Derivation and External Validation of a Case Mix Model for the Standardized Reporting of 30-Day Stroke Mortality Rates

Benjamin D. Bray, MRCP; James Campbell, LLB; Geoffrey C. Cloud, FRCP; Alex Hoffman, MSc; Martin James, FRCP; Pippa J. Tyrrell, FRCP; Charles D.A. Wolfe, FFPHM; Anthony G. Rudd, FRCP; on behalf of the Intercollegiate Stroke Working Party Group

Background and Purpose—Case mix adjustment is required to allow valid comparison of outcomes across care providers. However, there is a lack of externally validated models suitable for use in unselected stroke admissions. We therefore aimed to develop and externally validate prediction models to enable comparison of 30-day post-stroke mortality outcomes using routine clinical data.

Methods—Models were derived (n=9000 patients) and internally validated (n=18 169 patients) using data from the Sentinel Stroke National Audit Program, the national register of acute stroke in England and Wales. External validation (n=1470 patients) was performed in the South London Stroke Register, a population-based longitudinal study. Models were fitted using general estimating equations. Discrimination and calibration were assessed using receiver operating characteristic curve analysis and correlation plots.

Results—Two final models were derived. Model A included age (<60, 60–69, 70–79, 80–89, and ≥90 years), National Institutes of Health Stroke Severity Score (NIHSS) on admission, presence of atrial fibrillation on admission, and stroke type (ischemic versus primary intracerebral hemorrhage). Model B was similar but included only the consciousness component of the NIHSS in place of the full NIHSS. Both models showed excellent discrimination and calibration in internal and external validation. The c-statistics in external validation were 0.87 (95% confidence interval, 0.84–0.89) and 0.86 (95% confidence interval, 0.83–0.89) for models A and B, respectively.

Conclusions—We have derived and externally validated 2 models to predict mortality in unselected patients with acute stroke using commonly collected clinical variables. In settings where the ability to record the full NIHSS on admission is limited, the level of consciousness component of the NIHSS provides a good approximation of the full NIHSS for mortality prediction. (Stroke. 2014;45:3374-3380.)

Key Words: mortality ■ outcome assessment (health care)

Despite improvements in the prognosis of acute stroke seen during the past 20 years,1 it remains a condition with a high early mortality, with a risk of death of between 10% and 30% in the 30 days after stroke.2,3 In predicting outcomes after stroke, a variety of models have been developed that use information about patient characteristics to predict mortality and disability outcomes.4-6 These models may be used to inform decision making, provide more accurate estimates of prognosis, plan services, and to facilitate risk-adjusted comparisons between care providers. Patients, the public, clinicians, healthcare funders, and politicians are increasingly demanding to know that care provision is of high quality and therefore asking to be able to compare outcomes between providers. It is essential that if mortality is publicly reported there should be confidence that the data are accurately adjusted for case mix variability and that the differences genuinely reflect variations in care.

We aimed to develop and validate a model to predict the 30-day risk of death after admission with acute stroke from a large, prospective data set of unselected cases of acute stroke admitted to hospitals in England and Wales. The primary rationale was to use the resulting model to enable reporting of case mix–adjusted mortality outcomes for stroke services. None of the previously published algorithms were considered
Data Sources
Data were drawn from 2 sources: The Sentinel Stroke National Audit Program (SSNAP), the national stroke register of England and Wales,8 and the South London Stroke Register (SLSR). SSNAP is a prospectively collected register of patients admitted to hospital with acute stroke in England and Wales. Data are entered by clinical teams using a secure online web portal with built-in validation and data checking. Variables include demographic details, stroke phenotype, comorbidities, process of care, and clinical outcomes. Mortality status is identified by record linkage with the statutory national register of deaths. Data linkage was performed by a secure third party and the investigators were provided with an anonymized data set following removal of patient identifiers. Case ascertainment from stroke centers was estimated by comparison with stroke coding in administrative returns. Patients aged ≥18 years admitted with acute ischemic stroke or primary ICH from January 1, 2013 to June 30, 2013 were included in the data set. The data set was split at random into a derivation (one third) and an internal validation (two thirds) samples.

Data for external validation was drawn from the SLSR, a population-based study of all people with first-ever stroke in a defined population (357,308 inhabitants in 2010) of South London.10 Patients with stroke are identified using multiple overlapping sources of notification and reviewed at the time of stroke by a trained field worker. Detailed methods are published elsewhere.10 Patients aged ≥18 years admitted to hospital with ischemic stroke or primary ICH from 2005 to 2012 and with a documented National Institutes of Health Stroke Severity Score (NIHSS) on admission were included. Patients with subarachnoid hemorrhage were excluded from the analysis because they have a different clinical pathway and distinct epidemiology.

