Variations Between Countries in Outcome After Stroke in the International Stroke Trial (IST)

N.U. Weir, MBChB; P.A.G. Sandercock, DM; S.C. Lewis, PhD; D.F. Signorini, PhD; C.P. Warlow, MD; on behalf of the IST Collaborative Group

Background and Purpose—This study describes the large variations in outcome after stroke between countries that participated in the International Stroke Trial and seeks to define whether they could be explained by variations in case mix or by other factors.

Methods—We analyzed data from the 15 116 patients recruited in Argentina, Australia, Italy, the Netherlands, Norway, Poland, Sweden, Switzerland, and the United Kingdom. We compared crude case fatality and the proportion of patients dead or dependent at 6 months; we used logistic regression to adjust for age, sex, atrial fibrillation, systolic blood pressure, level of consciousness, and number of neurological deficits. We used the frequency of prerandomization head CT scan and prescription of aspirin at discharge to indicate quality of care.

Results—The differences in outcome (all treatment groups combined) between the “best” and “worst” countries were very large for death (171 cases per 1000 patients) and for death or dependency (375 cases per 1000 patients). The differences were somewhat smaller after adjustment for case mix (160 and 311 cases per 1000 patients, respectively). Process of care may have accounted for some but not all of the residual variation in outcome.

Conclusions—Adjustment for case mix explained only some of the variation in outcome between countries. The residual differences in outcome were too large to be explained by variations in care and most likely reflect differences in unmeasured baseline factors. These findings demonstrate the need to achieve balance of treatment and control within each country in multinational randomized controlled stroke trials and the need for caution in the interpretation of nonrandomized comparisons of outcome after stroke between countries. (Stroke. 2001;32:1370-1377.)

Key Words: case fatality rate ■ cerebrovascular disorders ■ disability evaluation ■ outcome ■ randomized controlled trials

Whether a patient who has had an acute stroke dies, survives in a disabled state, or recovers completely is determined by many factors. Outcome after stroke varies from country to country, and recent studies have sought to explain the variation.1–5 Some of the variation in outcome may be due to differences between countries in the process of stroke care.3,4,6–10 The International Stroke Trial (IST) was a multinational randomized controlled trial that investigated the impact of aspirin, heparin, both, or neither on the outcome of patients admitted to hospital with an acute ischemic stroke. Preliminary analyses (with all treatment groups combined) of the nearly 20 000 patients entered revealed that among the 37 participating countries, 6-month case fatality ranged from 0% to 33%, and the proportion dead or dependent ranged from 32% to 89%.11 In itself, this observation was perhaps not surprising, given the inevitable differences between countries in the baseline characteristics (case mix) of the patients entered. The IST used a randomization procedure that anticipated this possibility, and treatment allocation was prospectively balanced within countries to prevent any between-country differences in case mix from biasing the assessment of the effects of treatment. In this report we describe the large variations in outcome between countries in the IST in detail, and we use the trial data set to determine to what extent these differences might be explained by adjustment for important differences in case mix and to what extent other factors, particularly international differences in the treatment of acute stroke, might account for any variation in outcome that remains.

Subjects and Methods

Exclusion of Data From Countries Recruiting <500 Patients

When comparing patients from different countries, we wished to avoid the extreme variability in outcome that can occur with small sample sizes.12 We therefore arbitrarily chose to examine data from only those countries that had randomized >500 patients each. We combined data from the treatment and control patients in each

Received January 15, 2001; final revision received March 2, 2001; accepted March 5, 2001.

From the Department of Neurology, Royal Hallamshire Hospital, Sheffield (N.U.W.), and Department of Clinical Neurosciences, Western General Hospital, Edinburgh (P.A.G.S., S.C.L., D.F.S., C.P.W.), UK.

Correspondence to Professor P.A.G. Sandercock, Department of Clinical Neurosciences, Western General Hospital, Crewe Rd, Edinburgh, EH4 2XU, Scotland. E-mail pags@skull.dcn.ed.ac.uk

© 2001 American Heart Association, Inc.

Stroke is available at http://www.strokeaha.org

1370
country for all analyses, and our findings therefore do not describe variation between countries in the efficacy of aspirin or heparin in patients with acute stroke.

