Short report

Interobserver agreement in assessment of vestibulo-ocular responses

J H VAN DEN BERGE,* R BRAAKMAN,* H J A SCHOUTEN†

From the Departments of Neurosurgery* and Biostatistics,† University Hospital Rotterdam "Dijkzigt", Erasmus University, Rotterdam, the Netherlands

SUMMARY In 30 comatose head injured patients vestibulo-ocular responses were elicited by caloric stimulation. Inter- and intra-observer agreement of these responses was studied. The responses were shown in the form of a film to ten doctors who classified them into four categories: no reaction, tonic reaction, paradoxical nystagmus and nystagmus. The level of the coefficient Kappa (a chance-corrected measure of interobserver agreement) was 0.50; this is within the range of levels for most components of clinical examination reported in the literature.

Severe injury to the head often results in brainstem dysfunction. Whether or not this is the case can be indicated by eye movement. The prognosis in patients with severe head injury depends on age, on degree and duration of impairment of consciousness and brainstem dysfunction. It is no surprise, therefore, that spontaneous and elicited eye movements have been shown to be of prognostic value in patients with severe head injury.1–7 The practical prognostic value of these symptoms depends on the consistency of assessment. This can be determined by studying both inter- and intra-observer agreement. The aim of the present study was to determine the inter- and intra-observer agreement in rating vestibulo-ocular responses.

Patients and methods

Patient population and ocular signs studied

Vestibulo-ocular responses were studied in 30 comatose head-injured patients over a period of 18 months, coma being defined as a state of depressed consciousness in which the patient does not open his eyes, does not obey commands and does not speak words, irrespective of stimulus. The series is not consecutive, because the main observer (JHB) was not always present and because the film camera was not always available.

The responses were studied on the third or fourth day after admission. In all patients, the ears were irrigated with 120 ml ice-water and the effect studied. The right ear was treated first, and after a 5–10 minute interval, the left ear was irrigated in the same way. This caloric examination was always performed by JHB and the results filmed. The 30 film fragments obtained were combined into one film, lasting 50 minutes. Each fragment was preceded by a number and a sign, indicating whether the right or left ear had been irrigated. The moment at which a painful stimulus was applied to the nailbed(s) of the patient was indicated by a flash. Only the eyes were shown on the film, not the rest of the face, which made it possible to include the effect of caloric stimulation twice in two patients: one patient with nystagmus confirmed on the ENG and one patient with a tonic reaction without nystagmus on the ENG.

Ten doctors with many years' experience in neurological examination were then asked to attend the film session and to classify the vestibulo-ocular response in these 30 patients (or in fact 32 as far as they were concerned) into four categories: (1) No reaction, (2) Tonic reaction, (3) Paradoxical nystagmus, (4) Nystagmus. In cases of doubt, the higher category was chosen. Nystagmus was defined as a rhythmic movement of the eyes, occurring at least three times, with a minimal frequency of one beat per second, in either a horizontal, vertical, diagonal or rotational direction.

Paradoxical nystagmus was defined as a nystagmus occurring or increasing as a result of the painful stimulus. All other, non-rhythmic, movements of the eye(s) were defined as tonic reaction. There was no verbal or other contact between the examiners and the authors while the film was being shown.

Results

Two patients were shown twice: one in fragments 6
Table  Percentages of (randomly selected) patients classified identically by two (randomly selected) doctors

<table>
<thead>
<tr>
<th>Categories chosen by one doctor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>23</td>
<td>3</td>
<td>7</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>28</td>
<td>38</td>
</tr>
<tr>
<td>Total</td>
<td>10</td>
<td>37</td>
<td>7</td>
<td>47</td>
<td>100</td>
</tr>
</tbody>
</table>

Kappa = 0.50 with standard error 0.06.

Categories of vestibulo-ocular reactions, from which selection was made: (1) no reaction, (2) tonic deviation, (3) paradoxical nystagmus, (4) nystagmus.

and 15, and the other in fragments 20 and 25.

Small differences in the categorisation occurred frequently but sometimes even conflicting scores were given. For instance, in cases 13, 22 and 24, some observers saw no reaction at all, while others saw nystagmus. In patient 20, all ten observers classified the response as a postcaloric nystagmus. When, however, 10 minutes later, in case 25, the same fragment was re-shown, complete agreement no longer existed, and two observers allocated the classifications tonic reaction and paradoxical nystagmus, respectively. In the two patients whose responses were shown twice, a different category was chosen six times. Fifty per cent of the observers chose the same category for both patients in the repeated presentation.

The statistical analysis methods proposed by Schouten8,9 were applied to compute the percentages presented in the table. The coefficient Kappa was computed. This is a chance-corrected measure of interobserver agreement and has been lucidly explained by Sackett et al.10 Let us suppose that a (randomly selected) patient is to be classified by two doctors who are chosen at random from the total group of ten doctors. Then, according to the table, there is an estimated probability of 38% that both doctors use category 4, and there is an estimated probability of 7% that the first doctor uses category 2 and the second doctor uses category 4.

The conditional percentages can be interpreted as follows. Once one knows that the first (randomly chosen) doctor used category 3, there is an estimated probability of 28% that the second (randomly chosen) doctor will also use category 3 for the same patient. Note that the interpretation of the results no longer depends on the total number of doctors involved in the study, which makes it easier to compare the results from different studies.

Discussion

Clinical signs are subject to different interpretation by various observers. The causes for inter- and intra-observer disagreement have been summarised in a previous study.11 Teasdale and Jennett12 were the first to refer to observer disagreement in signs in comatose patients. Braakman et al.13 demonstrated a satisfactory agreement in the assessment of the motor response of the Glasgow coma scale, which was confirmed by Teasdale et al.14 The latter authors also found a satisfactory agreement in the assessment of eye opening and verbal response in comparison with other features.

Van den Berge et al.15 established interobserver agreement for various ocular signs in comatose patients. The Kappa values for various signs were: anisocoria 0-61; pupil reactions to light 0-57; spontaneous eye movements 0-46; cephalo-ocular responses 0-49. The agreement in assessment of pupil reaction to light and (an-) isocoria of pupils appeared to be moderately satisfactory, but the consistency in the interpretation of spontaneous eye movements and cephalo-ocular responses, is disappointing. In our study, the degree of disagreement found amongst residents was no greater than that amongst senior doctors. In this study too the agreement in interpretation of vestibulo-ocular responses was only moderate (Kappa = 0·50; ± 0·06).

The Kappa level found is within the range of levels for most components of the clinical examination, also non-neurological examinations.16

References

8 Schouten HJA. Measuring pairwise interobserver agreement when all subjects are judged by the same observers. Statistica Neerlandica 1982;36:45–61.


Berge JH van den, Schouten HJA, Boomstra S, Drunen Littel S van, Braakman R. Interobserver agreement in assessment of ocular signs in coma. *J Neurol Neurosurg Psychiatry* 1979;42:1163–8.

