Anaesthesia workload measurement devices: qualitative systematic review

Dalal S Almghairbi, Takawira C Marufu, Iain K Moppett

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmjstel-2017-000263).

Anaesthesia and Critical Care Research Group, Division of Clinical Neuroscience, Queen's Medical Centre, University of Nottingham, Nottingham, UK

Correspondence to
Dalal S Almghairbi, Anaesthesia and Critical Care Research Group, Division of Clinical Neuroscience, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK; msxda@nottingham.ac.uk

Received 3 October 2017
Accepted 24 February 2018
Published Online First
21 March 2018

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 by full text and 24 studies met the inclusion criteria. Six mental workload measurement tools were observed across included studies. Reliability for the Borg rating scales and Vibrotactile device was 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.

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. While 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). Criteria of methods are detailed intable 1. 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.

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 over load* OR taskload* OR burnout* OR stress* OR anxiety* OR fatigue] AND ([anaesthesia* OR anaesthetist*

Linked

► http://dx.doi.org/10.1136/bmjstel-2018-000330

Check for updates

To cite: Almghairbi DS, Marufu TC, Moppett IK. *BMJ Stel* 2018;**4**:119–123.

Table 1 Criteria for workload measurement methods[9]

Criteria	Explanation
Simplicity and usability	Minimum equipment, non-intrusive/non interference with performance, acceptable to subjects
Availability	Timely and sufficiently rapid for use
Performance characteristics	Sensitivity to changes, selectively sensitive, insensitive to other task demands, inter-rater and intrarater reliability, adequate floor and ceiling effects, face validity
Diagnostic	Indicating the source of workload variation

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. A total of 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 24[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] Flow chart of study selection is detailed in figure 1.

Characteristics of included studies are outlined in the online supplementary appendix 1.

Types of mental workload measurement tools

Six mental workload measurement tools were identified (table 2). Rating scales of workload were used in 16 of these 24 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 and Lee[23] used a bespoke rating scale and Vredenburgh *et al*[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 24 studies, 11 assessed only one of the four concepts (scoring/task-oriented: 2, task performance: 5, response time capabilities: 2 and physiological changes: 2); 10 studies assessed scoring and response times, 2 scoring and physiology; 1 study assessed scoring, response time capabilities and physiology. Vigilance was assessed in 13 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 are outlined in the online supplementary 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

