Predicting stroke outcome: Guy’s prognostic score in practice

Patrick Gompertz, Pandora Pound, Shah Ebrahim

Abstract
This study aimed to cross validate the Guy’s prognostic score in a new sample of patients of all ages admitted to hospital with a stroke and to devise and test a simpler version (the G-score). 361 consecutive acute patients with stroke who had been admitted to the acute hospitals in two adjacent health districts in East London were recruited and followed up for six months after the stroke. The G-score was derived by simplifying the weights used in calculating the Guy’s score. With the conventional threshold of 0 for the Guy’s score and 3 (out of 7) for the G-score, the sensitivity of both scores for predicting a bad outcome (death or Barthel score <13 out of 20 at six months) was 0·72 and specificity was 0·63. The likelihood ratio for the Guy’s score was 1·97 and for the G-score 1·95. Both versions of the score performed better than conscious level alone at predicting the outcome (sensitivity 0·47, specificity 0·73, likelihood ratio 1·74). Similar data are presented for different thresholds and prior probabilities. In addition, the G-score permits direct estimation of 95% confidence intervals for the probability of a bad outcome for five grades of stroke severity. Outcome prediction with multivariate techniques has the potential to improve and inform clinical decision making. The G-score should be used to define stroke severity for individual patients and for case mix adjustment.

(J Neural Neurosurg Psychiatry 1994;57:932-935)

“What are my chances, doctor?” In most cases, the clinician’s answer to this question would be based on personal experience and knowledge of the prognostic importance of some clinical features like unconsciousness or incontinence, but would have to be expressed in vague terms. Clearly, all clinicians would like to be able to answer this question more precisely. Accurate and early prediction of the outcome of acute stroke would also help early management and the planning of rehabilitation and has been shown to improve the management of patients with head injury.1 Formal assessment of stroke severity is also useful in clinical trials and for comparison of outcome between units, where it is essential to identify differences in case mix.

Whereas it may be sufficient when dealing with groups of patients to simply predict a “good” or “bad” outcome, the clinician at the bedside needs to be able to give a graded prognosis in the same way that some cancers are staged. The prognostication is never going to be perfectly accurate, so it is important to include some indication of the size of the error. Also, the significance of a particular test score will depend on the prior probability of a bad outcome and accuracy will be improved if this is taken into account. Even if a score is used to divide patients into good and bad prognostic groups, the appropriate threshold may differ in different circumstances. Data on the performance of a test at various thresholds are needed.

Several formal prognostic scores have been devised.2 Although they have been used to measure case mix for audit and clinical trials, no score is in common clinical use. Formal scores have not gained acceptance because of doubts about accuracy,3 their applicability to other groups of patients,4 and practical difficulties in their use. For example, the Guy’s prognostic score5 is obtained by giving weights to age and some important clinical features 24 hours after onset of stroke and adding them to a constant (table 1). The resulting score typically varies between −40 and +40. An estimate of the probability of functional independence at six months is then read off a graph. The validity of the score was examined by testing its ability to separate the patients from whom it was derived into good and bad prognostic groups (sensitivity 0·84, specificity 0·92) but no attempt was made to replicate the score performance in another group of patients, no evaluation of the probability estimates was published, and the original sample excluded patients over 76 years old. Also, the score is not simple to use in practice because the calculations are too difficult to do at the bedside.

<table>
<thead>
<tr>
<th>Table 1 Calculating the Guy’s score and G-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical feature or syndrome</td>
</tr>
<tr>
<td>Complete paralysis of worst limb</td>
</tr>
<tr>
<td>(Medical Research Council scale 0 or 1)</td>
</tr>
<tr>
<td>Hemiplegia + hemianopia + higher cerebral dysfunction*</td>
</tr>
<tr>
<td>Drowsy or comatose after 24 hours</td>
</tr>
<tr>
<td>Loss of consciousness at onset</td>
</tr>
<tr>
<td>Age (y)</td>
</tr>
<tr>
<td>76 + scores 3</td>
</tr>
<tr>
<td>Uncomplicated hemiparesis</td>
</tr>
<tr>
<td>Constant</td>
</tr>
</tbody>
</table>

*Higher cerebral dysfunction means dysphasia, perceptual, or cognitive impairment.
The aims of this study were to validate the Guy's score in a new sample of stroke patients of all ages and to devise and test a simpler version. Also we intended to provide data that would enable clinicians to adjust their prognosis for prior probability and to select appropriate thresholds for dividing patients into good and bad prognostic groups.

