Anaesthesia workload measurement devices - Qualitative systematic review

Dalal Almghairbi,

Anaesthesia and Critical Care Research Group,

Division of Clinical Neuroscience,

University of Nottingham (Queen’s Medical Centre Campus),

Nottingham University Hospitals NHS Trust,

Derby Road, Nottingham, NG7 2UH.

Dalal S. Almghairbi¹, Takawira C. Marufu² and Iain K. Moppett³

1 PhD student, Anaesthesia and Critical Care Research Group, Division of Clinical Neuroscience, Queen’s Medical Centre, University of Nottingham, UK.

2 PhD student, Anaesthesia and Critical Care Research Group, Division of Clinical Neuroscience, Queen’s Medical Centre, University of Nottingham, UK.

3 Professor and Honorary Consultant Anaesthetist, Anaesthesia and Critical Care Research Group, Division of Clinical Neuroscience, Queen’s Medical Centre, University of Nottingham, UK.

Corresponding author: Dalal S Almghairbi

Email: msxda@nottingham.ac.uk

Phone number: +44 (0) 115 82 31006

Abstract: 235

Main body: 2351

Number of References: 49

Figures: 1

Tables: 2

Appendix: 2
ABSTRACT

Background: Management of mental workload is a key aspect of safety in anaesthesia but there is no gold standard tool to assess mental workload, risking confusion in clinical and research use of such tools.

Objective: This review assessed currently used mental workload assessment tools.

Methods: A systematic literature search was performed on the following electronic databases; Cochrane, EMBASE, MEDLINE, SCOPUS and Web of Science. Screening and data extraction were performed individually by two authors. We included primary published papers focusing on mental workload assessment tools in anaesthesia.

Results: A total of 2331 studies were screened by title, 32 at full text and twenty - four studies met the inclusion criteria. Six mental workload measurement tools were observed across included studies. Reliability for the Borg rating scales and Vibrotactile device were reported in two individual studies. The rest of the studies did not record reliability of the tool measurements used. Borg rating scales, NASA-TLX and task oriented mental work load measurements are subjective, easily available, readily accessible and takes a few minutes to complete. However, the Vibrotactile and Eye-tracking methods are objective, require more technical involvement, considerable time for the investigator, and moderately expensive, impacting their potential use.

Conclusion: We found that, the measurement of mental workload in anaesthesia is an emerging field supporting patient and anaesthetist safety. The self - reported measures have the best evidence base

Key Words: workload, over load, anaesthesia, anaesthetist.
INTRODUCTION
Mental workload has been defined in terms of level of attention and resources required to meet objective and subjective performance criteria of an individual task, which may be mediated by task demands, external support and past experience [1]. It has been identified as key performance factor in various complex working environments [2]. Anaesthetists work in an environment where mental workload may impact on safety through slips, lapses and conflict [3]. Managed workload may have an impact on job satisfaction within the operating team, and provide cost effective care. Studies [4, 5] have demonstrated that workload may lead to increase anaesthetist stress (in part an imbalance between workload and resources) [6], burnout and fatigue. Both aviation and nuclear industries report associations between mental workload, system performance and safety [7, 8]. Mitigating effects of workload / capacity mismatch requires understanding of temporal, individual and contextual - sensitive changes in workload. However, although it might relatively be easy to identify times of increased or reduced mental workload, it is not a trivial issue to reliably quantify this [9]. Several methods have been proposed to measure anaesthesia workload based on metrics used in other industries, but to date none of the assessment tools available has been considered standard for use across healthcare. Whilst it is unlikely that any single tool will ever be optimal for all situations, it is important for researchers and practitioners to understand what various tools can and cannot be expected to do when used or reported in studies.
Current mental - workload measurements are conceived from four parallel conceptual frameworks (scoring / task oriented, task performance, response time capabilities and physiological changes). Subjective task oriented methodologies are multidimensional constructs of subscales that include: mental, physical, temporal demand, frustrations, effort and performance. The theory of such tools is that combination of specific elements is more
likely to describe the workload experienced by most operators doing the tasks [7]; and most importantly the operator is the best individual to rank / score how difficult the task is perceived. Other tools work from the premise that if a primary (clinical) task is associated with high level of workload, limited spare mental capacity will be available. This will be observed by delays in reacting to a secondary task – analogous to a computer slowing down if it has insufficient memory or processing power. It may also be observed in error rates or time taken for the primary tasks themselves. Physiological measurement tools assume that increases in mental workload are associated with observable changes in physiology, mediated via the autonomic nervous system [10]. This review assesses the theoretical framework and supporting data for currently used mental workload assessment tools in the anaesthetic environment, in order to allow researchers and practitioners to understand the relative merits of the available tools.

