PSYCHOMOTOR RECOVERY AFTER OUTPATIENT ANAESTHESIA: INDIVIDUAL IMPAIRMENT MAY BE MASKED BY GROUP ANALYSIS

S. HICKEY, A. J. ASBURY AND K. MILLAR

SUMMARY

Recovery from outpatient anaesthesia with propofol was followed in 10 patients, using a semantic recognition memory test (SemRT) (in a new implementation on a Psion hand-held microcomputer), choice reaction time (CRT) and critical flicker fusion threshold (CFFT). Group analysis of results revealed an effect on psychomotor performance as measured by the SemRT and CFFT (but not the CRT), 30 min after the end of the procedure. Performance on all three tasks had returned to baseline values within 60 min of completing the anaesthetic. Group analysis, however, masks individual impairment which may be clinically important. There was no statistically significant correlation between post-anaesthetic task performance and age, sex, dose of propofol, anxiety or depression score.

KEY WORDS

Tasks of simple and choice reaction time are well established as sensitive methods of evaluating performance in the recovery period following anaesthesia and sedation [1]. However, such tests assess rather basic cognitive processes and leave uncertain whether higher order functions might suffer extended residual impairment. For instance, while choice reaction time (CRT) performance might return to normal within 1 h of termination of anaesthesia, it is conceivable that memory retrieval processes might remain relatively impaired. In other words, a more sensitive test might reveal an otherwise unanticipated impairment.

It is possible to present simultaneously both elements of the above tasks (speeded decision making and memory retrieval) by requiring patients to perform a task of semantic recognition memory (SemRT) [2, 3]. Simply, the task assesses the speed with which the individual can access knowledge held in long-term semantic memory about the meaning of words and their relationships in order to make a binary decision. Typically, the computer-generated task follows a routine whereby the patient is presented with a series of phrases, each denoting a category of objects or things (e.g. “four-legged animals”); after the presentation of each category phrase, a test word is presented and the subject is required simply to decide whether the word describes a member of that category: in the case of the example above, the test word dog would evoke a “yes” response while hat would evoke a “no” response. The subject responds via the computer keyboard which permits measurement of the time taken to make the decision. The decision time represents two processes: the time taken to retrieve relevant information about the test word from long-term memory and the time to decode and perform the appropriate motor response.

It is possible to examine the memory retrieval time in greater detail by taking advantage of the known structure of semantic memory storage in order to vary the difficulty of the retrieval operation. The human semantic memory consists of our knowledge of the meaning of words and their relationships. The great advantage of testing semantic memory is that the subject retrieves information which has been learned and stored over a lifetime, hence avoiding the need to learn material for retrieval within the immediate clinical

STEPHEN HICKEY, F.C.A.N.A.E.S. (Clinical Shock Study Group, Department of Anaesthesia); A. JOHN ASBURY, F.C.A.N.A.E.S., PH.D. (University Department of Anaesthesia); Western Infirmary, Dumbarton Road, Glasgow G11 6NT. KEITH MILLAR, PH.D., C.PSYCHOL., F.B.P.S.S., Behavioural Sciences Group, University of Glasgow, 4 Lilybank Gardens, Glasgow G12 8QQ. Accepted for Publication: October 15, 1990. Correspondence to A.J.A.
or experimental setting. Within semantic memory, words are conceived as being stored in categories within which the words are stacked in conceptual hierarchies. Thus very familiar and commonly encountered instances of the category (e.g. “four-legged animal”—“dog”) are conceived as being of “high dominance” (HD) in the category and are readily retrieved spontaneously as category instances and are recognized quickly as such. In contrast, “low dominance” (LD) words are less common instances (e.g. “ram” or “cheetah” in the context of the example above) and take longer to recognize or recall spontaneously.

The difference in time taken to recognize or recall HD compared with LD words provides a measure of retrieval time in memory uncontaminated by motor response processes. Thus by manipulating the factor of retrieval dominance, one can examine the influence of various agents upon the speed of retrieval from long-term memory.

Previous research has shown that the SemRT is sensitive to changes in alertness or arousal. For instance, the time taken to retrieve LD information gradually reduces as physiological activation increases with diurnal changes in alertness [4]. Loss of sleep results in a deterioration in the speed of accessing LD information [5], and time to retrieve HD words decreases as arousal increases with loud noise stimulation [6].

