Novel method of measuring the mental workload of anaesthetists during clinical practice

A. J. Byrne1*, M. Oliver2, O. Bodger1, W. A. Barnett1, D. Williams2, H. Jones2 and A. Murphy1

1 School of Medicine, Swansea University, Swansea SA2 8PP, UK
2 Abertawe Bro Morgannwg University NHS Trust, Morriston Hospital, Swansea SA6 6NL, UK
* Corresponding author. E-mail: a.byrne@swansea.ac.uk

Key points
- Excessive cognitive load can lead an anaesthetist prone to making errors.
- A vibrotactile device and reaction times were used as surrogate measures of cognitive load.
- Increased reaction times were seen during stressful periods in routine anaesthetic practice.
- The method may be of use in providing an objective assessment of cognitive load of an anaesthetist.

Background. Cognitive overload has been recognized as a significant cause of error in industries such as aviation and measuring mental workload has become a key method of improving safety. The aim of this study was to pilot the use of a new method of measuring mental workload in the operating theatre using a previously published methodology.

Methods. The mental workload of the anaesthetists was assessed by measuring their response times to a wireless vibrotactile device and the NASA TLX subjective workload score during routine surgical procedures. Primary task workload was inferred from the phase of anaesthesia.

Results. Significantly increased response time was associated with the induction phase of anaesthesia compared with maintenance/emergence, non-consultant grade, and during more complex cases. Increased response was also associated with self-reported mental load, physical load, and frustration. These findings are consistent with periods of increased mental workload and with the findings of other studies using similar techniques.

Conclusions. These findings confirm the importance of mental workload to the performance of anaesthetists and suggest that increased mental workload is likely to be a common problem in clinical practice. Although further studies are required, the method described may be useful for the measurement of the mental workload of anaesthetists.

Keywords: medical error; mental workload; psychological factors; training

Accepted for publication: 16 June 2010

Mental workload is used to describe the amount of mental effort involved in performing any given task.1, 2 Given that there is a limit to the mental workload of any individual (mental capacity), mental workload is the proportion of that capacity in use at any point in time and will vary depending on the task being performed by the individual.3

The measurement of mental workload is considered important in many high-risk environments because of its implications for safety, crew size, and the effects of automation.4 Increased mental workload has been associated with poor performance. Jordan noted:

Mental workload is a measure of efficiency that has been widely used in assessing the usability of products where the time in which to carry out tasks is fixed and where error rates are low. This includes, for example, in-vehicle systems, systems in aircraft, and control panels for safety critical processes. The higher the level of mental workload when driving a car or operating a nuclear power plant, the greater the likelihood of an error occurring.5

Mental workload can be assessed concurrently in real time using several different methods, including Primary Task (response to task-related demands), Objective, physiological measures (heart rate and skin resistance), Objective secondary tasks (warning lights and mental arithmetic), and Subjective, psychological (rating scales).

The secondary (subsidiary) task paradigm6, 7 adds a minimally intrusive second task, whose performance is easily measured, to the primary task under study (i.e. administering anaesthesia or flying a plane).

The assumption is that if the workload associated with the primary task becomes excessive, there will be limited spare capacity. At this point, the performance of the secondary task will be impaired.

Thus, if secondary task performance is monitored continuously, any deterioration suggests that the mental workload of the operator is close to capacity and that their performance is likely to be impaired. This condition has been identified in anaesthetists through errors in simple mathematical problems, increased latency in identifying a warning light, and increased record keeping inaccuracies.2, 4, 8

Although less well investigated than in other industries, mental workload in clinicians is no less important to the occurrence and recovery from human error, and also to the performance of individuals during critical incidents.8–12
Effective measures of mental workload can provide us with the means to evaluate strategies for reducing mental workload and thereby preventing error and poor performance in routine practice.\textsuperscript{13–17} The aim of this study was to pilot the use of a novel secondary task to measure the mental workload of anaesthetists during routine clinical practice using a method described in a previous study.\textsuperscript{1}

Methods

Ethical approval was granted from a clinical research Ethics Committee, and after informed consent, volunteers were observed during 46 routine surgical cases, selected on an opportunistic basis over a 3 month period. Each subject was asked to adhere to their normal working practices during the study and to abandon the study if patient care might be compromised.

Before commencing the study, a computer-controlled vibrotactile device was attached to each subject’s arm using a soft material strap and holster. A signal was sent to this device at random intervals (10–90 s) using a Bluetooth\textsuperscript{TM} link, causing it to vibrate. Technical advice from manufacturers and a Department of Medical Electronics confirmed that this would not interfere with the function of any operating theatre equipment. The vibration was terminated either by the subject pressing a button on the device or pressing the whole device onto the side of their body. Subjects were asked to press the button whenever the device vibrated during the session, but were instructed to prioritize their primary task (patient care). The computer automatically recorded the time the signal was generated and the time taken for the subject to respond in milliseconds. All data were recorded as numbered text files. If the subject did not press the button immediately, the device continued to vibrate until the button had been pressed.