Patients and methods

Stroke was defined in the same terms as used in the Oxford Community Stroke Project, after the World Health Organisation definition "rapidly developing clinical symptoms and/or signs of focal, and at times global (applied to patients in deep coma and to those with subarachnoid haemorrhage) loss of cerebral function, with symptoms lasting more than 24 hours or leading to death, with no apparent other than that of vascular origin".

A consecutive series of patients with acute stroke admitted to the medical and geriatric wards of the acute hospitals in two adjacent health districts in East London was recruited over 12 months in 1990–1. In all cases, the Guy's prognostic score at 24 hours was recorded as well as age and demographic data. At six months, the survivors received a follow up questionnaire by post that included the Barthel score. We have established the reliability of the postal questionnaire in a sample of these patients.

Allen divided his patients into those with "good" outcomes (functionally independent in activities of general living with or without aids but able to walk unaided) and "bad" outcomes (unable to walk independently, or dead). For this study, a good outcome was defined as a Barthel score of 13 and above, which equates with functional independence. Less than this or death was defined as a bad outcome.

By reducing all the weights to 1 and dividing age into three categories (table 1), we devised a simplified version of the Guy's score, which we have called the "G-score" (to distinguish it from the original). This gives a score out of 7 with higher scores meaning greater severity.

We examined the sensitivity, specificity, and likelihood ratio (sensitivity/1 − specificity) for predicting a bad outcome of the Guy's score and G-score at different thresholds. The likelihood ratio is a useful indication of the validity of a test because it is independent of the prior probability of a bad outcome and if the prior probability is known, Bayes' theorem (posterior odds = prior odds × likelihood ratio) can be used to estimate posterior probabilities.

We used the analysis of receiver operating characteristic (ROC) curves to estimate the optimal cut off thresholds for predicting a bad outcome of the G-score with varying cost-benefit ratios of false and correct predictions. This is a graphical method for depicting the trade off between the true positive rate and the false positive rate of a test. The ROC curve is a plot of the true positive rate (sensitivity) against the false positive rate (1 − specificity). Each point represents the test results for a different threshold. The "best" threshold for a particular patient, taking into account the prior probability of a bad outcome and the importance of avoiding incorrect predictions is the one where the slope of the curve satisfies the relation:

\[ \text{Slope of ROC curve} = \frac{C}{B} \times \frac{1 - p[D]}{p[D]} \]

where \( p[D] = \) the prior probability of a bad outcome, \( C = \) the cost of a false bad prediction; \( B = \) the benefit of a correct bad prediction.

We also examined the sensitivity, specificity, and likelihood ratio of different G-scores for predicting a bad outcome and we derived posterior probabilities for varying prior probabilities. From the standard formula for the standard error of a proportion, we estimated 95% confidence intervals for the observed proportion with a bad outcome for each G-score.

Results

Three hundred and sixty one patients were recruited of whom 42 (12%) were lost to follow up (26 could not be traced, seven left the country, six refused to participate, three said they were unable to participate). Data were incomplete for a further five patients. Thus 314 patients of mean age 73 years were studied. After six months, 124 (39%) had died and 92 (29%) were disabled (Barthel score <13) totalling 216 patients (69%) with a bad outcome.

The conscious level (drowsy or comatose at 24 hours) alone had a sensitivity for predicting a bad outcome of 0·47, with a specificity of 0·73, giving a likelihood ratio of 1·74. Tables 2 and 3 show how the scores performed with different thresholds to predict a bad outcome. The ROC curve (see figure) enables formal selection of the most useful G-score threshold, taking into account the

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>1·00</td>
<td>0·01</td>
<td>1·01</td>
</tr>
<tr>
<td>20</td>
<td>0·97</td>
<td>0·10</td>
<td>1·08</td>
</tr>
<tr>
<td>10</td>
<td>0·85</td>
<td>0·42</td>
<td>1·46</td>
</tr>
<tr>
<td>0</td>
<td>0·72</td>
<td>0·63</td>
<td>1·97</td>
</tr>
<tr>
<td>-10</td>
<td>0·57</td>
<td>0·83</td>
<td>3·31</td>
</tr>
<tr>
<td>-20</td>
<td>0·34</td>
<td>0·95</td>
<td>6·62</td>
</tr>
<tr>
<td>-30</td>
<td>0·19</td>
<td>0·98</td>
<td>9·53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Threshold</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Likelihood ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0·95</td>
<td>0·14</td>
<td>1·11</td>
</tr>
<tr>
<td>2</td>
<td>0·82</td>
<td>0·45</td>
<td>1·49</td>
</tr>
<tr>
<td>3</td>
<td>0·72</td>
<td>0·63</td>
<td>1·95</td>
</tr>
<tr>
<td>4</td>
<td>0·55</td>
<td>0·86</td>
<td>3·86</td>
</tr>
<tr>
<td>5</td>
<td>0·33</td>
<td>0·96</td>
<td>8·05</td>
</tr>
</tbody>
</table>
estimated costs of a false prediction and the benefits of a correct prediction. With the prior probability of 0·69 found in this study, and assuming for illustrative purposes the cost of a false prediction to be five times greater than the benefit of a correct prediction, the best threshold corresponds to the point where the slope is 2:2, which is 4. If the costs and benefits are equal, the corresponding threshold would be where the slope is 0·45, which is 2.

