Errors on anaesthetic record charts as a measure of anaesthetic performance during simulated critical incidents

A. J. BYRNE, A. J. SELLEN AND J. G. JONES

Summary
We have measured the performance of 10 trainee anaesthetists during a single simulated anaesthetic during which there was a complex critical incident. Errors in the recording on the anaesthetic charts of the "patient's" oxygen saturation, heart rate, systolic and diastolic arterial pressures and end-tidal carbon dioxide concentrations were used as a measure of mental workload and hence performance. The critical incident was designed to be stressful and contained, in sequence, episodes of hypotension, arrhythmia and bronchospasm. Chart recording errors increased markedly during the critical incident (P < 0.01) and decreased subsequently when the "patient" had stabilized. More than 22% of the values charted during the simulation were in error by more than 25% of the real value, and errors in excess of 100% of the actual value were recorded. There was no evidence of a tendency to consistently underestimate the magnitude of abnormal values. This method is appropriate for assessing the performance of groups of anaesthetists during simulated critical incidents. It also raises questions on the accuracy of anaesthetic record charts when recording critical incidents. (Br. J. Anaesth. 1998; 80: 58–62)

Keywords: anaesthesia, audit; model, computer simulation; computers, simulation; anaesthetist, performance

"Mental workload" is a term often used to describe the amount of mental effort involved in performing any given task. If we consider that there is a limit to the rate at which information can be processed by the human mind ("mental capacity"), then a simple model of mental workload is that it is a function of the proportion of this capacity in use at any time. The proportion of capacity in use varies depending on the difficulty or demands of the task. It is also generally accepted that as we acquire experience of a new task and develop procedural skills, the mental workload associated with that task decreases, because our capacity increases. For example, learner drivers need all their abilities just to keep the car on the road, while taxi drivers are able, with ease, to drive, eat, navigate and talk simultaneously. Studies in fields such as aviation have shown that high mental workload is associated with poor performance. Casali and Weirwille noted that, "If requirements are moderately excessive, there may be a measurable degradation in the performance of simple, supportive tasks associated with flying, such as routine communications or engine instrument monitoring. Often, more immediate tasks receive the focus of attention, and other tasks are time shared or even ignored completely. Mental overload may further result in significant pilot errors in aircraft control, possibly culminating in an accident". The inference from these two observations is that a low mental workload during the performance of a complex task may be associated with expertise while a high mental workload for the same task may be associated with inexperience and loss of control. If the mental workload of a subject performing a standardized task can be measured, then it can provide an assessment of performance.

Several techniques have been used to measure mental workload, many relying on the use of a secondary task. In such studies, the subject is asked to perform two tasks, a primary task (which is regarded as essential) and a secondary task (which has a lower priority). In this situation, the mental capacity of the subject must be shared between the two tasks. If a subject performs the primary task well, their mental workload is low, allowing spare capacity to be allocated to the secondary task. Conversely, if the subject performs the primary task badly, their mental workload is high and there is little spare capacity available for the secondary task. Thus performance on the secondary task can be used as an indication of how much spare capacity is available and is therefore a measure of the subject's performance in the primary task.

The problem with this type of research in anaesthesia is the difficulty in finding a secondary task which subjects repeat frequently enough to obtain good measures of performance, but which does not distort performance of the primary task (i.e. does not interfere with normal working patterns). The most common methods used in other fields have involved choice reaction time, memory, monitoring and tracking tasks. In anaesthesia, Gaba and Lee studied the workload of anaesthetists during real anaesthetic cases by asking them to add together two numbers on the screen of a computer every 45 s. They were asked to observe when the numbers...
changed and enter their answer into the computer via the keyboard. The study showed that in times of stress, the anaesthetists took longer to add the numbers or failed to respond at all. Unfortunately, the addition of a monitor and a keyboard to an already complex theatre in addition to the risk of distracting the anaesthetist from their patient makes the technique unsuitable for routine use in real theatres. Further, the use of such techniques detracts from the realism of simulated events.

Our purpose in this study was to examine a secondary task technique, completion of the anaesthetic record chart, which can be more realistically incorporated into the simulation of an anaesthetic critical incident. This also has the potential of being used during anaesthetics involving real patients. During previous studies with the ACCESS simulator we noticed that anaesthetists’ perceptions of events during simulated anaesthetics were often very distorted. In particular, when anaesthetists were asked to complete an anaesthetic chart after the simulation had finished, the charted values showed large errors and the time scale was overestimated by up to 100%. The errors we observed were similar to those detected during other studies involving measurement of mental workload. It was therefore decided to investigate if chart error could be used as a measure of mental workload and hence performance.