Baseline Characteristics and Measures of the Process of Stroke Care
We extracted the following potential prognostic variables from the baseline data set: age, sex, delay from symptom onset to randomization, level of consciousness, presence of atrial fibrillation, systolic blood pressure on admission, presence of any infarction on CT scan, and presence or absence of 8 different neurological deficits. We also extracted data on 2 potential markers of the quality of the process of care: whether patients had had a CT scan before randomization and whether patients who were discharged alive from the participating hospitals were reported to be prescribed long-term aspirin.

Crude and Adjusted Outcome at 6 Months
We calculated the proportion who died from any cause (case fatality) and the proportion who were dead or needed help from another person for activities of daily living (death or dependency) at 6 months in each country (observed number). We used 2 previously described logistic regression models to adjust these data for important differences in case mix.13 The prognostic models were constructed and validated on 2 separate subsets of the IST data set and take the following variables into account: age, sex, systolic blood pressure, atrial fibrillation, level of consciousness, and total number of neurological deficits (see Appendix). We used the models to predict the probability of an outcome event for each patient and then calculated the total number of predicted events for each country as the sum of these probabilities (predicted number).

We expressed the adjusted outcome as a w score, a method that measures the difference in the number of observed and predicted events per 1000 patients treated within each country.14 15 We calculated the w score using the formula w=1000(o−p)/t, where o is the observed number of events, p the predicted number of events, and t the total number of patients per country. For example, if 500 patients are treated and a total of 100 deaths are predicted but 150 deaths are observed, then the w score is 1000×(150−100)/500=+100, that is, 100 more deaths than predicted per 1000 patients treated. However, we were principally interested in the difference in adjusted outcome between countries and calculated these simply by subtracting w scores. For example, if the w score for case fatality in country A was +50 and that in country B was −60, we estimated the absolute difference in adjusted case fatality between countries to be 90 deaths per 1000 patients treated. To directly compare outcomes before and after adjustment for case mix, we also calculated crude w scores by defining p in each country as tp, where p, is the proportion of patients that experienced the outcome in the entire study population (all countries combined).

Correction for Multiple Comparisons
Because the study necessarily involved multiple comparisons between countries, we calculated 99% CIs for the w scores (according to Parry et al14) to reduce the possibility that any observed differences might be due to chance.

Relationship Between Measures of the Process of Care and Adjusted Outcome After Stroke
We ranked the countries by the 2 adjusted outcomes and by the 2 measured items of the process of stroke care. Better outcome and better process of care are represented by lower numbers, and the converse is also true. We investigated the relationship between rankings with Spearman’s rank correlation coefficient.

Predictive Properties of the Prognostic Models
We determined the predictive properties of our prognostic models by estimating their calibration and discrimination.16 Calibration refers to the degree of bias of model predictions for groups of patients and can be estimated by plotting a calibration curve. To derive the calibration curves, we ordered the data set by ascending predictions of risk and then divided it into 10 equal groups (deciles). For each decile, we plotted the mean observed risk against the mean predicted risk. A model is well calibrated if, within each decile, the proportion of patients predicted to have an event and the proportion observed to have done so is the same, ie, if the calibration curve follows a 45° line. Discrimination refers to the ability of the model to differentiate between individuals who do and do not experience an event and may be estimated by calculating the area under the receiver operating characteristic (ROC) curve [a plot of the sensitivity against (1−specificity) of the model predictions]. Thus, a model that predicts case fatality with an area under the ROC curve of, for example, 80% will, in 80% of cases, correctly assign a higher risk of death to a randomly selected patient with a fatal outcome than to a randomly selected survivor.

Analyses were performed with the use of SAS (version 6.12).

Results
Patients Included in the Analysis
The IST recruited 19 435 patients from 37 countries. Nine countries recruited >500 patients: Argentina, Australia, Italy, the Netherlands, Norway, Poland, Sweden, Switzerland, and the United Kingdom. These countries contributed a total of 15 116 patients (78% of the trial accrual); of these, only 59 (0.4%) were lost to 6-month follow-up. The greatest losses to follow-up occurred in Argentina and the Netherlands (5% and 2%, respectively); in all other countries, outcome data were available in >99% of cases.

Baseline Characteristics and the Process of Care
The proportion of patients with each prognostic variable at baseline varied highly significantly between countries (all P<0.0001, χ2 tests) (Table 1), confirming our expectation that there would be variation between countries in the types of stroke patients considered eligible for the trial. Table 1 also shows substantial variation between countries in the process of stroke care, as judged by the proportion of patients who had a CT scan before randomization (range, 42% to 98%) and the proportion of patients prescribed aspirin for long-term secondary prevention at hospital discharge (range, 53% to 73%).