Table 4 shows how the different G-scores relate to the outcome found, the confidence intervals for the percentage achieving a good outcome for each score, and likelihood ratios for different G-scores. The likelihood ratios are different from those in table 3 because they refer specifically to the group of patients with that particular score, not those above that threshold.

The prior probability of a good or bad outcome is an estimate derived from simply knowing that the patient has had a stroke. From observational studies in stroke patients in hospital, estimates for a bad long term outcome vary from 36%5 to 66%.12 Additional information about severity of the stroke obtained from clinical history, examination, and investigation all have the potential for improving the accuracy of predictions of outcome. The likelihood ratio (LR) is a measure of this potential value of a piece of information. For example, if the prior probability of a bad outcome is 0·70 (odds = probability/1−probability = 2:33) and the patient was unconscious at 24 hours (LR = 1·74), this information improves prediction (prior odds × LR = posterior odds, 2·33 × 1·74 = 4·05), with an estimate of an 80% chance of a bad outcome. Prior probabilities will vary according to the type of patients seen (for example, referral centres will have worse prior probabilities).

Table 5 shows how different G-scores give different predictions when prior probability is changed. A score of 6 is strongly predictive of a bad outcome over a wide range of prior probabilities.

Discussion
In this study, the sensitivity, specificity, and likelihood ratio of the Guy's score and G-score are adequate but not as good as in the original study. This is not surprising as the original discriminant function was derived from the same patients used to test its validity. The validity of the G-score is, however, as good as that of the more complicated Guy's score.

Neither the items nor the weights used in the G-score were derived from multivariate analysis of the data in this study. The items of the G-score are the same as those that make up the Guy's score and the weights were arrived at arbitrarily to make the score easy to use. The estimates of sensitivity and specificity are therefore likely to be generally applicable.

The items used have the advantage of being usually recorded in most hospital records of patients with stroke, so that the score may be extracted from the notes.

Optimum cut off thresholds for scores can be chosen from ROC curves for groups of patients with differing prior probabilities and cost-benefit ratios. The examples we have described made use of the prior probability in this group of patients, but it would be possible to use the methods employed here to choose thresholds for groups with the wide range of prior probabilities reported.12 In this article we used arbitrary cost-benefit ratios to illustrate the method of choosing a threshold, but the cost of a false prediction and benefit of a correct prediction can be formally arrived at by utility theory.13

The G-score can be used to estimate 95% confidence intervals for the prognosis of patients with five grades of stroke severity. This could be very useful in patient communication, in guiding clinical decisions, in prospective research in rehabilitation, or in case mix analysis. Where prior probabilities are known to be different (for example, in a specialist centre), the prediction should be altered to take account of this by means of table 5.

The Glasgow coma scale,14 the APACHE II,15 and the Medicare mortality prediction...
Predicting stroke outcome: Guy’s prognostic score in practice

Address for correspondence: Sir Donald J. Robertson, MD, FICNeurop, FRCP, Professor, Department of Neurology, Guy’s Hospital, London SE1 9RT, United Kingdom. (Telephone: 01-219-7800 Ext. 1270; fax: 01-219-7794; e-mail: donald.robertson@ Guy’s NHS Trust.UK)

Systematic review of the literature

Predicting stroke outcome: Guy’s prognostic score in practice


Advice about prognosis, and use it as a means of adjusting for differences in case mix when making comparisons.

Predicting stroke outcome: Guy's prognostic score in practice.

P Gompertz, P Pound and S Ebrahim

*J Neurol Neurosurg Psychiatry* 1994 57: 932-935
doi: 10.1136/jnnp.57.8.932

Updated information and services can be found at:
http://jnnp.bmj.com/content/57/8/932

**Email alerting service**

These include:
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Topic Collections**

Articles on similar topics can be found in the following collections

*Stroke* (1449)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/