Statistical Analysis
The prognostic model was developed from the derivation sample (n=9000) and subsequently validated in the internal (n=18,169) and external (n=1470) validation data sets. Univariable associations between patient characteristics and 30-day mortality were assessed using χ² tests for categorical variables and Mann–Whitney tests for continuous variables. Continuous variables were described as medians with interquartile range.

The mortality model was fitted to the derivation data set using a process of backward elimination. The selection of the variables for initial inclusion was done following a literature review of stroke epidemiology and previously published stroke mortality models. A literature search for previously published models was performed in Pubmed using the following terms: Stroke AND (Mortality OR Death) AND Prediction. In addition to primary studies describing the derivation and validation of stroke mortality models, the search identified 1 systematic review.7 Variables available in both data sets and identified from previously published mortality models included demographic details (age, sex), stroke type, mode of admission, pre-stroke functional level, stroke severity (NIHSS, level of consciousness), and cardiovascular comorbidity (diabetes mellitus, congestive cardiac failure, hypertension, atrial fibrillation, previous stroke, or transient ischemic attack). Data of noncardiovascular comorbidities were not available in the derivation data set. The models were fitted using general estimating equations with logit links, taking into account clustering at the hospital level. Analytic 95% confidence intervals were computed. Various methods of handling continuous variables were assessed, including nonlinear transformations and categorization. Age was nonlinearly associated with mortality risk and was included as a 5-level categorical variable in the final models. Inclusion of the NIHSS as a continuous predictor produced better model fit than categorization. Nonsignificant variables (P<0.05) were removed from the model by backward elimination.

Complete data were available for all variables in the SSNAP data set apart from the full NIHSS score on admission, which was available for 76% of patients. A total of 373 patients with unspecified stroke type were not included. Two models were therefore fitted: 1 including the NIHSS (model A) and 1 including only the NIHSS level of consciousness component (model B), because this was available for 100% of patients.

The accuracy of the mortality predictions generated in the final models was assessed in the internal and validation data sets. The coefficients and variables from each of the models were used to calculate the predicted probability of death for each patient in the internal and validation data sets. Discrimination was assessed by receiver operating characteristic curve analysis and estimation of the c-statistic. Calibration was visually assessed using calibration plots, grouping patients into deciles of predicted mortality risk. The observed and predicted mortality rates were plotted by deciles of predicted risk. The correlation between observed and predicted mortality rates were assessed in scatter plots and the Pearson correlation coefficient estimated. Given the large sample sizes used in this study, we did not use Hosmer and Lemeshow goodness-of-fit tests to assess calibration, because the test is recognized to be sensitive to sample size.11 Analyses were performed in Stata 12 MP (Stata Corp).

Results
Data were available for 27,169 patients with acute stroke in the SSNAP data set and 1470 patients in the SLSR data set. The median age was higher in the derivation (77 years; interquartile range [IQR], 67–85 years) and internal validation data sets (77 years; IQR, 67–85 years) compared with the external validation data set (72 years; IQR, 59–81; Kruskal–Wallis P<0.0001; Table 1). There were also statistically significant differences in the distribution of stroke types between SSNAP and the SLSR. The proportion of ischemic strokes was 89.5%, 89.6%, and 86.3% in the derivation, internal validation, and external validation data sets, respectively. Atrial fibrillation was more frequent in the SSNAP data set (20.3% and 20.4%) than in the SLSR (16.7%; χ² P=0.003). Stroke severity was also overall lower in the derivation and internal validation data sets (median NIHSS, 4; IQR, 2–10 in both) than in the external validation data set (median NIHSS, 6; IQR, 3–14). Thirty-day mortality rates also differed significantly: 13.2% in the derivation, 12.0% in the internal validation, and 15.3% in the external validation groups (χ² P<0.0001).

The variables and coefficients of the 2 final models that were developed from the derivation data set are shown in Table 2. Model A included age in 5 categories (<60, 60–69, 70–79, 80–89 and ≥90 years), NIHSS on admission, the presence of atrial fibrillation, and stroke type (ischemic or primary ICH). Model B was similar but included the 4 levels of the NIHSS consciousness component.

