Observer variability in assessing impaired consciousness and coma

GRAHAM TEASDALE, ROBIN KNILL-JONES AND JAAP VAN DER SANDE

From the Departments of Neurosurgery and Medicine, Glasgow University and the Institute of Neurological Sciences, Southern General Hospital, Glasgow

SUMMARY Head-injured patients were examined by a number of observers whose assessments were compared. Considerable discrepancies occurred when overall “levels” of consciousness and coma were used, and also with some terms which are in common use. More consistent assessments were obtained by employing the “Glasgow Coma Scale,” which describes eye opening, verbal behaviour, and motor responsiveness. Nurses and general surgeons were as consistent as neurosurgeons when using this scale, and it was relatively resistant to language or cultural differences between observers. The practical reliability of the Glasgow scale enhances its value, both for monitoring individual cases and for making meaningful comparisons between series of patients with acute brain damage.

A wide variety of disorders may be complicated by impaired consciousness and coma. Assessments of the degree of impairment may be required for several purposes: to compare series of patients managed in different ways; to determine whether an individual patient requires or justifies a particular type of management; to provide a guide to ultimate outcome. Perhaps their widest application is in monitoring progress after head injury or intracranial surgery, in order to detect complications. Many different systems for observing and describing impairment of consciousness and coma have been proposed, a reflection of the importance ascribed to this activity in clinical practice. Whether employed to characterise series of cases or to monitor individual patients, a system should be reliable and consistent when used by different observers, who may have widely varying degrees of experience.

A practical scale for assessing impaired consciousness and coma has been developed in Glasgow (Teasdale and Jennett, 1974); it is based upon eye opening, and verbal and motor responsiveness. It has already proved useful in comparing a series of head injuries from different countries (Jennett et al., 1977), and it can be combined with other clinical data to predict outcome both after head injury (Jennett et al., 1976; Teasdale and Jennett, 1976), and also “non-traumatic” coma (Bates et al., 1977). As a practical tool for routine clinical monitoring it has been widely accepted (Plum, 1975; Verjaal and Van’t Hooft, 1975; Langfitt, 1978). Its reliability in formal observer error tests has been only briefly reported before (Teasdale and Jennett, 1976), and we now report results of detailed studies of the observer variability associated with this scale and with certain alternative terms often used to describe patients with acute brain damage.

Material and methods

The study was performed in three separate phases, each of which followed a similar pattern. Participating observers independently recorded certain features which they observed in a series of patients in a neurosurgical intensive care unit. The records of their observations were tabulated, and the variability within a particular set was determined by a method developed for this study. The variability rates associated with different observations, and also the performances of different groups of observers, were compared.

First of all we studied some frequently used terms and at the same time carried out a preliminary study of the elements of the Glasgow

Address for reprint requests: Graham Teasdale, University Department of Neurosurgery, Institute of Neurological Sciences, Southern General Hospital, Glasgow G51 4TF, Scotland.

Accepted 7 February 1977
Coma Scale. The second phase of the study consisted of a more detailed examination of the separate components of the scale. In the third part a specially prepared film of several patients' motor responses was scored by groups in Britain, Europe, and North America, in order to assess the reliability of the terms used in the motor component of the Glasgow Coma Scale as compared with commonly used alternatives. In the first two parts of the study the observers consisted of doctors of varying degrees of experience and also nurses.

STATISTICAL METHODS

Inter-observer disagreement rate

Inconsistent observations occur when two or more observers report different findings after examining the same patient. The frequency with which observers are at variance with each other is a measure of the lack of reproducibility of the particular observation under test. We have expressed this by calculating, for each feature observed, an inter-observer disagreement rate.

Disagreement rate using a scale of two alternatives

If observers are given the choice of allocating patients to one of only two classes, the disagreement rate is expressed by the proportion of observers taking a minority view. For example, if 10 observers examine a patient and eight allocate him to one class, the consensus is that the patient is in that class; the two observers taking the minority view constitute a 2/10 (0.2) disagreement rate.

Disagreement rate using scales of multiple choices

To compare the consistency of scales in which there are three or more possible states, an analysis which takes account of differences in the number of possible states and allows comparison with simple, two-choice scales is necessary. The analysis which follows was suggested by Mr Derek Teather, and has the advantage of being applicable even when some observations are missing.