All cases were studied in the operating theatres of a single health-care provider. Although the anaesthetist under study was instructed to care for the patient, a second anaesthetist was always present to monitor the situation and ensure patient safety.

After conclusion of the surgery, all anaesthetists were asked to complete a NASA TLX\textsuperscript{18} workload questionnaire. In each case, details of the grade of anaesthetist, difficulty of the case (simple/intermediate/complex), time of day, and any unexpected problems were recorded, although no personal details were recorded.

In previous studies, the ‘normal’ response time for each subject was determined by observation during a baseline, unloaded period.\textsuperscript{1, 2} However, to avoid a change in clinical practice, the response times were considered to be made up of two different populations; one population of ‘normal’ responses that would be normally distributed about a mean and another population of ‘abnormal’, delayed responses. The data from all subjects were therefore ranked and subjected to repeated Kalmogorov–Smirnov\textsuperscript{19} tests starting with the lowest values. The cut-off point was chosen as the lowest value at which the data were not normally distributed.

Any value below the cut-off point was considered to be normal and recorded as zero, with any delay recorded as the response time minus the cut-off. The standardized delay for each subject during any time period was calculated as the average delay.

Mann–Whitney’s $U$-test was used to compare two groups, with factor analysis used to determine the significance of differences between the multiple groups.

Results

All anaesthetists were able to use the device, and despite the detection of multiple Bluetooth\textsuperscript{TM} devices (mobile phones), and the presence of other electronic devices nearby, the data link appeared to be reliable. In a small number of cases, all data relating to that case were lost due to a laptop shutting down automatically, batteries in the device failing, or the subject moving out of range (>10 m) of the Bluetooth\textsuperscript{TM} link. These cases were excluded from the study.

Initial analysis of the data showed a small number of response times <150 ms. Tests showed that such values could only be recorded by depressing the button before the vibration started. Subsequent observation confirmed that some subjects rested the device on their side, continually depressing the button. Any such data were therefore excluded from later analysis.

Data from 46 cases were collected successfully from a range of anaesthetic grades shown in Table 1.

The cases were rated as simple (30), moderate (13), and complex (three), with 12 associated with complications. These included, ‘drip tissued and hypotension’, ‘blood loss’, ‘prone positioning for a block’, ‘inhalation induction in a child’, ‘failure of laryngoscope bulb’, ‘need to resite cannula’, ‘bradycardia’, ‘desaturation’, and ‘repeated boluses of propofol needed’. Two comments indicated problems pressing the button, ‘when blocking’ and ‘when moving patient in theatre’. One case included an unexpected asystolic cardiac arrest and the comment that the button had only been pressed, ‘because a nurse said the box was buzzing’.

<table>
<thead>
<tr>
<th>Table 1 Numbers of each anaesthetic grade taking part</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade</strong></td>
</tr>
<tr>
<td>Year 1</td>
</tr>
<tr>
<td>Year 2</td>
</tr>
<tr>
<td>Year 3</td>
</tr>
<tr>
<td>Year 4</td>
</tr>
<tr>
<td>Year 5</td>
</tr>
<tr>
<td>Year 6</td>
</tr>
<tr>
<td>Year 7</td>
</tr>
<tr>
<td>Staff grade</td>
</tr>
<tr>
<td>Consultant</td>
</tr>
</tbody>
</table>

Byrne et al.
Results from most subjects approximated to a normal distribution with the mean response time for each individual of 695 ms with a range of 328–2517 ms. Inspection of the response times for the case involving cardiac arrest showed a single, delayed response of 78 947 ms which coincided with the arrest. This case was considered to be atypical and as this single very large value would have caused large changes to the average response times, this case was excluded from further statistical analysis. The limit of ‘normal’ response times for all subjects was 945 ms.

The standardized delay of each subject varied with phase of anaesthesia from a mean (range) of 301 ms (0–7531 ms) during induction, 51 ms (0–443 ms) during maintenance, to 141 ms (0–2572 ms) during emergence (Fig. 1). The difference between induction and maintenance was statistically significant \( P<0.05, Z=-2.150, n=45 \); however, the difference between maintenance and emergence was not \( P>0.5, Z=-0.641, n=45 \).

Simple cases were associated with less standardized delay (range), 40 ms (0–234 ms) than moderate/complex cases, 210 ms (4–1664 ms) \( n=45, Z=-2.728, P<0.006 \).

Initial analysis of the result by anaesthetic grade showed no obvious relationship with standardized delay; however, inspection of the data revealed selection bias, in that intermediate and complex cases tended to be completed by more experienced anaesthetists. Analysis of the response times for simple cases only showed that mean delay decreased with experience (Fig. 2), although the change was not statistically significant.