Both models A and B showed good discrimination in both the SSNAP and SLSR validation data sets (Figure 1; and see online-only Data Supplement), with model A overall having slightly higher c-statistics. For model A, the c-statistics in the
internal validation data set were 0.86 (95% confidence interval, 0.85–0.88) and 0.87 (0.84–0.89), respectively. For model B the corresponding c-statistics were 0.82 (0.81–0.83) and 0.86 (0.83–0.89; Table 3).

Both models also were well calibrated across the whole range of mortality risk (Figures 2 and 3; see online-only Data Supplement), correctly classifying patients across a wide range of mortality risk. In external validation, model A tended to predict lower mortality in the highest risk patients than was observed in the external validation cohort but overall observed mortality scaled appropriately across deciles of predicted risk. Both models showed a high degree of correlation

Table 1. Characteristics of the Cohorts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Derivation</th>
<th>Internal Validation</th>
<th>External Validation</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients with stroke included in SSNAP: January to June 2013</td>
<td>Patients with stroke included in SSNAP: January to June 2013</td>
<td>Patients with stroke recruited to the South London Stroke Register, 2005–2012</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>9000</td>
<td>18,169</td>
<td>1,470</td>
<td>...</td>
</tr>
<tr>
<td>Age (median, IQR), y</td>
<td>77 (67–85)</td>
<td>77 (67–85)</td>
<td>72 (59–81)</td>
<td>0.0001</td>
</tr>
<tr>
<td>Women (%)</td>
<td>50.2</td>
<td>50.4</td>
<td>52.6</td>
<td>0.82</td>
</tr>
<tr>
<td>Stroke type (%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Ischemic</td>
<td>89.5</td>
<td>89.6</td>
<td>86.3</td>
<td>...</td>
</tr>
<tr>
<td>ICH</td>
<td>10.5</td>
<td>10.4</td>
<td>13.7</td>
<td>...</td>
</tr>
<tr>
<td>Atrial fibrillation (%)</td>
<td>20.3</td>
<td>20.4</td>
<td>16.7</td>
<td>0.003</td>
</tr>
<tr>
<td>Previous stroke/TIA (%)</td>
<td>26.8</td>
<td>26.9</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Median NIHSS (IQR)</td>
<td>4 (2–10)</td>
<td>4 (2–10)</td>
<td>6 (3–14)</td>
<td>0.0001</td>
</tr>
<tr>
<td>NIHSS consciousness (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0: alert</td>
<td>83</td>
<td>82.5</td>
<td>74.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>1: verbal response</td>
<td>9.9</td>
<td>10.7</td>
<td>13.7</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>2: pain response</td>
<td>3.9</td>
<td>4.1</td>
<td>6.3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>3: unconscious</td>
<td>3.2</td>
<td>2.7</td>
<td>5.8</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>30-Day mortality (%)</td>
<td>13.2</td>
<td>12.0</td>
<td>15.3</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

ICH indicates intracerebral hemorrhage; IQR, interquartile range; NIHSS, National Institutes of Health Stroke Severity Score; SSNAP, Sentinel Stroke National Audit Program; and TIA, transient ischemic attack.

Table 2. Variables and Coefficients of the Final Models

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>95% CI</th>
<th>Coefficient</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;60</td>
<td>0.000</td>
<td>NA</td>
<td>0.000</td>
<td>NA</td>
</tr>
<tr>
<td>60 to 69</td>
<td>0.624</td>
<td>0.046 to 1.21</td>
<td>0.657</td>
<td>0.25 to 1.06</td>
</tr>
<tr>
<td>70 to 79</td>
<td>1.033</td>
<td>0.51 to 1.55</td>
<td>1.252</td>
<td>0.89 to 1.62</td>
</tr>
<tr>
<td>80 to 89</td>
<td>1.488</td>
<td>0.98 to 1.99</td>
<td>1.613</td>
<td>1.26 to 1.97</td>
</tr>
<tr>
<td>≥90</td>
<td>1.781</td>
<td>1.24 to 2.32</td>
<td>2.127</td>
<td>1.75 to 2.50</td>
</tr>
<tr>
<td>NIHSS</td>
<td>0.137</td>
<td>1.26 to 1.49</td>
<td>0.000</td>
<td>NA</td>
</tr>
<tr>
<td>Atrial fibrillation (%)</td>
<td>0.425</td>
<td>0.20 to 0.65</td>
<td>0.000</td>
<td>NA</td>
</tr>
<tr>
<td>Stroke type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischemic</td>
<td>0.000</td>
<td>NA</td>
<td>1.585</td>
<td>1.41 to 2.76</td>
</tr>
<tr>
<td>ICH</td>
<td>0.870</td>
<td>0.68 to 1.17</td>
<td>1.252</td>
<td>0.89 to 1.62</td>
</tr>
<tr>
<td>Constant</td>
<td>−5.250</td>
<td>−5.75 to −4.75</td>
<td>3.564</td>
<td>3.25 to 3.88</td>
</tr>
</tbody>
</table>