Let the ordinal numbers associated with the extreme states of a feature, described by a multi-component scale, be noted by 1 and Top, and let N observers examine the patient and record the value of their observations (0 to n). The consensus state, C, is the mode of the N observations.

The disagreement rate, DR, for one patient is then given (all differences taken as positive) by:

\[
DR = \frac{\sum_{1}^{N} |O - C|}{N \times \max (C - 1, Top - C)}
\]

The principle of the calculation is shown in the following example, where 10 observers have assigned a patient's response to one of five classes, denoted by scores 1-5:

<table>
<thead>
<tr>
<th>Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The consensus, scored by seven observers, is class 3 and the disagreement rate:

\[
DR = \frac{(0 \times 1) + (1 \times 1) + (3 \times 0) + (1 \times 1) + (1 \times 2)}{10 \times \max (3 - 1, 5 - 3)}
\]

\[
= \frac{4}{20} = 0.2
\]

The disagreement rate has a range of 0 to 0.5. The aggregate disagreement rate for a group of observers, for several different observations, is obtained by summing the top and bottom parts of the equation before a final division.

Non-parametric statistical tests have been used (Siegel, 1956).

Investigation 1

There were 27 observers—eight nurses, six house officers, and 15 registrars or consultants in neurosurgery.

Sixteen patients were studied; on any day only two patients were examined and each was seen by all observers within a three hour period. Observers first wrote a brief description of the patient, and then, on a standard pro forma, recorded the presence or absence of various "levels" of impaired consciousness and of certain motor abnormalities (Table 1). On a second pro forma, seven of the neurosurgeons recorded additional assessments of certain ocular features and of the three responses of the Glasgow Coma Scale—eye
opening, verbal, and motor responsiveness. No general discussion about the terms being tested took place nor, at this stage, did we provide standard definitions.

**RESULTS**

**Observer's own description**

Preliminary interviews with participants revealed widely varying views about description and classification of impaired consciousness. This confusion was reflected by the terms they applied to the patients in the study, showing considerable inconsistencies and ambiguities. The same patient could be described as "somnolent," "difficult to arouse," or "deeply comatose" by different observers.

**Commonly used terms (Table 1)**

There was considerable variability in assessing whether a patient was "conscious" or "unconscious." The disagreement rate observed (0.191) indicates that almost one out of five observers disagreed with the remainder. This undoubtedly reflects the ambiguity of these terms; definition of a clear distinction between them has long eluded both physicians and philosophers. Some improvement in consistency was achieved by using an intermediate grade of "semi-conscious," but variability was again high when four categories of impaired consciousness and coma were employed.

Disagreement about the four motor abnormalities was less than for overall judgments of impairment of consciousness. Disagreement rates for the presence or absence of individual abnormalities, however, do not reflect the discrepancies resulting from using these terms as part of a graded scale of responsiveness. By constructing a hierarchical motor abnormality scale, with flaccidity the most abnormal response and absence of any specific abnormality scoring highest, we obtained an estimate of this variability and found a relatively high disagreement rate (0.176).

**Ocular reflexes, motor, and verbal responsiveness**

(Table 2)

Assessing pupil responses to light was the most reliable of all observations studied. In contrast, there was more than three times as much disagreement about assessments of pupil equality. The difference probably reflects the seven neurosurgeons having to make "value" judgments when assessing minor differences in pupil size.

Judged by their consensus scores, only four of the 16 patients blinked, only two orientated on their surroundings, and only three patients gave verbal responses. There was considerably less disagreement about the state of the remaining, negatively scored patients than about these few patients giving positive responses. As a result the aggregate disagreement rates for this whole set of patients probably underestimates the variability inherent in assessments of these features. This reservation does not apply to the disagreement rates for eye opening and motor responses for which the various positive responses were well represented among the patients we studied.

Disagreement about limb responses in this part of the study (0.109) was distinctly less than the disagreement among the seven neurosurgeons when their observations were analysed by the hierarchical motor abnormality scale shown in Table 1.

**Investigation 2**

Six nurses, seven neurosurgeons, and five general surgical trainees acted as observers. In contrast to the first investigation, participants were given standard definitions for each observation being assessed. The definitions were derived from the published description of the Glasgow Coma Scale (Teasdale and Jennett, 1974).