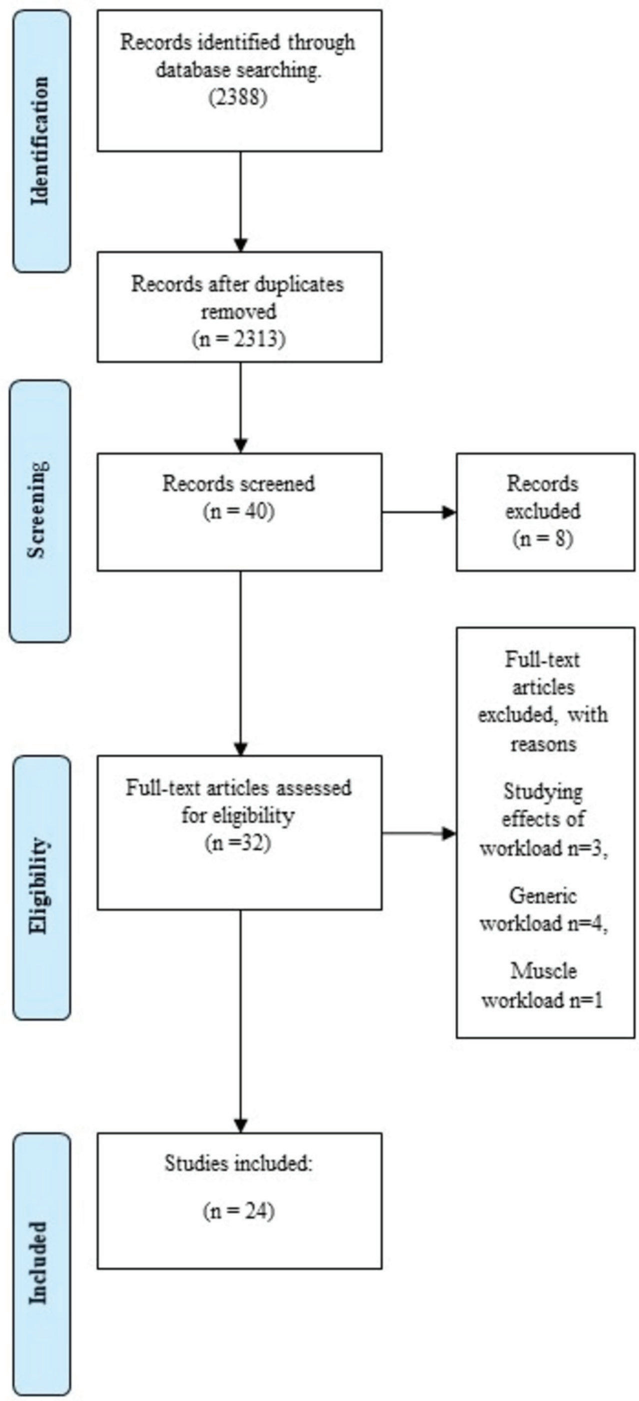

Figure 1 Flow diagram of included and excluded studies.

 Almghairbi DS, *et al*. *BMJ Stel* 2018;**4**:119–123. doi:10.1136/bmjstel-2017-000263

Table 2 Tools used to measure mental workload in anaesthesia

Tools	Measurement	Outcome
Borg workload	Self-reported workload score	6 (no exertion), 20 (maximum exertion)
NASA-Task Load Index	Self-rating workload chart in six subscales mental, physical, temporal demand, frustration, effort and performance/success	Score ranges from 0 (minimal) to 10 (maximum possible)
Tasks	Response time	Primary task (response to task-related demands) Secondary task performance (measure of spare capacity) Absolute response time of mathematical addition task The rate of correct answers to the mathematical questions every 5 s
Vibrotactile device	Response time to a vibration on the upper arm	Absolute response time (ms) Response time (ms) above statistically derived 'threshold'
Eye-tracking device (EyeSeeCam)	Pupil diameter and eye movements	Absolute (mm) and change in pupil diameter Mean duration (seconds) of fixation episodes Amplitude (degrees) of saccades
Autonomic responses	Heart rate Respiratory rate	Absolute and relative changes in HR Heart rate variability: time, frequency and entropy-derived metrics

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 the 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 tasks (directly related to anaesthesia care) was used as a measure of workload. Alternatively, a secondary task was used to increase workload and/or assesses 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 (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 interindividual 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 to workload with marked interindividual variability.

RELIABILITY

Two studies reported reliability of the tools' measurement used. One study[16] reported reliability (α=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 with 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 unobstrusive multimodal monitoring of autonomic physiology through chest harnesses. Johnstone *et al*[49] reported that this multivariable technology was demonstrated to be reliable and valid in a laboratory setting. However, in the workplace, interindividual 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.