Given the sensitivity of semantic recognition to changes in arousal, it would seem important to consider if the task might prove useful in assessing impairment of function in the post-anaesthesia period. Despite the above positive findings, application of the task in previous studies relevant to anaesthesia have shown uncertain results. Ghoneim, Hinrichs and Mewaldt reported an overall slowing in semantic classification time following oral diazepam 0.1-0.3 mg kg⁻¹, but no selective effect as a function of retrieval dominance [7]. The lack of selective effect may reflect the authors’ choice of HD and LD words which did not differ sufficiently in their dominance values (norms are available which assign numerical dominance values to several thousand words for numerous semantic categories [8]). In other words, the difficulty of the retrieval task may have been slight and insufficient to show drug-induced impairment.

A further methodological difficulty may attach to the negative results of Moss and his co-workers who used an SemRT in a battery of tests to investigate the recovery times from anaesthesia induced with methohexital and maintained with either halothane or alfentanil [9]. While these researchers found no significant effect of each agent on retrieval from semantic memory, one should note that their patients were not tested until 2 h after the anaesthetic agents had been discontinued. It is therefore conceivable that testing occurred too late to detect initial impairment. The researchers deferred testing because their patients lacked the sensori–motor coordination to perform the task presented by BBC microcomputer.

In this paper we introduce a new method of measuring semantic classification time, using a Psion hand-held computer, and provide data on the learning curve with this system. We also present data from a clinical study where this implementation of the SemRT, and two more commonly used tests of psychomotor function, were used to follow the pattern of recovery from outpatient general anaesthesia in 10 patients.

SUBJECTS AND METHODS

Study of the SemRT in unsedated volunteers to determine the learning curve

Subjects. Ten volunteers participated in the study and each completed a Hospital Anxiety and Depression (HAD) questionnaire before performing the task. The questionnaire assessed seven common symptoms of anxiety and seven of depression. For each question the subject’s answer scored 0–3, with the higher score indicating increasing symptoms. The final total for both anxiety and depression score ranges from 0 to 21: a score of 7 or less is normal, 8-10 is a borderline score, and 11 or more suggests clinically relevant anxiety or depression [10]. This standardizes the SemRT data, as anxiety increases arousal and may influence performance [11].

Stimulus materials for the SemRT. Ten context categories were chosen from Battig and Montague’s list of category norms which define numerically the “dominance” (frequency in language) of a word when it occurs in a specific context [8]. Examples of these categories include “Is it a colour?”, “Has four legs?”, and “Is a vegetable?”. Associated with each category are 12 test words; six words should generate the answer “yes”, and six the answer “no”. Of the
words which generate the answer “yes”, three were chosen as HD test words, and three as LD test words, having mean dominance values of 1.7 and 29.4, respectively. The six words which generate the answer “no” are referred to as “non-category” (NC) test words. For example, in the category “Is it a colour?”, “red” is an HD member, “mauve” a LD member, and “spider” a NC member.

The difference in time taken to recall HD compared with LD words provides a measure of retrieval time in memory uncontaminated by motor response processes. Thus, by manipulating the factor of retrieval dominance, one may examine the influence of various agents upon the speed of retrieval from long-term memory.

Task implementation on Psion XP micro-computer. The SemRT programme required the user only to identify and press keys corresponding to “yes” and “no”, these being demonstrated by the experimenter in an earlier briefing session. As the Psion keys are small, the system is designed to allow several adjacent keys to respond to “yes” and “no” in order to accommodate subjects with large fingers. Before the SemRT itself, preliminary instructions scroll across the screen, prompted by key presses by the subject. The final message is “Are you ready?—press the “Y” key to go on.” At this point the subject positions his fingers over the Y and N keys.

After the subject presses the Y key, the first category is presented, chosen randomly from the memory. After 0.5 s, a test word appears, chosen randomly from the list of 12 words associated with the category. The subject responds to the test word as quickly and accurately as possible. The latency period is measured to an accuracy of 5 ms by the computer. A 0.5-s pause follows each response, during which the screen clears. The task continues until the subject has responded correctly to 10 HD and 10 LD test words. In this process he may respond to a variable number of NC words, usually about 20. The duration of the task depends, therefore, on the performance of the subject, but averages 2 min.