 Twelve patients were studied. As in investigation 1, only two patients were examined on any single day. However, observations were recorded on each patient on two successive days so that altogether 24 sets of patient data were available. The pro forma used to record observations contained six sections (Table 3), the motor responses of the arms being recorded in four subsections to facilitate comparison of different methods of stimulation and assessment.

---

**Table 2. Variability among seven neurosurgeons in assessment of "ocular features," verbal, and motor responses in 16 patients**

<table>
<thead>
<tr>
<th>Observations assessed</th>
<th>Categories</th>
<th>Disagreement rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pupils</td>
<td>Equal/unequal&lt;br&gt; Reacting/non-reacting</td>
<td>0.153&lt;br&gt; 0.043</td>
</tr>
<tr>
<td>Blinking when threatened</td>
<td>Present/absent</td>
<td>0.105</td>
</tr>
<tr>
<td>Orientation to environment</td>
<td>Present/absent</td>
<td>0.091</td>
</tr>
<tr>
<td>Eye opening to stimulation</td>
<td>(1) Not at all; (2) to pain; (3) to sound; (4) spontaneous&lt;br&gt; (5) Not at all; (6) to sound; (7) to pain; (8) spontaneous</td>
<td>0.143&lt;br&gt; 0.109</td>
</tr>
<tr>
<td>Speech</td>
<td>(1) Absent; (2) incomprehensible; (3) confused; (4) lucid and orientated</td>
<td>0.054</td>
</tr>
<tr>
<td>Response of most normal limb</td>
<td>(1) Flaccid; (2) extension; (3) abnormal flexion; (4) withdrawal; (5) localising; (6) obeying</td>
<td>0.109</td>
</tr>
</tbody>
</table>
Disagreement about the "best motor response" in this phase of the study was less than about the first limb in the first phase, but the difference was not significant. Because of the difficulty in distinguishing types of flexion response for routine clinical purposes no distinction is made between them. Disagreement is further reduced (Table 4) and the resulting rate for the "best motor response" is significantly lower than for either eye opening (P<0.001 Wilcoxon test) or verbal responsiveness (P<0.02 Wilcoxon test).

It is useful, sometimes, to obtain a numerical index of the overall severity of a patient's condition. This can be achieved by allocating numbers to the individual components of the scale, more normal responses scoring higher. Although the assumption that the various components and steps on the scale are of equal magnitude may not be entirely valid, in practice summing the three components to a total score yields useful information. The observers' records were analysed according to this notation (Table 4). Disagreement about the resulting total scores was only 0.076, half the level of variability found with overall assessments of impaired consciousness in the first investigation.

Nurses and general surgeons were at least as consistent as neurosurgeons, and there were no significant differences between the various groups of observers (Table 5) either in respect of any single observation or when observations were combined and overall rates calculated.

**Table 3** Variability among 18 observers in assessments of eye opening, verbal, and motor responses in 12 patients on two successive days

<table>
<thead>
<tr>
<th>Observation assessed</th>
<th>Categories</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye opening</td>
<td>(1) Absent; (2) to pain; (3) to sound; (4) spontaneous</td>
<td>0.092</td>
<td>0.086</td>
<td>0.089</td>
</tr>
<tr>
<td>Verbal response</td>
<td>(1) Nil; (2) incomprehensible; (3) inappropriately; (4) confused; (5) oriented</td>
<td>0.088</td>
<td>0.094</td>
<td>0.091</td>
</tr>
<tr>
<td>Obeying commands</td>
<td>Yes/no</td>
<td>0.073</td>
<td>0.089</td>
<td>0.081</td>
</tr>
<tr>
<td>Response to supraorbital stimulation</td>
<td>(1) Nil; (2) extension; (3) abnormal flexion; (4) normal flexion; (5) localising</td>
<td>0.198</td>
<td>0.186</td>
<td>0.191</td>
</tr>
<tr>
<td>Response to fingertip stimulation</td>
<td>1-4 as previous category</td>
<td>0.129</td>
<td>0.146</td>
<td>0.138</td>
</tr>
<tr>
<td>Weakness of either arm</td>
<td>Right/left/neither</td>
<td>0.135</td>
<td>0.187</td>
<td>0.160</td>
</tr>
</tbody>
</table>

**Table 4** Variability among 18 observers in assessment of best motor response and coma scale total in 12 patients on two successive days