Predicted mortality is calculated for individual patients using standard methods to generate predictions from the coefficients of logistic models. CI indicates confidence interval; ICH, intracerebral hemorrhage; NA, not applicable; and NIHSS, National Institutes of Health Stroke Severity Score.
We have derived and validated 2 simple models to predict 30-day mortality in unselected patients admitted with acute stroke to hospitals in the United Kingdom. One model uses the complete NIHSS on admission and the other uses the 4 levels of the NIHSS consciousness component. Both models showed good discrimination and were well calibrated in internal and external data sets, although the model including the full NIHSS score demonstrated slightly better discrimination. The variables included in the models are likely to be collected routinely on patients with stroke in many healthcare settings and so potentially have wide applicability. In settings where training or resources limit the collection of the full NIHSS, information on consciousness level provides a good proxy for the full NIHSS in predicting 30-day mortality.

The study was not intended to provide new insight into the epidemiology of stroke: it is well recognized that older age, higher NIHSS, hemorrhagic stroke, and atrial fibrillation are all associated with worse prognosis. Atrial fibrillation is a marker for cardioembolic stroke, which is associated with the poorest outcomes in ischemic stroke. In the current study, atrial fibrillation was also a predictor for higher mortality in ICH. The reason for this is not clear, but it may be that anticoagulation-associated ICH is associated with poor outcomes or that atrial fibrillation is a marker for cardiovascular comorbidity.

The main contribution of this study is that it developed relatively simple and parsimonious models to make accurate predictions of the risk of 30-day mortality in unselected populations of stroke. Because they use variables routinely recorded during the assessment of patients with stroke and have been validated in both ischemic and ICH stroke, the models are likely to be comparatively straightforward to implement in the comparison of mortality rates between stroke care providers. Developing standardized and validated methods for adjusting mortality rates is essential when comparing the outcomes of stroke services so that apparent variation does not just reflect differences in case mix.

A variety of models have been developed previously to predict outcomes after stroke. A recent systematic review identified 17 models derived and externally validated in stroke cohorts. More recent studies have generated ≥1 further externally validated model. Seven of these models predict
mortality after stroke. As with the models developed in this study, most include age and some measure of stroke severity. The most widely validated model is the simple six variable (SSV) model,\(^\text{15}\) which includes age, living alone, independence prestroke, Glasgow coma scale verbal score, ability to lift both arms, and ability to walk. The discrimination of the models derived in the present study is similar to that reported for the simple six variable, which has been found in external validation to have a c-statistic of 0.73 to 0.86.\(^\text{15-17}\) A more recently developed model to predict outcomes is the iScore, which has been externally validated in cohorts from several countries.\(^\text{18,19}\) In addition to age and stroke severity (measured by the Canadian Neurological Scale), it includes comorbidity, serum glucose level, and prestroke dependence. Discrimination of the iScore is good and again similar to the models developed in this study, with c-statistics of 0.85 to 0.87.\(^\text{18,19}\); however, it has only been validated in patients with ischemic stroke. Several previous studies have incorporated the NIHSS\(^\text{13,20}\) and demonstrated the strong association of this scale with stroke outcomes. In the United Kingdom, commonly used mortality measures for stroke care (eg, Dr Foster Intelligence)\(^\text{21}\) use administrative data. These are however limited in that routine coding contains no information on stroke severity and the models have not been externally validated in stroke: their accuracy is therefore unclear and likely to be less than models based on clinical data sets.