Variations in Outcome at 6 Months (Observed and Predicted)
Table 2 shows the variation between countries in the proportion of patients who had died by 6 months (range, 12% to 30%) and in the proportion who were dead or dependent at 6 months (range, 42% to 79%). The proportion of patients predicted dead and the proportion predicted dead or dependent can be taken to summarize the case mix of each cohort, and, not surprisingly, these also varied between countries (Table 2).

Crude and Adjusted Case Fatality at 6 Months
Figure 1 shows the w scores for case fatality for each country before and after adjustment for case mix. The figure clearly illustrates the substantial variation in crude case fatality between countries, the most extreme difference being between Sweden and Poland (171 more deaths per 1000 patients treated in Poland). After adjustment for case mix, some of the differences in case fatality between pairs of countries altered considerably. For example, adjustment for case mix reduced...
the differences in case fatality between Switzerland and Sweden and between the United Kingdom and the Netherlands by approximately 70%. However, the most striking finding shown in Figure 1 is that, on the whole, adjustment for important differences in case mix had little influence on the variation in case fatality between countries, which remained very substantial. For example, at 6 months, the difference between Sweden and Italy was 69 deaths per 1000 patients and between the United Kingdom and Poland was 57 deaths per 1000 patients; at its most extreme, 160 more patients were dead at 6 months per 1000 treated in Poland than in Sweden despite adjustment for case mix.

Crude and Adjusted Death or Dependency at 6 Months

Figure 2 shows the w scores for death or dependency at 6 months before and after adjustment for case mix. As with case fatality, the crude number of patients dead or dependent at 6 months varied considerably between countries, with the largest difference between Sweden and the United Kingdom (375 more patients were dead or dependent per 1000 treated in the United Kingdom). Adjustment for case mix markedly reduced some of the between-country differences, for example, that between Switzerland and Italy was reduced by 86% and that between Argentina and Australia was reduced by 60%. Again, however, despite adjustment for case mix, the overall variation in death or dependency between countries remained very substantial. For example, at 6 months, the difference between Sweden and Poland was 86 patients dead or dependent per 1000 treated and between Switzerland and the United Kingdom was 146 patients dead or dependent per 1000 treated; at its most extreme, 311 more patients were dead or dependent

### TABLE 1. Characteristics at Baseline of the 15 116 Patients Randomized in the IST From the 9 Countries That Recruited >500 Patients and Their Process of Care in Hospital

