7  Lin, LC, Yang, JT, Weng HH, Hsiao CT, Lai SL, Fann WC: Predictors of early clinical deterioration after acute ischemic stroke. Am J Emerg Med 2011;29:577–581.
8  Kwon H, Lim J, Park H, Lee Y: Hypertriglyceridemia as a possible predictor of early neurological deterioration in acute lacunar stroke. J Neurol Sci 2011;309:128–130.
9  Dávalos A, Toni D, Iweins F, Lesaffre E, Bastianello S, Castillo J: Neurological deterioration in acute ischemic stroke: potential predictors and associated factors in the European Cooperative Acute Stroke Study (ECASS). Stroke 1999;30:2631–2636.
10 Royston P, Moons KG, Altman DG, Vergouwe Y: Prognosis and prognostic research: developing a prognostic model. BMJ 2009;338:1373–1377.
11 Harrell FE, Lee KL, Mark DB: Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–387.
12 Hosmer DW, Lemeshow S: Applied Logistic Regression, ed 2. New York, Wiley, 2000.
13 Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al: Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:157–160.
14 Jones HR, Millikan CH, Sandok BA: Temporal profile (clinical course) of acute vertebrobasilar system cerebral infarction. Stroke 1980;11:173–177.
15 Yamamoto H, Bogousslavsky J, van Melle G: Different predictors of neurological worsening in different causes of stroke. Arch Neurol 1998;55:481–486.
16 Counsell C, Dennis M, McDowall M, Warlow C: Predicting outcome after acute and subacute stroke: development and validation of new prognostic models. Stroke 2002;33:1041–1047.
17 Moons KG, Altman DG, Vergouwe Y, Royston P: Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:1318–1320.
18 Barber M, Wright F, Stott DJ, Langhorne P: Hemostatic function and progressing ischemic stroke: D-dimer predicts early clinical progression. Stroke 2004;35:1421–1425.