One of the strengths of the study was that the models were validated in a cohort of patients identified in a population-based register. The cohort is therefore more likely to truly represent a population of unselected patients with stroke than cohorts from other sources, such as randomized controlled trials or patients identified through routine administrative coding. In addition, the characteristics of the patients in the external data set differed significantly from the derivation data set. The differences in the populations between the SLSR and SSNAP might reflect that the SLSR data are from an earlier time period, more complete ascertainment in the population-based SLSR, or the relatively young, multiethnic inner city population of the SLSR denominator population.\(^\text{10}\) That the models made almost as accurate predictions in a stroke population with different
characteristics than were used to derive the models increases the external validity of the models. Another advantage of the study is that both the derivation and validation data sets included unselected patients with ischemic stroke or primary ICH, and thus is likely to be more representative of the broad range of patients with stroke admitted to hospital. The sample size was large (larger than the sum total of all 17 previous studies identified in systematic review), includes patients admitted with acute stroke to all hospitals in England and Wales, and reflects the outcomes of contemporary stroke care. Many of the previously developed models were based on data from single sites or interventional cohorts, or were derived from historic cohorts of patients who had stroke at a time before the widespread use of interventions proven to improve outcomes after stroke, such as organized stroke unit care. The validity of these models in predicting the outcomes of contemporary patients with stroke is therefore uncertain.

**Limitations**
The models were derived on data collected during routine care, and although participating teams are encouraged to enter data on all consecutive stroke admissions, we cannot exclude selection bias in the sample. Ideally models are validated in a prospective cohort specifically set up to test the validity of the model. One weakness is therefore that we used a retrospective cohort for external validation. Both samples were drawn from the UK population and external validation in other settings would be important to consider before using the models elsewhere, particularly in populations with a higher rate of ICH. Mortality outcomes were limited to all-cause mortality and did not take into account clinical decisions that might contribute to mortality rates, such as withdrawal of active treatment. The choice of variables to be included in the models was determined by their availability in SSNAP; some potentially important variables that may have improved the precision of the model (eg, levels of comorbidity) were not available. Patients with undetermined stroke type were not included in the model, and therefore these models may not be suitable for use in low-income countries where it may not be possible to determine whether the stroke was ischemic or the result of a primary ICH.

**Conclusions**
This study has derived and externally validated 2 new models to predict 30-day mortality after stroke, using a large sample of contemporary patients with stroke. They show good accuracy in predicting 30-day mortality in unselected patients with ischemic stroke or primary ICH, are simple to calculate and use variables that are recorded on admission in many stroke centers. These prognostic models permit the validated comparison of adjusted 30-day mortality outcomes between stroke centers in the United Kingdom.

**Acknowledgments**
We would like to thank the many hundreds of individuals who have contributed to Sentinel Stroke National Audit Program (SSNAP). The audits would not be possible without their hard work and dedication. We would also like to thank the Intercolloquiate Stroke Working Party for their guidance and work to support the audits and the patients who have participated in the South London Stroke Register. We also thank Hannah Evans for her helpful comments on the methodology of the study. Ethical approval of SSNAP was granted by the Ethics and Confidentiality Committee of the National Information Governance Board. Further ethical approval was not sought.

**Sources of Funding**
Sentinel Stroke National Audit Program is commissioned by the Healthcare Quality Improvement Partnership on behalf of the Department of Health in England. No specific funding from any source was sought for this study. The research was supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy’s and St Thomas’ National Health Service (NHS) Foundation Trust and King’s College London. Dr James is supported by the NIHR Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for the South West Peninsula. This article presents independent research funded (in part) by the NIHR CLAHRC. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health in England. No funding bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the article.

**Disclosures**
A. Hoffman is funded by the Healthcare Quality Improvement Partnership. P.J. Tyrrell is a coinvestigator on a grant to investigate relative outcomes in 2 different models of urban stroke service reorganization in the United Kingdom. The other authors report no conflicts.

**References**


Derivation and External Validation of a Case Mix Model for the Standardized Reporting of 30-Day Stroke Mortality Rates
Benjamin D. Bray, James Campbell, Geoffrey C. Cloud, Alex Hoffman, Martin James, Pippa J. Tyrrell, Charles D.A. Wolfe and Anthony G. Rudd
on behalf of the Intercollegiate Stroke Working Party Group

Stroke. 2014;45:3374-3380; originally published online October 7, 2014;
doi: 10.1161/STROKEAHA.114.006451

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/45/11/3374

Data Supplement (unedited) at:
http://stroke.ahajournals.org/content/suppl/2014/10/07/STROKEAHA.114.006451.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

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

Subscriptions: Information about subscribing to Stroke is online at:
http://stroke.ahajournals.org//subscriptions/
Supplemental Figure I. Receiver-Operator-Curves for Model B in derivation, internal validation and external validation datasets
Supplemental Figure II. Calibration plot for Model B in derivation, internal validation and external validation datasets. Bars are 95% confidence intervals.
Supplemental Figure III. Observed-versus predicted 30 day mortality for Model B in derivation, internal validation and external validation datasets.