Contributors DSA and IKM made substantial contributions to conception and design of the study. DSA and TCM created search strategies and data extraction. DSA participated in drafting the article and IKM in revising it critically for important intellectual content. IKM gave final approval of the version to be submitted.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests 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. TCM was a PhD student supported by a grant from the Sir Jules Thorn Charitable Trust. IKM 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.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Young MS, Stanton NA. Mental workload: theory, measurement, and application. Karwowski W, eds. *International encyclopedia of ergonomics and human factors*. 1. London: Taylor & Francis, 2001:507–9.
2 Wickens CD. Multiple resources and mental workload. *Hum Factors* 2008;50:449–55.
3 Glavin RJ. Drug errors: consequences, mechanisms, and avoidance. *Br J Anaesth* 2010;105:76–82.
4 Lindfors PM, Nurmi KE, Meretoja OA, *et al*. On-call stress among Finnish anaesthetists. *Anaesthesia* 2006;61:856–66.
5 Nyssen AS, Hansez I, Baele P, *et al*. Occupational stress and burnout in anaesthesia. *Br J Anaesth* 2003;90:333–7.
6 Bakker AB, Demerouti E. The Job demands-resources model: state of the art. *J Manag Psychol* 2007;22:309–28.
7 Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proc Hum Factors Ergon Soc Annu Meet* 2006;50:904–8.
8 Hwang S-L, Yau Y-J, Lin Y-T, *et al*. Predicting work performance in nuclear power plants. *Saf Sci* 2008;46:1115–24.
9 Cain B. A review of the mental workload literature. 2007 http://www.dtic.mil/dtic/tr/fulltext/u2/a474193.pdf (accessed on 25 Jul 2016).
10 O'Donnell R, Eggemeier F. In: Boff K, Kaufman L, Thomas JP, eds. *Workload assessment methodology handbook of perception and human performance*. New York: Wiley, 1986.
11 Byrne AJ, Sellen AJ, Jones JG. Errors on anaesthetic record charts as a measure of anaesthetic performance during simulated critical incidents. *Br J Anaesth* 1998;80:58–62.
12 Byrne AJ, Sellen AJ, Jones JG, *et al*. Effect of videotape feedback on anaesthetists' performance while managing simulated anaesthetic crises: a multicentre study. *Anaesthesia* 2002;57:176–9.
13 Byrne AJ, Oliver M, Bodger O, *et al*. Novel method of measuring the mental workload of anaesthetists during clinical practice. *Br J Anaesth* 2010;105:767–71.
14 Byrne AJ, Murphy A, McIntyre O, *et al*. The relationship between experience and mental workload in anaesthetic practice: an observational study. *Anaesthesia* 2013;68:1266–72.
15 Byrne A, Tweed N, Halligan C. A pilot study of the mental workload of objective structured clinical examination examiners. *Med Educ* 2014;48:262–7.
16 Davis DH, Oliver M, Byrne AJ. A novel method of measuring the mental workload of anaesthetists during simulated practice. *Br J Anaesth* 2009;103:665–9.
17 Schulz CM, Schneider E, Fritz L, *et al*. Eye tracking for assessment of workload: a pilot study in an anaesthesia simulator environment. *Br J Anaesth* 2011;106:44–50.
18 Schulz CM, Skrzypczak M, Schneider E, *et al*. Assessment of subjective workload in an anaesthesia simulator environment: reliability and validity. *Eur J Anaesthesiol* 2011;28:502–5.
19 Schulz CM, Schneider E, Kohlbecher S, *et al*. The influence of anaesthetists' experience on workload, performance and visual attention during simulated critical incidents. *J Clin Monit Comput* 2014;28:475–80.
20 Weinger MB, Herndon OW, Zornow MH, *et al*. An objective methodology for task analysis and workload assessment in anesthesia providers. *Anesthesiology* 1994;80:77–92.
21 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 Weinger MB, Reddy SB, Slagle JM. Multiple measures of anesthesia workload during teaching and nonteaching cases. *Anesth Analg* 2004;98:1419–25.
23 Gaba DM, Lee T. Measuring the workload of the anesthesiologist. *Anesth Analg* 1990;71:354–61.