<table>
<thead>
<tr>
<th>Country</th>
<th>n</th>
<th>Male (%)</th>
<th>Age ≥70 y</th>
<th>Systolic BP ≥160 mm Hg</th>
<th>Fully Conscious</th>
<th>AF*</th>
<th>Enter Systolic BP ≥160 mm Hg Fully Conscious</th>
<th>Enter Systolic BP ≥160 mm Hg Fully Conscious Within 24 h of Onset</th>
<th>Visible Infarct on CT Scan</th>
<th>Discharge on Aspirin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>545</td>
<td>310 (60)</td>
<td>267 (51)</td>
<td>280 (54)</td>
<td>411 (79)</td>
<td>58 (11)</td>
<td>234 (45)</td>
<td>235 (45)</td>
<td>165 (18)</td>
<td>399 (77)</td>
</tr>
<tr>
<td>Australia</td>
<td>597</td>
<td>355 (60)</td>
<td>352 (59)</td>
<td>288 (48)</td>
<td>438 (74)</td>
<td>67 (12)</td>
<td>296 (50)</td>
<td>368 (62)</td>
<td>232 (45)</td>
<td>516 (87)</td>
</tr>
<tr>
<td>Italy</td>
<td>3437</td>
<td>1932 (56)</td>
<td>2190 (64)</td>
<td>1891 (55)</td>
<td>2747 (80)</td>
<td>598 (19)</td>
<td>1598 (46)</td>
<td>2242 (65)</td>
<td>1045 (43)</td>
<td>2421 (70)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>728</td>
<td>371 (52)</td>
<td>394 (55)</td>
<td>455 (64)</td>
<td>574 (81)</td>
<td>73 (10)</td>
<td>361 (51)</td>
<td>505 (71)</td>
<td>224 (36)</td>
<td>615 (86)</td>
</tr>
<tr>
<td>Norway</td>
<td>526</td>
<td>281 (53)</td>
<td>352 (67)</td>
<td>316 (60)</td>
<td>425 (81)</td>
<td>33 (6)</td>
<td>285 (54)</td>
<td>235 (45)</td>
<td>222 (45)</td>
<td>492 (94)</td>
</tr>
<tr>
<td>Poland</td>
<td>759</td>
<td>385 (51)</td>
<td>382 (51)</td>
<td>432 (57)</td>
<td>540 (72)</td>
<td>189 (25)</td>
<td>425 (56)</td>
<td>474 (63)</td>
<td>213 (56)</td>
<td>383 (51)</td>
</tr>
<tr>
<td>Sweden</td>
<td>636</td>
<td>344 (55)</td>
<td>473 (75)</td>
<td>407 (65)</td>
<td>532 (84)</td>
<td>95 (15)</td>
<td>260 (41)</td>
<td>276 (44)</td>
<td>288 (47)</td>
<td>617 (98)</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1631</td>
<td>842 (52)</td>
<td>1130 (69)</td>
<td>971 (60)</td>
<td>1129 (69)</td>
<td>263 (16)</td>
<td>988 (61)</td>
<td>1119 (69)</td>
<td>540 (45)</td>
<td>1207 (74)</td>
</tr>
<tr>
<td>UK</td>
<td>6257</td>
<td>3223 (52)</td>
<td>4229 (68)</td>
<td>3192 (51)</td>
<td>4610 (74)</td>
<td>1202 (21)</td>
<td>3594 (57)</td>
<td>3380 (54)</td>
<td>1681 (64)</td>
<td>2607 (42)</td>
</tr>
<tr>
<td>Total</td>
<td>15 116</td>
<td>8043 (53)</td>
<td>9769 (65)</td>
<td>8232 (55)</td>
<td>11 406 (76)</td>
<td>2578 (18)</td>
<td>8041 (53)</td>
<td>8634 (59)</td>
<td>4610 (50)</td>
<td>9257 (61)</td>
</tr>
</tbody>
</table>

BP indicates blood pressure.

*The presence of atrial fibrillation (AF) was not recorded in the 822 patients randomized from those countries in the pilot phase of the trial.

†Patients with a visible infarct are expressed as a percentage of those with a prerandomization CT scan.

‡Expressed as a percentage of cases that were discharged alive. Data describing discharge on long-term aspirin not collected in 292 cases.

### TABLE 2. Variation in the Observed and Predicted Number of Patients Who Were Dead or Were Dead or Dependent at 6 Months Between Countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Patients Dead at 6 Months, No. (%)</th>
<th>Patients Dead or Dependent at 6 Months, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>Observed</td>
<td>Predicted*</td>
</tr>
<tr>
<td>Australia</td>
<td>117 (23)</td>
<td>95 (18)</td>
</tr>
<tr>
<td>Italy</td>
<td>86 (15)</td>
<td>133 (23)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>664 (20)</td>
<td>744 (22)</td>
</tr>
<tr>
<td>Norway</td>
<td>128 (18)</td>
<td>141 (20)</td>
</tr>
<tr>
<td>Poland</td>
<td>69 (13)</td>
<td>114 (22)</td>
</tr>
<tr>
<td>Sweden</td>
<td>223 (30)</td>
<td>172 (23)</td>
</tr>
<tr>
<td>Switzerland</td>
<td>78 (12)</td>
<td>136 (22)</td>
</tr>
<tr>
<td>UK</td>
<td>366 (22)</td>
<td>464 (28)</td>
</tr>
<tr>
<td>Total</td>
<td>1711 (27)</td>
<td>1643 (26)</td>
</tr>
</tbody>
</table>

*We derived the total number of predicted events per country by summing the individual predicted probabilities generated by our prognostic models.
per 1000 treated in the United Kingdom than in Sweden even after adjustment for case mix.

Association of the Process of Care With Adjusted Outcome

The ranking of countries by adjusted outcome and by the proportion of patients that received each item of care is shown in Table 3. Higher adjusted case fatality was strongly correlated with a lower rate of CT scanning before randomization ($r = 0.78$) and also with a lower rate of prescription of long-term aspirin ($r = 0.53$). The correlation between adjusted death or dependency and the 2 processes of care variables was weaker ($r = 0.40$ in both cases).