Methods

Ten trainee anaesthetists completed a single simulation of an anaesthetic using the anaesthetic computer-controlled emergency situation simulator (ACCESS), a full scale, low-cost anaesthesia training simulator. Each anaesthetist was introduced to the simulation as if taking over a routine case from a colleague. The “patient” was described as a 25-year-old female undergoing an anterior cruciate ligament repair under general anaesthesia. She was said to be asthmatic and to be recovering from an episode of viral myocarditis. When the simulation was started, the anaesthetist was asked to behave as if it were a real case, including keeping an accurate, contemporaneous anaesthetic record, but they were not made aware of the purpose of the study. Two changes were made from previous simulations to aid the collection of data. First, the anaesthetic chart had a time scale marked at 2.5-min intervals rather than the usual 5 min to increase the number of data points. This increased the workload of those studied (as does any secondary task method), but we considered that it would not have significantly altered the result because completion of the anaesthetic chart is such a routine part of patient care during anaesthesia. Second, the “patient” and monitors were set to continue through a series of abnormalities irrespective of any of the anaesthetists’ actions. This ensured that all anaesthetists were exposed to exactly the same information from the simulator over the same time scale and that task demands would be held constant from simulation to simulation. The simulation lasted 25 min and started with 5 min of normal baseline readings. Hypotension then developed over the next 5 min with systolic arterial pressure decreasing from 90 to less than 60 mm Hg. A supraventricular tachycardia (SVT) of 160 beat min⁻¹ then started and arterial pressure increased to normal. Five minutes later the SVT terminated and at the same time bronchospasm developed. After another 5 min, the bronchospasm reduced and the simulation ended with 5 min of “normality”. Although the scenario seemed rather contrived, we have noticed during previous studies that those being tested become so involved in treating their “patient”, that the behaviour of the simulator was rarely questioned. We expected therefore that the anaesthetists would treat each abnormality as it arose, as if it were a real case.

At the end of the simulation each anaesthetist was allowed to enter any omitted data on their anaesthetic chart. A print-out of the values shown on the simulator’s monitor was then obtained from the computer. The charted values of heart rate, systolic arterial pressure, diastolic arterial pressure, oxygen saturation and end-tidal carbon dioxide values obtained from the anaesthetic chart and the real values from the simulator print-out were compared and the difference between each of the paired (charted and real) values was calculated. To standardize the error for each variable, the median error for each of the variables was then calculated for all 10 anaesthetists at the first (2.5 min) time. These values are shown in table 1 and were taken to represent the performance of the anaesthetists in the un stressed, baseline state. All subsequent errors were divided by the baseline value to provide a standardized error. An error greater than 1 therefore represents a performance worse than baseline and less than 1 better than baseline. At each time the median standardized error, range and interquartile range for all 10 anaesthetists was calculated.

If a reading was missing from the anaesthetic chart it was allocated the same percentage error as the other readings at that time. If, at a particular time, none of the five variables was charted, it was assumed that the readings were unchanged from the previous entry and the error calculated accordingly.

Table 1 Median errors for all 10 anaesthetists at the first time point

<table>
<thead>
<tr>
<th>Variable</th>
<th>Median error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart rate (beats per minute)</td>
<td>0.9</td>
</tr>
<tr>
<td>Systolic arterial pressure (mm Hg)</td>
<td>5.4</td>
</tr>
<tr>
<td>Diastolic arterial pressure (mm Hg)</td>
<td>3.0</td>
</tr>
<tr>
<td>Oxygen saturation (%)</td>
<td>0.3%</td>
</tr>
<tr>
<td>End-tidal carbon dioxide (kPa)</td>
<td>0.08</td>
</tr>
</tbody>
</table>

To analyse the way in which anaesthetists were using each of the monitors, the data for each monitor were separated and the median standardized error for all 10 anaesthetists was calculated at each time. For statistical analysis, data for all variables for the first time were taken to represent the baseline, unstressed state. The median standardized error and interquartile range were then calculated. Data from the other times were classified into one of the four quartile ranges of the first time. This method was chosen because the data were positively skewed and so unsuitable for parametric analysis. A chi-square analysis was carried out, with P < 0.05 taken as significant. Manipulation of the data and drawing of the graphs were carried out using a standard spreadsheet (EXEL, Microsoft Corporation).
Results

All anaesthetists treated their “patient” in an appropriate manner and successfully completed the simulation. All anaesthetists eventually used a sympathomimetic drug to treat the hypotension. This treatment was then blamed for the production of the supraventricular tachycardia. Arrhythmia usually prompted discontinuation of the volatile anaesthetic and the resulting reduction in depth of anaesthesia was blamed for the bronchospasm. The final phase of normality was accepted as evidence of the success of the anaesthetist’s treatment with either aminophylline, epinephrine or salbutamol.

Data from a single anaesthetist are shown in table 2. Data for all five variables were available from both the anaesthetic charts and the simulator in all cases. The real values shown on the monitors are at the top and the values charted by the anaesthetist are shown below. Calculations were performed as described in the methods section. For clarity all calculated values are shown as integers. Actual calculations were performed using a spreadsheet and the output imported directly into graph drawing software to produce the figures.

Figure 1 shows the median standardized error, interquartile range and range for the five variables, recorded by all 10 anaesthetists. There was a clear increase in median charting error with the onset of the critical incident and a subsequent decrease, although the range was wide. Compared with the results in the first time period, all subsequent results were significantly worse ($P < 0.01$).