**Table 1** Criteria for workload measurement methods

**METHODS**

A systematic literature search was performed across five databases; Cochrane, EMBASE (via Ovid platform), MEDLINE (via Ovid platform), SCOPUS and Web of Science. The search strategy was piloted and tailored to the individual databases. A combination of Medical Subject Headings (MeSH) and free text terms was used to increase sensitivity for identification of potential studies. Search terms used ([workload* OR work load* OR overload* OR taskload* OR burnout* OR stress* OR anxiety* OR fatigue] AND ([anaesthesia* OR anaesthetist* OR anaesthesiology*]) were initially run individually and then combined in each database. References list of identified papers and previous reviews were checked for further data and citing articles were also sought from Google Scholar.

We included primary published research papers focusing on mental workload assessment tools used in anaesthesia and / or impact of workload on anaesthetists and patient safety up to
February 2018. Studies in other languages were considered if an English translation version was available. We excluded reviews, descriptive articles, letters to the editor and opinions.

**Data Extraction**

DA and TM both performed the search strategy and independently identified potential studies for inclusion at abstract stage. Data extraction was performed individually with disagreements resolved by the discussion and involvement of a third researcher (IM). Data extracted from individual studies included: sample size, study region, date of study, validity and reliability of workload measurement tool. Data are reported as presented in the studies. Due to heterogeneity of study results presentation, populations and tools used, no attempt has been made to pool data.

**RESULTS**

**Overview**

Primary search produced 2388 articles. 2313 studies were screened by title after removal of duplicates; 40 were eligible for abstract screening. Thirty-two studies were considered for full text analysis and twenty-four [11-34] met the inclusion criteria with outcomes given with adequate data to assess study results. Most studies screened at full text phase looked at effects of workload in general, not its measurement, and therefore were excluded [35-42].

**Figure 1:** Flow diagram of included and excluded studies

Characteristics of included studies are outlined in **Appendix 1**.

**Types of mental workload measurement tools**

Six mental workload measurement tools were identified (Table 2). Rating scales of workload were used in sixteen of these twenty-four studies. The Borg rating (6 to 20) scale was used in nine [19-22, 24-28] and NASA - TLX in five [13, 15, 28, 30, 32]; Gaba, [23] used a bespoke rating scale and Vredenburgh, [33] used a survey instrument, both similar to NASA - TLX. Analysis of primary task performance (tasks required as part of anaesthesia care) was
assessed in four studies [11, 12, 23, 33], and performance of secondary tasks (arithmetic) was reported in three studies [23, 29, 34]. Of the twenty-four studies, eleven assessed only one of the four concepts (scoring / task oriented - 2, task performance - 5, response time capabilities - 2 and physiological changes – 2); ten studies assessed scoring and response times, two scoring and physiology; one study assessed scoring, response time capabilities and physiology. Vigilance was assessed in thirteen studies using either the Vibrotactile device [13-16] or response to a randomly illuminated light [20-22, 24-27, 34] or alarm - sound response latency [32]. Physiological monitoring was reported in five papers [17-19, 22, 28] from three studies. Pupil responses were reported in two papers from the same study [17, 19] and one from a different study [18].

Type of tools used for each study outlined in Appendix 2.

The Borg rating scale is widely used to assess workload in anaesthesia. It assess workload using a perceived exertion numerical scale from 6 (no exertion at all) to 20 (maximum exertion). Its use in anaesthesia settings such as clinical environments, simulated critical incidents and non-simulated critical incidents showed similar results. Essentially, during routine cases of anaesthesia the Borg rating scale showed an increase in workload values during induction and emergency compared with maintenance. Similarly, comparatively high Borg rating values were observed during complex and moderate cases compared with simple cases. More experienced anaesthetists appeared to have spare mental capacity during standardised primary anaesthetic tasks as observed in their reporting of lower workload values from standardised incidents in comparison with less experienced anaesthetists. It could be argued the Borg scale is unidimensional and do not explicitly differentiate between physical and mental workload. Nevertheless observed results from its use across included studies support the Borg rating scale to be an effective tool workload measurement tool.
The NASA - TLX workload measurement tool uses six categories: effort; mental demand; physical demand; temporal demand; frustration; and performance to measure workload. Variations in NASA - TLX were comparable with Borg rating scale and/or physiological monitoring tools [18, 28], when both tools were used. When used with physiological measuring tools, significant positive correlation was observed between physiological outcome measures (heart rate, heart rate variability, and pupil size) and NASA - TLX scores [28]. For NASA - TLX domains across all included studies, higher workload scores were consistently observed in temporal, mental and physical domains not changed. However, there were inconsistent results for the frustration and performance domains across studies.