<table>
<thead>
<tr>
<th>Observation assessed</th>
<th>Categories</th>
<th>Day 1</th>
<th>Day 2</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best motor response to any stimulus</td>
<td>(1) Nil; (2) extension; (3) abnormal flexion; (4) normal flexion; (5) localising; (6) obeying</td>
<td>0.075</td>
<td>0.087</td>
<td>0.081</td>
</tr>
<tr>
<td>Best motor response (reduced scale)</td>
<td>(1) and (2) as above; (3) flexion; (4) localising; (5) obeying</td>
<td>0.046</td>
<td>0.064</td>
<td>0.055</td>
</tr>
<tr>
<td>Coma scale total range 3-14</td>
<td>Eye opening (1-4) + Motor response (1-5) + Verbal response (1-5)</td>
<td>0.079</td>
<td>0.073</td>
<td>0.076</td>
</tr>
</tbody>
</table>

**RESULTS**

There were no significant differences between the disagreement rates obtained on the two days. The significance of differences between the observations were tested by using day 1 and day 2 and different groups of observers as a repetition, and applying the Kruskal-Wallis test ($x^2$, (3 df) = 23.2, P<0.001).

Disagreement about eye opening was considerably less than in the first study (0.089 versus 0.143), consistency probably being improved because standard definitions were supplied to the second group of observers. Disagreement about verbal responses occurred at similar rates in the two investigations. In contrast to the first investigation, however, the majority of the second group of patients gave some verbal response, and this difference may have masked any beneficial effect resulting from providing definitions.

Observations of motor responses elicited by a supraorbital painful stimulus were more liable to disagreement than responses to fingertip stimulation (P<0.01 Wilcoxon test). The former was also a less effective type of stimulus; four patients were considered by a majority of observers to be unresponsive to supraorbital stimulation, yet none failed to respond in some way to fingertip pain. Whatever the site of stimulation, confusion between abnormal and normal flexion was a major source of disagreement. When applying the Glasgow Coma Scale, observers are now advised to stimulate the fingertip initially but to record the best response obtained from either arm to any stimulus. Table 4 shows that, as a result, disagreement about motor responsiveness can be reduced.
The film showed the responses of 14 patients to painful stimulation. It was viewed in the Institute of Neurological Sciences, Glasgow, by nurses, junior doctors, general surgical trainees rotating through the unit, neurosurgeons, and visitors from abroad. For some of the latter and also for some of the permanent staff in the unit, their native tongue was not English. The film was also scored by groups of clinicians involved in the care of head injury patients in the Netherlands (the Academic hospitals of Groningen and Rotterdam), and in North America (Los Angeles County Hospital). These audiences provided a spectrum of observers of differing expertise, experience, and background, both with respect to training and language. Before recording their observations of the film, observers were given definitions of responses as in investigation 2. Responses to supraorbital and fingertip stimulation were recorded separately for each limb.

RESULTS
The results permit several interesting comparisons. The rates of disagreement among Glasgow observers (Table 6) were less than half those for the corresponding assessments of motor responses in the clinical investigations. Different patients were studied but the groups were closely comparable in the frequency with which patients showed the various types of motor response. The greater reproducibility of observations of the film reflects the elimination of variations in examination technique. While the film indicated that fingertip stimulation was again the most certain way of evoking a response, observations of responses to supraorbital stimulation were now more consistent, indicating that the disparity between them in the second investigation was due to observers having applied supraorbital stimulation with different degrees of vigour.

As was found in the clinical studies, counting only the best response shown by a patient and then making no distinction between varieties of flexor response, considerably reduced disagreement.

While there were no differences between the performances of the various professional groups who saw the film in Glasgow, there were differences between the disagreement rates of the doctors in the various centres.

The majority of Glasgow observers were accustomed to using the coma scale; this was not the case for the audiences in the Netherlands or North America. Indeed for most members of the latter two groups the projection of the film was their first exposure to the scale. The disagreement rates for the different groups show (Table 7) that the North Americans' lack of familiarity with the scale was accompanied by an increase in variability, and there was a further increase in the Dutch observers. Foreign graduates working in Glasgow, however, were at least as reliable as British observers. Assessments of the purposefulness of motor responses, in contrast, were less reliable among all groups, and were particularly influenced by an observer's native tongue. Further studies are in progress in Holland, using translations of the scale (Braakman et al., 1978).