Figure 2 shows the median standardized error for three of the variables, as charted by all 10 anaesthetists over the period of the simulation. The data for systolic and diastolic arterial pressures were intermediate and have been omitted for clarity. There was a wide variation in error, but interestingly, the error in charting heart rate was lower than that of end-tidal carbon dioxide during the normal, hypotension and arrhythmia phases. In contrast, during the episode of bronchospasm (from 15 to 20 min), the error in heart rate charting increased above that of end-tidal carbon dioxide.

Discussion

We have shown that it is possible to use error in anaesthetic chart recording as a measure of mental workload and hence performance. Importantly, the method is not intrusive and is less likely to interfere with the normal activities of anaesthetists than methods of measuring mental workload described previously. Our method avoids the inter-observer variability which caused problems with a rating scale tested previously by another group.8 It also does not require the use of videotaping, which is expensive and time consuming to analyse.
This study may be criticized for not involving real anaesthetized patients and for not taking into account the many reasons why those studied may have chosen not to fill in the chart accurately. However, the secondary task method has been used in some industries for many years and has been shown to identify subjects who were subsequently successful in their tries for many years and has been shown to identify anaesthetists to chart abnormal values to avoid later criticism. However, examination of table 1 shows that where the error was greater than 10%, 11 of the charted values erred on the side of normality while 12 were more abnormal than the real values. This supports the view that errors are not introduced deliberately by the anaesthetist, but are caused by excessive mental workload. Another problem is that even with standardization, the variables studied have different normal ranges and variability so that, for example, an error in oxygen saturation of 10% may be considered to be much more significant than a similar error in systolic pressure, making comparisons between variables difficult. It is true that the errors for each of the variables studied are not directly comparable and simply standardizing the values results in loss of information. We suggest that interpretation of the results always requires a detailed knowledge of the circumstances under which the data were collected and of the psychological theory on which this method is based. However, our method has the advantage of producing a clear and easily comprehended result and the error of each variable can easily be studied individually if further information is needed.

The results in figure 1 show that chart error was clearly related temporally to the critical incident. This association was highly significant, even in this small study. However, the wide variability in error indicates that to provide a valid assessment of an individual, their performance over a range of tasks needs to be studied.

Figure 2 shows that the error was not divided equally between different monitored variables. Other research has indicated that, as situations become more complex and stressful, subjects tend to shed tasks and information so as to concentrate on what is perceived as important. It is interesting that the error in charting heart rate was generally low and then increased markedly during the period of bronchospasm, much more than that of end-tidal carbon dioxide. The inference is that the anaesthetists usually attended to the heart rate, but during a respiratory problem their attention was shifted elsewhere. This is perhaps unsurprising but almost impossible to quantify with any other method. The observation that charting error changes with the task suggests that it may be possible to assess new ways of presenting information during anaesthesia and allows the evaluation of new monitors without putting patients at risk.

As our previous study, indicated, our perception of events can differ widely from reality. While this study was designed to remove the previously demonstrated distortion of time associated with increased workload, we were still able to demonstrate an increase in chart recording error with the onset of a simulated critical incident. This is a strong indication that mental overload had occurred or was imminent in those tested. Solutions to such a problem are not clear at present. Techniques such as analysis of charting error may allow us to evaluate a variety of possible solutions in a scientific way. Thus if training and changes to equipment or the theatre environment lead to reduced mental workload, then it is likely that the chances of future errors is also reduced.

We must stress that the use of chart error is not necessarily appropriate for use in assessing the performance of individual anaesthetists, in real operating theatres, by retrospective analysis of their anaesthetic charts. First, the operating room is a complex environment and the care of the patient involves a variety of tasks, any one of which may rightly take precedence over completion of the anaesthetic chart. Second, the limitations of the monitors in current use may mean that we cannot always accept their output as a true reflection of the condition of the patient. Third, performance is not synonymous with competence. Poor performance cannot be equated with standards of patient care. However, this method might be applied to real anaesthetics, not for assessing individuals, but rather for assessing performance of groups of anaesthetists. For example it might be used to compare the results from a large group of anaesthetists across theatres, to assess the impact of equipment configurations or theatre layouts.

This study has another important implication. At present, the accuracy of the anaesthetic chart is important in the legal definition of acceptable standards of patient care. This study indicates that in a real critical incident, the accuracy of any anaesthetic chart is highly questionable. Further, the more stressful the situation, the less likely the chart is to be accurate. Thus the relevance of anaesthetic chart accuracy to the legal process must be questioned.

Finally, a recent editorial, commented that the increasing safety of anaesthesia is progressively reducing the value of studies such as the Confidential Enquiry into Maternal Deaths because deaths caused by anaesthesia have become so uncommon. To continue this progress, we need techniques which
can detect poor performance even when it only has the potential to cause harm to patients. We believe that the use of secondary task techniques and chart error in particular may provide a way of assessing potential safety problems before errors and accidents occur.

Acknowledgements

We thank the Association of Anaesthetists for the provision of a research grant, Ambu International (UK) Ltd and Ohmeda Ltd for the loan of equipment, and Mr H. Cooke for advice and support.

References