Tasks were used in two ways. First, accuracy of completion of a primary task (directly related to anaesthesia care) was used as a measure of workload. Alternatively, a secondary task was used to increase workload and/or assess spare capacity. A positive correlation between subjective workload and primary workload density was reported [20, 21]. Workload assessment using secondary task measurements reported high levels of workload. Alarm - response latency (using either a light or Vibrotactile device) was used to measure vigilance among anaesthetists. Response latency increased significantly at times of increased workload, whether during routine or crisis induced variation. The results of increasing workload with secondary tasks was inconsistent. Some studies reported a statistically significant increase in mean response time and others not; this may have been due to individuals’ performance variability (inter-interindividual variation).

Physiological workload measurement tools uses physiological responses as a surrogate marker of mental workload including the assessment of subjects’ visual focus of attention. Heart rate alone had inconsistent relationship with workload with marked inter-individual variability. More advanced metrics of autonomic function (derived from heart rate variability)
showed better discriminative ability in one study [28]. Pupil changes were inconsistently related with workload with marked inter-individual variability.

RELIABILITY

Two studies reported reliability of the tools measurement used. One study [16] reported reliability ($\alpha = 0.922$) for the Vibrotactile device. Another study [27] reported reliability of the Borg scale which found to be moderate (concordance coefficient = 0.55). Several studies [20-22, 25, 26] reported moderate to good correlation between self-reported and observer workload rating using the Borg scale.

DISCUSSION

Four techniques of measuring mental workload were found in our review; scoring, task performance, response time capabilities and physiological measures. The Borg rating scale measuring method was the most commonly used method and showed consistent association with expected variations in workload. Correlation between self-reported and observed rating was moderate to good.

Areas with increased workload such as the emergency department, intensive care unit and operating room are associated with a significantly higher rate of medical errors compared to other departments [43]. While less well studied compared with other high risk industries, mental workload among anaesthetists is presumably related to incidents and recovery from human error [44]. Observed results support the concept that anaesthesia is similar to other high risk industries such as aviation and nuclear power where human behaviour has positive and negative impacts on safety and performance. Workload is believed to be a key contributing factor for human error [3]. NASA - TLX and Borg scale mental workload measurements are subjective cognitive workload measures that can be performed with a paper and a pencil and are easily available, readily accessible, and only take a few minutes to complete [7]. They have been widely studied and used in other settings [9]. The NASA - TLX questionnaire has been
extensively validated in the aviation and nuclear industries. It has been found to have diagnostic capability for subjective overload [45]. It does not interfere with primary tasks and affords opportunity for operators’ individual perception of workload to be explored in depth. The Borg rating scale was the single most common measure used in the identified studies. It consistently demonstrates discriminant ability between expected contexts of varying workload: phases of anaesthesia, routine versus and non-routine simulated scenarios.

The interpretation and ability of these scores to quantify performance is uncertain. Subjectivity is integral to the approach [1] and it is unclear whether they are measuring mental workload or ‘stress’ [46]. All of the studies reported individual NASA-TLX subscales as well as the weighted summary score, though formal analysis was less commonly carried out. Qualitatively, the temporal and mental domains were consistently associated with workload and physical domains not. Results were inconsistent for the frustration and performance domains, suggesting that NASA-TLX should be reported at domains level.

Vigilance (response latency) and physiological data are objective methods that require more technical involvement, considerable time for investigator, and are moderately expensive. Physiological measurement methods are perceived as complex, intrusive and less likely to be part of daily routine use [47]. With advances in technology, equipment is becoming much more portable, capable and acceptable [9]. The Vibrotactile device is unobtrusive and allows operators to move easily and can be freely used in different clinical practice settings without interfere with normal working practice. However, the definition of delay or threshold varies between studies which hampers interpretation.