Factor analysis showed small, significant increased response time during complex/moderate cases compared with simple ones \( P<0.006 \) and with higher TLX scores for mental workload \( P<0.045 \), physical workload \( P<0.004 \), and frustration \( P<0.01 \). There was no relationship between response times and TLX scores for temporal pressure, self-rated performance, and effort.

**Discussion**

This study achieved its primary aim, in that it was possible to measure the response times of anaesthetists to a vibrotactile device in an operating theatre environment and that an increased measured workload was associated with induction of anaesthesia. Although the technology appears to be reliable, operator error did cause data loss and the Bluetooth data link may prove less reliable in other environments.

The methodology also appeared to be successful, in that measured mental workload was associated with the induction phase of anaesthesia and during more complex cases where primary task workload is likely to be high, similar to the findings of other studies.\(^5\)\(^\text{20}\)

The relationship of mental workload with the level of experience is also consistent with the findings of others.\(^11\)\(^15\) The finding of lower workload in first-year trainees was unexpected, but may reflect either increased supervision or that they were not yet aware of the demands of the task, ‘unconscious incompetence’.\(^21\)

It was expected that the presence of complications would result in a higher measured workload. However, the short duration of many of the complications would be unlikely to greatly alter the average response time. Analysis of individual results suggests that there may be short-term changes in response to specific events, but this would require further study with direct observation and timing of events, as has been used in other studies.\(^4\)\(^20\) The correlation between response times and the subjective workload assessed by the TLX questionnaire\(^18\) was also consistent with the findings in other settings,\(^22\)\(^\text{24}\) although the differences were small.

This method has limitations, in that subjects may have made a conscious decision to press the button more slowly, in the same way that patients may consciously report pain scores inaccurately. Further, the delay during induction and emergence could be ascribed to the increased physical activity making it more difficult to press the button. This would also compromise the statistical methods used, which rely on the assumption that the data consist of two distinct groups. However, the same method has demonstrated increased response times during simulated patient consultations, which did not involve any physical activity,\(^25\) and we would still contend that the responses recorded represent either ‘normal’ responses or ‘abnormal’ responses due to increased mental or, in some cases, physical restriction. The validity of this method would therefore be increased by...
removing from analysis any data relating to periods during which subjects report an inability to press the button, although this would require the increased expense of direct observation by a trained researcher.

The study provided data from anaesthetists with a wide range of abilities and under a very wide range of workloads. However, the data suggest that there may be significant differences between the response times of individuals and the anonymity imposed on the study for ethical reasons did not allow us to determine whether this variation was stable among individuals or not. Although we describe a method for determining a limit to the ‘normal’ response of an individual, the power of this method may be greatly increased if it were possible to determine the normal range of response times for each individual before their assessment in the workplace.

In particular, the observation that during an unexpected cardiac arrest, a consultant anaesthetist was unaware of the vibration while responding appropriately during the rest of the case supports the theory that the main cause of the delay is increased mental workload.

It is important to note that the theoretical basis of this study suggests that a delayed response will only occur during times when the subject is close to or exceeding their mental capacity. These data would therefore suggest that excessive mental workload is common during ‘routine’ anaesthetic practice and that although the anaesthetists studied did not make observable errors, their high mental workload would make errors more likely.

A previous study found that 64% of anaesthetists reported errors due to excessive workload. Without validated assessment tools, we cannot measure the performance of medical practitioners and therefore cannot properly evaluate strategies to make patient care safer. This may require the use of multiple, simultaneously measured variables, including subjective and objective methods, as used in other industries rather than the subjective opinions in current use.

In conclusion, the importance of human factors to the safe practice of medicine is becoming increasingly apparent. This study has demonstrated the practicality of using a vibrotactile device and TLX questionnaire to measure mental workload in a clinical environment. These data are consistent with the findings of other studies and with subjects developing cognitive overload during routine delivery of anaesthesia.

Conflict of interest
None declared.

References
8 Weinger M. Multiple measures of anaesthesia workload during teaching and non-teaching cases. Anesth Analg 2004; 98: 58–62
9 Bertram DA, Hershey CO, Opila DA, Quinir O. A measure of physician mental workload in internal medicine ambulatory care clinics. Med Care 1990; 28: 458–67

Fig 2 Standardized delay shown by experience (trainee year, staff grade, and consultant).
20 Weinger MB, Herndon OW, Gaba DM. The Effect of electronic record keeping and transesophageal echocardiography on task distribution, workload, and vigilance during cardiac anesthesia. Anesthesiology 1997; 87: 144–55
22 Charabati S, Bracco D, Mathieu PA, Hemmerling TM. Comparison of four different display designs of a novel anaesthetic monitoring system, the ‘integrated monitor of anaesthesia (IMATM)’. Br J Anaesth 2009; 103: 670–7
26 Wickens CD. Multiple resources and mental workload. Hum Factors 2008; 50: 449–55
27 Gray WD. Cognitive architectures: choreographing the dance of mental operations with the task environment. Hum Factors 2008; 50: 497–505