Performance of the Predictive Models

Figure 3 shows that both prognostic models were well calibrated (both calibration plots follow a 45° line). Both models also showed moderately good discrimination with an area under the ROC curve of 0.79 in each case. For both outcome states, however, a substantial proportion of the predicted probabilities (33% for death, 57% for death or dependency) was outside the range of the observed data.
dependency) lay between 25% and 75%, i.e., the models failed to place substantial numbers of individuals into groups either very likely or very unlikely to have an outcome event.

**Discussion**

The principal aim of this study was to determine the extent to which the large differences in crude outcome between countries participating in the IST might be explained by differences in the baseline characteristics of the patients entered into the trial. The answer to this question is clearly complex. For certain pairs of countries, it appears that the differences in crude outcome could be explained very largely by differences in case mix, particularly by differences in age, sex, proportion with atrial fibrillation, and baseline stroke severity (measured by level of consciousness, number of neurological deficits, and systolic blood pressure) of the patients entered into the trial. For example, most of the large differences in case fatality between Sweden and Switzerland and in death or dependency between Switzerland and Italy were explicable in this way. However, for the majority of countries, adjustment for the factors in our prognostic models led only to modest changes in outcome, and, as a result, the overall variation in case fatality and in death or dependency between countries remained substantial. This finding was quite a surprise. We had anticipated that, with a prospective design, the availability of detailed and important prognostic data, virtually complete follow-up, and the use of a validated prognostic model, we would have been able to account for most of the overall variation in case fatality and in death or dependency between countries.

Given the well-described international differences in the treatment of acute stroke and speculation that these differences may contribute to differences in stroke outcome between countries, we considered whether differences in care between countries in the IST might explain their residual differences in outcome. Our observation that lower case fatality was strongly correlated with a higher proportion of patients with a head CT scan and with a higher proportion prescribed aspirin on discharge lends some support to this possibility. Similarly, it is notable that the Scandinavian countries were early advocates of the organization and specialization of stroke care and, at the time of the IST, the majority of Swedish and Norwegian patients were likely to have been treated in a stroke unit. This is in sharp contrast to the United Kingdom, Argentina, and Poland, the countries with the highest case fatality, where the majority of patients would have received conventional ward care. Such differences in care may well underlie at least some of the difference in outcome between the 2 groups of countries.

**Table 3. Rankings of Countries According to Outcome Adjusted for Case Mix and by Provision of the 2 Items of Process of Care**

<table>
<thead>
<tr>
<th>Country</th>
<th>Case Fatality at 6 Months</th>
<th>Death or Dependency at 6 Months</th>
<th>CT Before Randomization</th>
<th>Discharged on Long-Term Aspirin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Norway</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Australia</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Switzerland</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Italy</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Netherlands</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>UK</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Argentina</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Poland</td>
<td>9</td>
<td>2</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

Better outcomes and better process of care are assigned lower numbers.

**Figure 3. Calibration plots for the models predicting case fatality and death or dependency at 6 months.**
On the whole, however, the evidence from these data that differences in care might explain some of the residual differences in outcome is weak. Stroke interventions, especially those relating to secondary prevention, may quite properly be withheld from patients who survive in a very poor functional state. Our simplistic measurements of the quality of care do not take this into account and portray all cases in which the intervention is withheld as an error. Our finding that higher case fatality is strongly correlated with apparently worse provision of care was therefore perhaps inevitable. The weaker correlation between death or dependency and our measures of the process of care also argues against a significant impact of quality of care on the differences in outcome between countries. In particular, if a difference in the proportion of patients treated on a stroke unit was an important reason for the difference in outcome, how is one to reconcile Norway’s apparently excellent “performance” as measured by case fatality but its comparatively poor performance when measured by the combined outcome of death or dependency? Similarly, how does one explain the opposite findings for Argentina and Poland? Perhaps, one might argue, where there is better care, patients otherwise destined to die are more likely to survive but do so in a dependent state (eg, Norway), and where there is inferior care, patients with a poor prognosis die and therefore cannot be counted as dependent (eg, Argentina and Poland). However, if this were so, how would one explain the ranking of Sweden as the country with the best performance on both measures of outcome or the considerably worse performance of the United Kingdom when measured by death or dependency than when measured by case fatality alone?