24 Slagle JM, Weinger MB. Effects of intraoperative reading on vigilance and workload during anesthesia care in an academic medical center. *Anesthesiology* 2009;110:275–83.
25 Slagle JM, Porterfield ES, Lorinc AN, *et al*. Prevalence of potentially distracting noncare activities and their effects on vigilance, workload, and nonroutine events during anesthesia care. *Anesthesiology* 2018;128:44–54.
26 Merry AF, Webster CS, Hannam J, *et al*. Multimodal system designed to reduce errors in recording and administration of drugs in anaesthesia: prospective randomised clinical evaluation. *BMJ* 2011;343:d5543–019.
27 Merry AF, Hannam JA, Webster CS, *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.
28 Martin J, Schneider F, Kowalewskij A, *et al*. Linear and non-linear heart rate metrics for the assessment of anaesthetists' workload during general anaesthesia. *Br J Anaesth* 2016;117:767–74.
29 Sato H, Miyashita T, Kawakami H, *et al*. Influence of mental workload on the performance of anesthesiologists during induction of general anesthesia: a patient simulator study. *Biomed Res Int* 2016;2016:1–5.
30 Prottengeier J, Petzoldt M, Jess N, *et al*. The effect of a standardised source of divided attention in airway management: A randomised, crossover, interventional manikin study. *Eur J Anaesthesiol* 2016;33:195–203.
31 Garden AL, Robinson BJ, Kappus LJ, *et al*. Fifteen-hour day shifts have little effect on the performance of taskwork by anaesthesia trainees during uncomplicated clinical simulation. *Anaesth Intensive Care* 2012;40:1028–34.
32 de Man FR, Erwteman M, van Groeningen D, *et al*. The effect of audible alarms on anaesthesiologists' response times to adverse events in a simulated anaesthesia environment: a randomised trial. *Anaesthesia* 2014;69:598–603.
33 Vredenburgh AG, Weinger MB, Williams KJ, *et al*. Developing a technique to measure anesthesiologists' real-time workload. Proceedings of the human factors and ergonomics society annual meeting. Los Angeles, CA:SAGE Publications Sage CA, 2000.
34 Cao CG, Weinger MB, Slagle J, *et al*. Differences in day and night shift clinical performance in anesthesiology. *Hum Factors* 2008;50:276–90.
35 Notcutt WG, Down MP. Anaesthetists and stress. *Anaesthesia* 1996;51:1072–3.
36 Peltola L, Mayer J, Cook J, *et al*. Increased anaesthetic workload associated with increased maternal age. *Int J Obstet Anesth* 2012;21:100–1.
37 Chiron B, Michinov E, Olivier-Chiron E, *et al*. Job satisfaction, life satisfaction and burnout in French anaesthetists. *J Health Psychol* 2010;15:948–58.
38 Willoughby L, Morgan R. Neuroanaesthetists' experience of workload-related issues and long-duration cases. *Anaesthesia* 2005;60:151–4.
39 Carvalho B, Coghill J. Obstetric anaesthesia workload and time of day. *Int J Obstet Anesth* 2004;13:126–7.
40 Hagau N, Pop RS. Prevalence of burnout in Romanian anaesthesia and intensive care physicians and associated factors. *J Rom Anest Terap Int* 2012;19:117–24.
41 Howard SK, Gaba DM, Smith BE, *et al*. Simulation study of rested versus sleep-deprived anesthesiologists. *Anesthesiology* 2003;98:1345–55.
42 Caldiroli D, Molteni F, Sommariva A, *et al*. Upper limb muscular activity and perceived workload during laryngoscopy: comparison of Glidescope(R) and Macintosh laryngoscopy in manikin: an observational study. *Br J Anaesth* 2014;112:563–9.
43 Cao A, Chintamani KK, Pandya AK, *et al*. NASA TLX: software for assessing subjective mental workload. *Behav Res Methods* 2009;41:113–7.
44 Leedal JM, Smith AF. Methodological approaches to anaesthetists' workload in the operating theatre. *Br J Anaesth* 2005;94:702–9.
45 Soekkha HM. *Aviation Safety: Human factors, system engineering, flight operations, economics, strategies, management: VSP*, 1997.
46 Meshkati N. Heart rate variability and mental workload assessment. *Adv Cogn Psychol* 1988;52:101–15.
47 Schulz CM, Schneider E, Fritz L, *et al*. Visual attention of anaesthetists during simulated critical incidents. *Br J Anaesth* 2011;106:807–13.
48 De Rivecourt M, Kuperus MN, Post WJ, *et al*. Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight. *Ergonomics* 2008;51:1295–319.
49 Johnstone JA, Ford PA, Hughes G, *et al*. Bioharness(™) multivariable monitoring device: Part. II: reliability. *J Sports Sci Med* 2012;11:409–17.