If any individual response time exceeds 3000 ms, the result is withdrawn automatically from the programme’s internal statistical analysis, and filed under a separate group called “blocks”. The “block” phenomenon was noticed during early trials of this system, where occasionally the subject developed inability to respond, and this required a short time to resolve. This phenomenon has been recorded commonly in serial reaction time performance since the initial observation by Bills in 1931 [12]. The number of blocks during a run is expressed as a percentage of the total number of correct and incorrect keystrokes. At the end of the task, the Psion displays the results, which are stored in a Rampak with the date and time. The information in the Rampak can be transferred subsequently to an IBM-compatible computer for manipulation by word-processing and spreadsheet programmes.

Design. During investigation of the learning curve, the volunteers were tested six times at 10-min intervals. To minimize auditory distraction, they listened to white noise through a set of headphones at a volume adjusted to a comfortable level.

Recovery from general anaesthesia

Subjects. We studied 10 patients (four male) attending for outpatient urological investigation under general anaesthesia. Ethics Committee approval and informed written consent were obtained. Any patient receiving regular sedatives or beta-blockers was excluded. Each patient completed an HAD score questionnaire before operation.

Stimulus material, and test design. In addition to performing the SemRT, each subject was studied using the Leeds psychomotor testing system. This compact portable device measures choice reaction time (CRT) and critical flicker fusion threshold (CFFT) [1].

For CRT, the subject viewed a test board containing a central touch-sensitive button and six other touch-sensitive buttons corresponding to six small lights placed equidistant from the central button. The lights were illuminated randomly, and the subject touched the appropriate button to extinguish the light as soon as it illuminated. Two time intervals are measured: the first from illumination of the light until the subject’s finger leaves the central button (the recognition time, CRT-R), and the second the total time from illumination of the light until it is extinguished by the subject. The difference between the two is the motor time (CRT-M), as the distance travelled in all cases is the same. The reaction time was taken as the mean of 20 responses.
### TABLE I. Characteristics of volunteers

<table>
<thead>
<tr>
<th>No.</th>
<th>Sex</th>
<th>Age (yr)</th>
<th>Time of day</th>
<th>HAD score</th>
<th>Occupation</th>
<th>Keyboard familiar?</th>
<th>HD mean (ms)</th>
<th>LD mean (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>48</td>
<td>16:00</td>
<td>A 6</td>
<td>Secretary</td>
<td>Yes</td>
<td>742</td>
<td>661</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>29</td>
<td>14:00</td>
<td>D 6</td>
<td>Immunologist</td>
<td>Yes</td>
<td>656</td>
<td>725</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>26</td>
<td>16:00</td>
<td>A 4</td>
<td>Nurse</td>
<td>No</td>
<td>506</td>
<td>551</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>29</td>
<td>09:30</td>
<td>A 6</td>
<td>Lab. Technician</td>
<td>Yes</td>
<td>521</td>
<td>690</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>31</td>
<td>14:40</td>
<td>D 6</td>
<td>Doctor</td>
<td>No</td>
<td>512</td>
<td>621</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>30</td>
<td>16:00</td>
<td>A 3</td>
<td>Nurse</td>
<td>No</td>
<td>621</td>
<td>650</td>
</tr>
<tr>
<td>7</td>
<td>M</td>
<td>30</td>
<td>12:00</td>
<td>A 12</td>
<td>Pipefitter</td>
<td>No</td>
<td>812</td>
<td>886</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>33</td>
<td>15:00</td>
<td>A 7</td>
<td>Lab. Technician</td>
<td>Yes</td>
<td>532</td>
<td>631</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>32</td>
<td>11:00</td>
<td>A 10</td>
<td>Doctor</td>
<td>Yes</td>
<td>557</td>
<td>555</td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>43</td>
<td>14:00</td>
<td>A 4</td>
<td>Medical Physicist</td>
<td>Yes</td>
<td>631</td>
<td>700</td>
</tr>
</tbody>
</table>

It has been suggested that the CFFT is a sensitive test of CNS arousal [1]. The subject viewed a set of four light-emitting diodes set in a square pattern, at a distance of 1 m under consistent light conditions. The frequency of flickering increased slowly until the subject was no longer able to distinguish between intermittent and constant light, at which point he pressed a button to record threshold. The task was repeated with the frequency decreasing until the subject first noticed intermittent light. The subject performed the task twice in each direction, and the mean frequency of the four runs was taken as the score.