A stronger argument that differences in stroke care are not the major cause of the residual differences in outcome is the sheer size of the absolute differences between countries. The differences in the proportion of patients dead or dependent between the United Kingdom and the other 8 countries were between 150 and 300 events per 1000 patients treated. These absolute differences in outcome are 2 to 4 times larger than the treatment effect of stroke unit care18 and twice as large as the benefit of giving thrombolysis within 3 hours of the onset of stroke.19 The differences in outcome are therefore much larger than might plausibly be explained by the differential use of even the most efficacious of known interventions. Indeed, at the time of the IST, thrombolysis was not in routine use. Thus, while it remains plausible that differences in medical treatment may account for some of the residual differences in outcome, other factors are likely to account for much more.

As with all observational studies, these other factors are chance, bias, and residual variation in case mix (confounding). Given our use of large samples and the narrow 99% CIs, chance is highly unlikely to explain much of the residual variation in outcome between countries. Similarly, given the nearly complete follow-up and the unambiguous nature of death, biased measurement cannot explain the residual variation in case fatality. However, measurement error may underlie some of the residual variation in death or dependency. Functional status is difficult to define and measure, and it is known that international comparisons are prone to bias.20,21 The fact that the range of differences between countries was greater for death or dependency than for death alone suggests that such bias might have operated here. In non–English-speaking countries, the follow-up questions were translated into the local language without back-translation to check for alterations, and subtle but important differences in interpretation may have been introduced.22,23 The method used to collect outcome varied between countries. In some it was by postal questionnaire, in others by telephone, and in others still by a combination of the 2 methods. These differences might also have influenced response. Cultural differences in the perception of disability and dependence may also have led patients in different countries to report dependency differently despite similar degrees of impairment in function.21,22,24,25 In general, however, the impact of these second order biases is modest. They would also be unlikely to explain the marked difference in death or dependency between patients in the United Kingdom and Australia, countries that used exactly the same outcome questionnaire and method of follow-up and that experience reasonably similar cultures. Furthermore, empirical research suggests that people in Sweden and the United Kingdom, the countries with the greatest difference in the proportions dead or dependent, value health states very similarly.26 It seems likely, therefore, that most of the residual variation in outcome between countries in the IST must be due to unmeasured variation in case mix.

The reason for the marked variation in case mix between countries in the IST is that the fundamental entry criterion for the trial was simply that the clinician had to be substantially uncertain whether or not to treat a given patient with aspirin, heparin, both, or neither. Variation between countries in “uncertainty,” in the types of physicians participating in the trial, and in the types of stroke patients routinely admitted to hospital9 are therefore all likely to have played a part. Although detailed, our prognostic models might have better accounted for the marked differences in case mix if they had adjusted for other recognized prognostic variables, such as prestroke functional status and living arrangements, comorbid conditions such as heart failure and diabetes mellitus, post-stroke urinary incontinence, hyperglycemia, and the size of the stroke lesion on brain imaging.27 However, they certainly could not have accounted for the (probably many) important differences in case mix that are not currently understood. The limitations of our models are illustrated by the fact that for large numbers of patients the predictions of risk are in the middle, nonconfident range, ie, they are not particularly good at separating patients into high- and low-risk groups. This in turn implies that they explain only a small part of the total variability in outcome (analogous to having a low r² statistic in linear regression). This is the case despite the fact that both models show excellent calibration and moderately good discrimination, 2 widely quoted measures of model performance. This observation highlights the important difference between a prognostic model having a good fit and providing clinically useful predictions.28

Regardless of their explanation, our findings have a number of implications. First, they emphasize the potential limitations of adjusting nonrandomized comparisons of
stroke outcomes for differences in case mix, even with high-quality and complete data from large numbers of patients and especially when the differences in unadjusted outcome are very large. These observations may be of particular relevance to those attempting to draw inferences about the quality of care from nonrandomized comparisons of stroke outcome, whether between hospitals, regions, or countries, and especially if their data are retrospective, incomplete, or less completely adjusted than our data. Second, investigators need to be aware that simply because case mix adjustment has been performed with the use of models that show excellent calibration and moderately good discrimination, a considerable amount of variation in outcome may remain to be explained. Third, therefore, before conclusions about the quality of stroke care are drawn, the plausibility of ascribing any residual differences in outcome to variation in the use of currently understood stroke interventions should be carefully considered. Fourth, those wishing to measure between-country differences in functional outcome after stroke should recognize that this task is prone to various subtle forms of bias.