Practical implications of observer variability
While the three investigations demonstrated that the Glasgow scale is more reliable than some

Table 6 Disagreement among observers in Glasgow about filmed motor responses of patients

<table>
<thead>
<tr>
<th>Stimulation site</th>
<th>Categories</th>
<th>Disagreement rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fingertip</td>
<td>(1) Nil; (2) extension; (3) abnormal flexion; (4) normal flexion</td>
<td>0.105</td>
</tr>
<tr>
<td>Supraorbital</td>
<td>1–4 as above; (5) localising</td>
<td>0.075</td>
</tr>
<tr>
<td>Best response</td>
<td>1–5 as above</td>
<td>0.078</td>
</tr>
<tr>
<td>either site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Best response</td>
<td>(1) Nil; (2) extension; (3) flexion; (4) localising</td>
<td>0.031</td>
</tr>
<tr>
<td>reduced scale</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7 Disagreement among groups of doctors in three countries about filmed motor responses of patients

<table>
<thead>
<tr>
<th>Observer groups</th>
<th>Disagreement rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Glasgow scale; (1) Nil; (2) extension; (3) flexion; (4) localising</td>
</tr>
<tr>
<td>Glasgow (1) English native tongue</td>
<td>0.031</td>
</tr>
<tr>
<td>Glasgow (2) native tongue not English</td>
<td>0.013</td>
</tr>
<tr>
<td>Los Angeles</td>
<td>0.064</td>
</tr>
<tr>
<td>Groningen</td>
<td>0.087</td>
</tr>
</tbody>
</table>
alternative systems, there is, as with any method of clinical assessment, a degree of variability inherent in its use.

For practical purposes we need to know the reliability of a recorded difference, whether between two patients or between two assessments of the same patient. Using data from the second clinical study, we determined the distribution of observations around the consensus for patients’ total coma scores. The distribution peaked sharply around the consensus, and Table 8 provides an estimate of the confidence with which a difference between two observers’ findings may be taken to indicate a “real” clinical difference rather than being a reflection of inter-observer variability. Although there is a chance of only one in the coma scale total may be due to inter-observer variability, if the change in either direction is greater than one, it is almost 90% certain that this reflects a “real” change or difference. Moreover, the more profoundly consciousness is impaired, the more reliable is an observed change, in this circumstance, the overall score comes to depend mainly upon motor responsiveness, the most consistent of the three components of the scale.

Table 8 Distribution of observations around consensus: 18 observers assessing coma scale total in 10 patients on successive days

<table>
<thead>
<tr>
<th>Number of observations</th>
<th>0</th>
<th>7</th>
<th>61</th>
<th>179</th>
<th>53</th>
<th>7</th>
<th>4</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>0</td>
<td>2.1</td>
<td>6.3</td>
<td>18.3</td>
<td>53.8</td>
<td>15.9</td>
<td>2.1</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Discussion

Assessment of patients with acute brain damage has been advanced considerably by techniques such as computerised axial tomography and continuous monitoring of intracranial pressure. Yet it is not practical to repeat any radiological investigation at frequent intervals, and intracranial pressure measurements are still applied to only a minority of the patients in a minority of centres; studies of evoked potentials have similar limitations at present. It will be necessary for some time to continue depending to a great extent upon repeated clinical assessments of “conscious level,” which, therefore, need to be made with as little variability as possible.

There is no generally accepted method for analysing the results of studies of observer variability. The method we employed has two essential advantages: it is not invalidated by missing data, an inevitable complication of a study using so many observers on different days, and it facilitates comparisons between scales with different numbers of components. Further studies, using alternative methods of analysis, are planned.

The reasoning which led to the incorporation of assessments of three separate aspects of behaviour into the Glasgow scale, rather than relying on one aspect or employing overall levels and grades, has been discussed previously (Teasdale and Jennett, 1974). The results of our studies show that assessments of patients’ responsiveness by the specific terms which were incorporated into the scale are more reliable than overall judgments about degrees of impairment of consciousness and coma. Similarly, Hall and his colleagues (1976) found that, while agreement about the presence or absence of specific symptoms after surgery for peptic ulcer was “excellent,” application of an overall system of grading results (Visick) gave unacceptable variation.