Eye-tracking provides objective surrogate measures of physiological responses (pupil diameter) and other behaviours. However, pupil diameter varies by individual, light intensity, time of day and caffeine, all of which are relevant to anaesthetists. Studies in aviation found that the influence of mental effort on eye movement measurement is highly reliant on specific
task characteristics [48] and therefore it is not suitable to use as a method to assess workload in aviation.

Heart rate appears to be inconsistent in its ability to discriminate workload. Advances in technology allow relatively straightforward and unobtrusive multimodal monitoring of autonomic physiology through chest harnesses. Johnstone et al [49] reported that this multi-variable technology was demonstrated to be reliable and valid in a laboratory setting. However, in the workplace, inter-individual variability, therapeutic drug use, caffeine and physical workload all potentially act as confounders. Changes within a low physical load situation may be reflective of mental workload, but this ideal may not exist in most scenarios where mental workload coincides with increased movement.

STUDY LIMITATIONS

Presentation of results, and the methods used to generate or infer differences in workload, differed across the studies, making direct comparisons difficult. There were limited data comparing tools in the same setting.

Despite the attraction of objective measurement – some approaches directly interfere with clinical practice, making their widespread use questionable.

Self-reported measures have theoretical problems – subjectivity, lack of clarity of what is actually being measured, and time delay. But on the evidence presented in this review they do appear to have validity for measuring workload. They are sensitive to changes in workload in a variety of contexts and seem acceptable to users.

CONCLUSION

The measurement of mental workload in anaesthesia is an emerging field supporting patient and anaesthetist safety. Self-reported measures have the best evidence base to date.
FUNDING STATEMENT
This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

AUTHOR CONTRIBUTIONS
DA and IM make substantial contributions to conception and design of the study.
DA and TM created search strategies and data extraction
DA participate in drafting the article and IM revising it critically for important intellectual content.
IM gave final approval of the version to be submitted

CONFLICTS OF INTEREST
This work forms part of DA’s PhD thesis, which is supported by scholarship from the Libyan Ministry of Higher Education and university of Zawia. TM was a PhD student supported by a grant from the Sir Jules Thorn Charitable Trust. IM is a member of the NICE topic expert group for Quality Standards for hip fracture, Deputy Director of the National Institute of Academic Anaesthesia (NIAA) Research Council and holds grants from the National Institute for Health Research and the Association of Anaesthetists of Great Britain & Ireland and Royal College of Anaesthetists through the NIAA for trials in hip fracture.
REFERENCES


26. Merry et al. Multimodal system designed to reduce errors in recording and administration of drugs in anaesthesia: prospective randomised clinical evaluation. *BMJ* 2011; **343**: 92-019

27. Merry et al. Retesting the Hypothesis of a Clinical Randomized Controlled Trial in a Simulation Environment to Validate Anesthesia Simulation in Error Research (the VASER Study). *Anesthesiology* 2017; **126**: 472-81.


Table 1: Criteria for workload measurement methods [9]

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplicity and usability</td>
<td>minimum equipment, non-intrusive / non-interference with performance, acceptable to subjects</td>
</tr>
<tr>
<td>Availability</td>
<td>timely and sufficiently rapid for use</td>
</tr>
<tr>
<td>Performance characteristics</td>
<td>sensitivity to changes, selectively sensitive, insensitive to other task demands, inter- and intra-rater reliability, adequate floor and ceiling effects, face validity</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>indicating the source of workload variation</td>
</tr>
<tr>
<td>Tools</td>
<td>Measurement</td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Borg workload</td>
<td>Self-reported workload score</td>
</tr>
<tr>
<td>NASA - Task Load Index</td>
<td>Self-rating workload chart in six subscales mental, physical, temporal demand, frustration, effort and performance / success.</td>
</tr>
<tr>
<td>Tasks</td>
<td>Response time</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Vibrotactile device</td>
<td>Response time to a vibration on the upper arm</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye - tracking device</td>
<td>Pupil diameter and eye movements</td>
</tr>
<tr>
<td>(EyeSeeCam)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Autonomic responses</td>
<td>Heart rate</td>
</tr>
<tr>
<td></td>
<td>Respiratory rate</td>
</tr>
</tbody>
</table>

Table 2: Tools used to measure mental workload in anaesthesia