**Method.** Each subject performed the SemRT and the CRT task four times before anaesthesia; the CFFT was measured once before anaesthesia. Each task was performed at 30, 60 and 90 min after the anaesthetic agents were discontinued. While performing the SemRT, each subject listened to white noise through headphones, in order to reduce distraction.

**Anaesthetic technique.** Each patient was fasted and ASA grade I or II. No premedicant drug was given. Anaesthesia (administered by a non-investigator) was induced with propofol until verbal contact was lost, and maintained with nitrous oxide in oxygen supplemented by enfurane \((n = 8)\) or halothane \((n = 2)\). The patients breathed spontaneously via a Bain system.

**Statistics**

The effect of anaesthesia on the psychometric tasks and the SemRT was examined using confidence intervals [13]. For each task the result immediately before anaesthesia was taken as the baseline. The median change from baseline, and the 95% confidence intervals (95% CI) around the median were calculated from the results of the post-anaesthesia tasks using the Wilcoxon one sample test. In order to examine the individual responses to anaesthesia as measured by these tasks, a summary statistic was used for each patient: an "area under the curve" (AUC, arbitrary units) was calculated incorporating change from baseline and time, with the time scale used being 0 (before anaesthesia) to 3 (90 min after anaesthesia).

### RESULTS

**Learning curve**

Characteristics of the volunteers are displayed in table I. The mean response times and 95% confidence intervals for HD, LD and NC words are displayed in figure 1. The learning curve can be seen clearly. Overall, it appears that practice was complete after four sessions.

An overall improvement in score was seen between the first and the second trials, with no further improvement in performance thereafter. Five of the volunteers experienced "blocks" during the first trial, with only one during the second trial, and none thereafter (table II).

**Recovery from general anaesthesia**

The mean age of the patients studied was 42 yr (range 31–58 yr) and mean weight was 73 kg (range 60–102 kg). The duration of anaesthesia was 5–15 min (mean 7 min), and total dose of propofol used was 1.47–3.08 mg kg\(^{-1}\) (mean 2.46 mg kg\(^{-1}\)).
OUTPATIENT ANAESTHESIA

FIG. 1. Mean (95% CI) changes in response time over six trials with semantic recognition memory test. ◊ = High dominance words; • = low dominance words; □ = neutral words.

FIG. 2. Median (95% CI) change in response time from baseline values (0) after outpatient anaesthesia, as measured by SemRT. ◊ = High dominance words; • = low dominance words; □ = neutral words.

FIG. 3. Median (95% CI) change in response time from baseline values (0) after outpatient anaesthesia, as measured by CRT task. • = CRT-Recognition; ◊ = CRT-motor.

**Group analysis**

**SemRT task.** In the initial 30-min period of recovery from anaesthesia, the response times to HD and LD words were increased by medians of 115 ms (95% CI = 5, 233) and 118 ms (95% CI = 29, 243), respectively, compared with baseline (pre-anaesthesia) performance (fig. 2). The fact that the CI do not encompass the baseline value of zero would imply significant (P < 0.05) impairment at this stage (see [13]). In the case of NC words, the median response time increase of 165 ms is associated with a CI which does encompass the baseline zero (95% CI = −8, 372).

The fact that the increase in median response time to HD and LD words was identical implies that the anaesthetic had no selective effect as a function of retrieval difficulty. By 60 min, response times to all three classes of words had returned to baseline.

**CRT task.** At each test period the 95% CI for change in CRT performance encompassed the baseline value of zero (fig. 3), therefore the slight increase in median response time in the 30–60 min post-anaesthesia period may be considered not significant.

**CFF task.** In the 30-min period following recovery, the median threshold value and its CI were beyond the baseline zero (fig. 4), implying significant depression of the threshold at this time, although it returned to baseline thereafter.

**Individual results**

There was considerable inter-individual variation in the performance of SemRT, CRT and

<table>
<thead>
<tr>
<th>Trial</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean score (%)</td>
<td>90.4</td>
<td>94.8</td>
<td>93.5</td>
<td>97.3</td>
<td>95.6</td>
<td>96.5</td>
</tr>
<tr>
<td>Range</td>
<td>71–98</td>
<td>86–100</td>
<td>74–100</td>
<td>88–100</td>
<td>87–100</td>
<td>89–100</td>
</tr>
<tr>
<td>No. of volunteers having &quot;block&quot;</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
CFFT tasks following anaesthesia (table III). We could find no statistically significant correlation between post-anaesthetic task performance and age, sex, dose of propofol, anxiety or depression score (Spearman rank correlation test).