Self-reported ratings appear to be the best for workload measurement

Craig S Webster,[1,2] Jennifer M Weller[1,3]

DEFINING WORKLOAD AND WHY IT IS IMPORTANT TO MEASURE IT

What is workload, why does it matter and how do we measure it? Workload is commonly defined in terms of the demands required to complete a task and the ability of an individual to meet those demands.[1] Workload increases as task demand increases relative to the finite mental and physical resources an individual has at their disposal to fulfil a task. Overload occurs when the task demand exceeds an individual's resources. In the clinical context, high workload can increase stress for the clinician and can precipitate errors in action and decision-making, with obvious consequences for patient safety.[2 3] Poorly designed clinical work environments and equipment interfaces can also substantially increase workload, and long-term stress due to high workload can lead to burnout for healthcare personnel.[4–6] A better understanding of excess workload and when and how it occurs is therefore important for the quality of life of healthcare staff, the design of clinical procedures and equipment and, ultimately, for the safety of patients. However, it is clear that observable behaviours can only be a proxy measure for total workload since a substantial part of most forms of human work also involves mental activity, which cannot be directly observed. Our propensity to prefer objective or observable measures of phenomena comes from the physical sciences where the physical or chemical processes under study do not involve mental aspects. This preference for the objective was also evident during the heyday of behaviourism, only a few decades ago, where mental phenomena were shunned as a subject of scientific inquiry simply because of the difficulties involved in measuring them.[7] Accessing mental events typically required subjective self-report, and such subjective measures were not considered scientific at the time.

The systematic review paper by Almghairbi *et al*[8] is of interest because it appraises the tools currently available to measure workload in anaesthesia, including self-reported and more objective and technical measures. After screening 2313 papers, only 24 met the inclusion criteria of the review, and these papers identified six contemporary workload measurement tools. Of the six workload tools, two were based on self-report, two were based on response time to an external stimulus and two were based on physiological measures such as heart rate and pupil diameter. Due to the heterogeneity of the data presented in the included workload studies, Almghairbi *et al*[8] employed a qualitative synthesis approach to assess the evidence for different workload measures and concluded that the two self-reported workload measurement tools, namely, the Borg Workload Scale and the NASA Task Load Index (TLX), demonstrated the best evidence of validity and sensitivity in the measurement of workload in anaesthesia.

SELF-REPORT IS A GOOD MEASURE OF WORKLOAD

Perhaps, this result should not be so surprising, despite the fact that it may not satisfy those who desire a purely objective measure of workload. Any sensitive measure of workload has a lot to do with how in control, or on the other hand, how overwhelmed, an individual is feeling in relation to their current activities. As such, workload seems like a phenomenon that would privilege self-report over other measures. This is doubly the case when the potential confounders of any kind of non-self-reported measure of workload are considered. As Almghairbi *et al*[8] point out, factors such as individual differences in the tolerance of stress, or the number of cups of coffee or other substances an individual has consumed, can all make large differences in the experience of workload, but may not be reliably reflected in any kind of physiological or behavioural measures.

EXPERTISE AND WORKLOAD

Expertise is also a substantial confounding factor, one which is likely to be particularly relevant in healthcare. An expert who appears very busy may be so practised at a complex activity that they can effortlessly execute sequences of actions without experiencing especially high workload, while a novice attempting to do less may appear to be achieving little, but feel overwhelmed by it. Workload measurement therefore may be a lot like pain assessment—a reliable and valid measure may only be possible by asking an individual to self-report on some kind of quantitative scale.[9] In the case of pain, this is often a 100 mm Visual Analogue Scale with anchors of 'no pain' and 'worse possible pain'. The Borg scale and NASA-TLX both function in a similar way, by employing numeric rating scales which yield ordinal quantitative data based on self-report (although a scale broken into subscales in the case of the NASA-TLX). Hence, a self-reported measure of workload would not be biased by expertise, since the expert (or novice) would be able to accurately report their *experienced* workload level, rather than using the apparent level measured by an external behavioural or physiological proxy measure. The converse would also be true—if you plan to use a behavioural or physiological measure of workload, you should also ask participants to rate their expertise level (and any other obvious biases), so that you can correct for this in your analysis. In practical terms, the method used will come down to the specifics of the study being conducted. Behavioural and physiological proxy measures can allow continuous recording of data, without interrupting workflow, whereas self-report requires at least momentary interruption. However, to many, the result of Almghairbi *et al*[8] review will be good news, as the Borg and NASA-TLX workload tools are the easiest of the six contemporary workload tools to administer, since they do not require any technical equipment. Borg ratings can be made verbally at intervals and recorded, while the NASA-TLX requires only pen and paper to complete.