Finally, our findings also have implications for the design of multinational randomized controlled trials. The variations in outcome between groups of patients in different countries in the IST do not affect the validity of the overall trial results because the trial used a method of allocation that ensured balance within countries (minimization). This design enabled the trial to detect the effect of a treatment (a reduction in death or dependency of 10 cases per 1000 treated) that was 30 times smaller than the largest difference in outcome between countries. If, however, a multinational study did not ensure balanced allocation within countries, then treatment effects might be spuriously generated or obscured. For instance, in an imaginary trial of a truly ineffective treatment using the same study population as the IST, if by chance the proportion of patients randomized to drug or to placebo happened to be 2:1 in Sweden and, for the same number of patients, 1:2 in the United Kingdom, then an apparent benefit of the truly ineffective drug would be observed even though the trial would have allocated equal numbers to each intervention.

In summary, this study demonstrates the potential limitations of using analyses of observational data to explain international differences in outcome after stroke. Differences in the quality of care, chance, measurement error, and cultural bias may account for some of the residual variation in outcome between countries in the IST, but most of the unexplained variation is likely to reflect the difficulty of achieving perfect case mix adjustment. To avoid these biases, multinational randomized controlled trials in stroke must ensure a balance of treatment and control within each country as well as in the overall trial. Furthermore, those wishing to draw inferences from nonrandomized comparisons of outcome after stroke should consider the issues raised in this report carefully.

Appendix

Statistical Calculations
We calculated the predicted risk of an outcome event \( P \) for each patient using the logistic regression models, as follows:

\[
P = e^{y/(1 + e^y)},
\]

where \( y \) is the linear predictor of the model and \( e \) is the exponential constant. We derived the linear predictor \( y \) of each model as follows.

Case Fatality at 6 Months
\[
y = -7.3529 + (0.0603 \times \text{age}) + (0.1637 \times \text{sex}) + (0.5130 \times \text{AF}) + (0.9533 \times \text{level of consciousness}) + (0.3272 \times \text{number of neurological deficits}),
\]

where \( \text{AF} \) is atrial fibrillation.

Dead or Dependent at 6 Months
\[
y = -1.5288 + (1.059 \times \text{age}) + (0.2988 \times \text{sex}) + (0.3066 \times \text{sex}) + (0.1963 \times \text{AF}) + (1.0634 \times \text{level of consciousness}) + (0.1012 \times \text{SBP}) + (0.2291 \times \text{SBP})^2 + (0.4249 \times \text{number of neurological deficits}),
\]

where \( \text{SBP} \) is systolic blood pressure.

Coding of Variables
\[
\text{Age} = (\text{age} - 70)/20; \quad \text{sex} = (\text{sex variable})^2
\]

Sex: female = 1, male = 0
Level of consciousness: drowsy/unconscious = 1, fully conscious = 0
\( \text{SBP}^2 = (\text{SBP} - 160)/60; \quad \text{SBP}^2 = (\text{SBP}^2 \times 2)/100 \)
Number of neurological deficits = observed number except 0 coded as 1 and 8 coded as 7 (covariate treated as a continuous variable because it has a linear relationship with each outcome)

Acknowledgments

The IST was supported by grants from the Medical Research Council, the Stroke Association, and the European Commission. Dr. Weir was supported by a Wellcome Trust Research Training Fellowship. We are grateful to the trial collaborators and to the patients who participated in the IST.

References


Variations Between Countries in Outcome After Stroke in the International Stroke Trial (IST)

on behalf of the IST Collaborative Group

Stroke. 2001;32:1370-1377
doi: 10.1161/01.STR.32.6.1370

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2001 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/32/6/1370

An erratum has been published regarding this article. Please see the attached page for:
/content/33/4/1171.full.pdf

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/
The article “Variations Between Countries in outcome After Stroke in the International Stroke Trial (IST)” by Weir et al. 1 contained an error in the Appendix. In the model that predicts death at 6 months, age should be entered unaltered and not in a standardized format, as is suggested. The standardized format should be used only in the model that predicts death or dependency at 6 months.