**DISCUSSION**

Group analysis of psychomotor and recovery performance suggests that our anaesthetic technique (essentially, propofol) was associated with return of patient performance to pre-anaesthesia achievement within 60 min of discontinuing the anaesthetic. Examination of the 95% CI around the median change from baseline suggests that impaired performance was present 30 min after the anaesthetic for SemRT and CFFT, but not for the CRT task. Although the size of the group was small, this was accounted for in calculation of confidence intervals; the data might then suggest that the SemRT and CFFT are more sensitive indicators of post-anaesthetic performance impairment than is CRT testing.

Our results differ slightly from those of a similar study by MacKenzie and Grant, who investigated the postoperative recovery properties of propofol compared with thiopentone and methohexitone [14]. In their study, they detected no significant impairment of psychomotor function measured by CFFT and overall CRT at 30 min following anaesthesia with propofol. The CRT-M, however, was impaired at this time, but had returned to normal by 60 min. Inclusion of the SemRT in the investigation of recovery of higher order mental function in the present study would appear to confirm that this anaesthetic seems to be excellent for outpatients.

The hazards of presenting grouped results in this manner have been discussed recently by Matthews and his colleagues [15]. They observed that presentation of the results as a curve joining means or medians may not be a good descriptor of a typical curve for an individual. When each time point is analysed separately, no account is taken of the fact that the procedure ignores the way in which individuals respond with time: these two factors give a misleading impression of individual behaviour. We also feel that "conventional" analysis of the results may conceal important individual variations. Some of the problems in presenting inter-individual variability adequately in studies of this type have been highlighted by Millar [16]. When assessing "street fitness" after day case anaesthesia, the individual's worst performance is more important than that of a theoretical group. Table III highlights this problem, with the considerable inter-individual variation in performance being obvious from the

<table>
<thead>
<tr>
<th>Table III. Individual performances as expressed by the &quot;area under the curve&quot; (in arbitrary units) as defined in the text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient No.</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>
AUC values presented. While the grouped results for these tests show very little effect of the anaesthetic, recovery may be extremely unpredictable, and raises questions about the procedure for discharging patients following day-case anaesthesia (e.g. all patients leave after 90 min). We suggest that each patient requires individual assessment.

Investigation of the postoperative recovery period using psychometric tests helps the anaesthetist to develop a safe outpatient anaesthetic technique, assuming that performance in such tests equates with aptitude for real life skills. This is the subject of continuing debate [16–19]. It should be remembered also that, while impairment in psychometric testing suggests that the patient is not fully recovered, the converse cannot be assumed. We suggest that a further potential hazard may arise when a patient is assumed to be fully recovered simply because he was anaesthetized using a technique which is “known” to have no significant group effect. One postal survey revealed that, despite instructions to the contrary, 73 % of car owners drove within 24 h of outpatient anaesthesia, and 9 % drove themselves home alone [20]. Given this information, the desirability of ensuring complete recovery from drug effects is obvious.

Clearly no single test of psychomotor function can evaluate fully all the facets of patient recovery from anaesthesia and sedation [17]. We feel that the SemRT implemented in this way is a useful addition to the battery of tests already available. It is an indicator of state of arousal [4], providing us with a measure of reaction time and retrieval time, and therefore potentially allows us to quantify effects of drugs on basic and higher order cognitive functions.

In our implementation using the compact, hand-held Psion, we have shown that the task can be used earlier in the period of recovery from anaesthesia than had been expected [9]. The SemRT is learnt easily, with little improvement in response times after four practice runs. Further advantages of the SemRT in this implementation are the ability of the Psion to analyse and store data from the test period, presenting it in a neat and easy to use form, and the fact that, although each test only takes about 2 min to perform, up to 40 responses are obtained, making the data from this 2-min test more statistically viable than would normally be the case.

The implementation has some drawbacks: for example, the small screen and characters demand good eyesight, and the characters, as displayed by the Psion, do not have conventional descenders, which may lead to difficulty. However, none of the present patients or volunteers complained that this was a problem. The small keys may be difficult for some patients, and would certainly be unsuitable for any patient suffering from distal arthropathy. The learning curve demands trial runs before the patient receives any medication, and may delay the patient’s procedure. More experience with the system may provide answers to these problems.

REFERENCES