Observer variability, with rates of disagreement usually varying between 20% and 35%, has been documented in many aspects of clinical practice (Fletcher and Oldham, 1964; Conn et al., 1965). Neurological observations are at least as liable to variability as any other. McCance et al. (1968) assessed the reproducibility of plantar responses in a geriatric population where the chance of an extensor plantar response being found again at a second examination by either the same observer or another observer was only 50%. Sisk et al. (1970) examined discrepancies in duplicate histories and examinations of patients suspected of cerebral vascular disease. In spite of having formulated a manual of definitions, they found considerable inconsistencies in eliciting symptoms and concluded that the “probability of two neurologists eliciting positive history of the same symptoms is a random phenomenon.” Discrepancies in examination findings were also considerable but were more susceptible to improvement after redefinition and discussion. Price (1976) presented a limited study of the consistency of a considerably more complex method of assessing “conscious level” than the practical Glasgow scale.

There are two main sources of observer variability: variations in the response given by a patient, either because his condition is fluctuating or because he has been examined in a different way; and variations in interpretation of the same response by different observers. The results we obtained reflect the influence of both factors but also demonstrate how attention to either may improve reliability.
Observer variability in assessing impaired consciousness and coma

Elimination of variable interpretations when using scales composed of several, ordered components requires that patients' states fall naturally into corresponding subdivisions. In practice, sharp distinctions exist only rarely and the nearer a patient lies to the boundary line between two categories, the greater the difficulty in allocating him to either class; the more clearly defined the divisions, the less the need for observers to exercise judgment and the greater the reproduceability of their assessments. For many of the patients in these studies there was almost no disagreement between observers. A minority of patients, whose responses may, in reality, have lain near the boundaries between the different components of the scale, proved to be the major sources of variability. Disagreement about these cases might have been reduced by even more detailed definitions and discussions about the responses. While these could be appropriate in some circumstances, for example between the participants of a collaborative study, an increase in complexity of the scale might have hindered its general practical usefulness.

The more consistent observation of the film by doctors working in Glasgow than those in either North America or Holland indicates that experience in the use of the scale reduces variability. This is also suggested by the differences in variability between the first and second clinical phases of the study but we have not made a systematic study of the effect of training on disagreement rates. Moreover, we did not attempt to train observers to a high pitch of consistency before our studies as our concern was to achieve a realistic assessment of the reliability of observations made during the course of ordinary clinical practice.

Observer variability is likely to be most evident when, as in the present study, comparisons are made between single assessments by different observers. Four observers scored the film on two occasions, several weeks apart; discrepancies between an observer's first and second assessments (intra-observer variability) were considerably less than the rate of disagreement among the whole group of observers. In the clinical situation, when the same observer re-examines a patient, or when a series of consistent recordings have been obtained by different observers, the significance that can be attached to any change is enhanced.

Complete elimination of variability is an impossible ideal but, provided that sufficient care is taken in eliciting and interpreting responses, the precision of the Glasgow scale is adequate for most purposes. No alternative method has been shown to equal its capacity for reliable, practical use in a wide variety of circumstances.

Unrecognised discrepancies in methods of assessment may explain many of the conflicts and controversies about the results of therapy (Feinstein, 1967). Hall and his colleagues (1976) suggested that differences between alternative operations for peptic ulcer might be obscured by the imprecision of the Visick system for grading results and used the analogy of an adverse signal: noise ratio in electronics, the signal (the difference between operations) being lost in the noise (observer variability). Until a treatment with a strong enough "signal" emerges, the many controversies surrounding the management of brain damage from trauma or vascular disease may be resolved only by the use of sufficiently precise assessments of the severity of the initial damage. We believe that the Glasgow Coma Scale is a basis for such assessments.

We thank Mr Derek Teather, Senior Lecturer in Statistics, School of Mathematics, Computing and Statistics, Leicester Polytechnic, for valuable advice about the analysis of these data; also our many colleagues who willingly acted as observers. These studies were performed as a contribution to an International Collaborative Study of Coma Prognosis supported, at different times, by the Royal College of Physicians of London, the Nuffield Provincial Hospitals Trust, the British Council, and the National Institutes of Health of the USA, contract number 1-NS-4-2328.

References


Hall, R., Horrocks, J., Clamp, S. E., and de Dombal,


Observer variability in assessing impaired consciousness and coma.
G Teasdale, R Knill-Jones and J van der Sande

*J Neurol Neurosurg Psychiatry* 1978 41: 603-610
doi: 10.1136/jnnp.41.7.603