In “Development of a Novel, Weighted, Quantifiable Stroke Scale: Japan Stroke Scale” by Gotoh et al.2 the categories of the motor system on the Japan Stroke Scale (Figure 2, item 10) were incorrectly assigned letters (A, B, C, D, and E). They should be indicated with numerals (1, 2, 3, 4, and 5). In addition, the constant of −14.71 should be inserted at the bottom of the scale so that readers can calculate the total score. The authors regret these oversights. The corrected Japan Stroke Scale appears below:

1. Level of Consciousness (scored on Glasgow Coma Scale)

<table>
<thead>
<tr>
<th>Eyes Open:</th>
<th>Best Verbal Response:</th>
<th>Best Motor Response:</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 spontaneous</td>
<td>5 oriented</td>
<td>6 obey commands</td>
</tr>
<tr>
<td>3 to speech</td>
<td>4 confused</td>
<td>5 can localize pain</td>
</tr>
<tr>
<td>2 to pain</td>
<td>3 inappropriate</td>
<td>4 flexion withdrawal</td>
</tr>
<tr>
<td>1 none</td>
<td>2 incomprehensible</td>
<td>3 flexion to pain</td>
</tr>
<tr>
<td></td>
<td>1 none</td>
<td>2 extension to pain</td>
</tr>
</tbody>
</table>

E: 15 B: 14 - 7 C: 6 - 3

2. Language

1 verbal command for the patient to make a fist on the healthy side
2 have the patient name an object such as a “watch”
3 have the patient repeat familiar words such as “cherry blossoms”
4 have the patient give his/her address and name (Yes, No)

A: 4/4 B: 3/4 or 2/4 C: 1/4 or 0/4

3. Neglect

A: patient biceps line in middle
B: patient usually biased line towards “healthy” side of body
C: anosognosia or denial of body part

A: 0.42 B: 0.85 C: 1.27

4. Visual Loss or Hemianopsia

A: no visual loss
B: hemianopsia present

A: 0.84 B: 1.68 C: 2.33

5. Gaze Palsy

A: normal
B: gaze preference or difficulty with far lateral gaze
C: gaze palsy or persistent deviation

A: 0.31 B: 0.62 C: 0.93

6. Pupillary Abnormality

A: both reactive
B: one reactive
C: neither reactive

A: 0.31 B: 0.62 C: 0.93

7. Facial Palsy

A: normal
B: asymmetry on forced grimace
C: asymmetry or drooping at rest

A: 0.08 B: 0.15 C: 0.23

8. Plantar Reflex

A: normal
B: equivocal
C: positive (Babinski) and/or Chaddock’s sign

A: 0.08 B: 0.15 C: 0.23

9. Sensory System

A: normal
B: partial loss (mild sensory abnormality)
C: strong loss

A: 0.08 B: 0.15 C: 0.23

10. Motor System

<table>
<thead>
<tr>
<th>Hand:</th>
<th>A: normal</th>
<th>B: can make a circle with the thumb and 4th finger of the affected side</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A: 0.33</td>
<td>B: 0.66</td>
</tr>
<tr>
<td></td>
<td>B: 1.31</td>
<td>C: 0.99</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Arm:</th>
<th>A: normal</th>
<th>B: can make a straight arm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A: 0.66</td>
<td>B: 0.21</td>
</tr>
<tr>
<td></td>
<td>B: 1.97</td>
<td>C: 0.97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Leg:</th>
<th>A: normal</th>
<th>B: can raise a straight leg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A: 1.15</td>
<td>B: 2.31</td>
</tr>
<tr>
<td></td>
<td>B: 3.46</td>
<td>C: 0.97</td>
</tr>
</tbody>
</table>

6. Pupillary Abnormality

A: both reactive
B: one reactive
C: neither reactive

A: 0.31 B: 0.62 C: 0.93

The Japan Stroke Scale.
In “Association Between Influenza Vaccination and Reduced Risk of Brain Infarction” by Lavallée et al, the incorrect hospital was listed as the affiliation for Drs Lavallée, Perchaud, Grabli, and Amarenco. The correct affiliation is as follows: the Department of Neurology and the Formation de Recherche en Neurologie Vasculaire (Association Claude Bernard), Lariboisiere Hospital, Denis Diderot University, Paris, France.

In “Ultrastructural Changes of Neuronal Mitochondria After Transient and Permanent Cerebral Ischemia” by Solenski et al, the name of Gregory A. Helm, MD, was spelled incorrectly. In addition, the clinical affiliation for Charles G. diPierro, MD, was listed incorrectly. Dr diPierro is affiliated with the Department of Neurosurgery, Rikshospitalet/The National Hospital, University of Oslo, Norway.