WORKLOAD, EQUIPMENT DESIGN AND PATIENT SAFETY

An accurate tool to measure workload is important in healthcare, as it allows the demands of key tasks involved in care to

[1]Centre for Medical and Health Science Education, University of Auckland, Auckland, New Zealand
[2]Department of Anaesthesiology, University of Auckland, Auckland, New Zealand
[3]Department of Anaesthesia, Auckland City Hospital, Auckland, New Zealand

Correspondence to Dr Craig S Webster, Centre for Medical and Health Sciences Education, University of Auckland, Auckland 1142, New Zealand; c.webster@auckland.ac.nz

be quantified and better understood. As touched on by Almghairbi *et al*[8] in their paper, workload measurement has a range of applications in many research areas, particularly systems design and patient safety. It is important to determine whether changes to equipment intended to add functionality or to improve efficiency have not instead inadvertently increased workload demands on clinicians, by making devices overly complex and so more difficult to use safely.[4 10] For example, modern anaesthetic machines, patient monitors and infusion pumps provide considerably more complex functionality than previous models, often requiring the anaesthetist to navigate through nested menus and to be aware of multiple functional modes in which the device will behave in quite different ways.[4 5] Searching for the desired patient data on a screen with a reconfigurable display, or for the correct menu option, or for some indication of which mode the device is currently in, can consume a considerable amount of an anaesthetist's time. This can add complexity and workload in ways that may not make up for the advantages that the new device features were intended to offer.[11] Eye tracking techniques may be used effectively in such circumstances as a proxy measure for workload. Such techniques allow a continuous measurement of the location of eye fixation and hence locate the parts of devices or their interfaces which may be involved in excessive task demand (but caveats about the mental aspects of work and the biases of proxy measures apply).[12–14]

MEASURING WORKLOAD IN TEAMS

Workload research in healthcare remains relatively new compared with other industries, and so robust measurement tools are welcome.[15–17] However, current workload measures remain a function of an individual engaged in work, and so many fascinating questions remain to be answered in terms of how workload operates at the level of the clinical team. For example, are there aspects of clinical teams which can adversely increase individual and collective workload? Do members of well-functioning teams have lower workload measures simply because everyone works so well together and tasks progress smoothly? You might expect that members of an operating room team who work together frequently would have a substantially aligned understanding of how key tasks in common procedures should proceed. In fact, in a recent high-fidelity simulation study, our group found that clinical team members agreed, on average, only 87% of the time about when key tasks should be done, and only 70% of the time about who should do them.[18] It would seem to make intuitive sense that teams would experience higher workload if they were in relatively low agreement on how tasks should proceed, as low agreement would be expected to lead to delays and duplication of efforts—but we currently do not know the answer to this question.

CONCLUSION

After screening 2313 papers in their systematic review, Almghairbi *et al*[8] conclude that the weight of current evidence suggests that two self-reported measures, namely, the Borg Workload Scale and the NASA-TLX demonstrate the best validity and sensitivity in the measurement of workload in anaesthesia. Workload may be a phenomenon that uniquely privileges self-report, since so much of human work involves mental activity which cannot be objectively measured. However, more objective proxy measures of workload such as those based on behavioural or physiological measurements can be useful in certain experimental circumstances. For example, better understanding the task demands involved in the use of modern anaesthetic monitors is an activity that would be suitable to the use of eye tracking techniques as a proxy measure of workload. The future of workload research in healthcare more generally may involve moving beyond workload assessments of individuals doing work, to workload assessments of dynamic clinical teams doing collective work. Such developments not only will involve objective behavioural and physiological proxy measures of workload but will also likely involve the development of new self-report measures of the direct experience of workload.

Contributors Both authors jointly wrote the editorial.

Competing interests None declared.

Provenance and peer review Commissioned; internally peer reviewed.

To cite Webster CS, Weller JM. *BMJ Stel* 2018;**4**:124–125.

Received 25 March 2018
Accepted 27 March 2018
Published Online First 20 April 2018

Linked

► http://dx.doi.org/10.1136/bmjstel-2017-000263

BMJ Stel 2018;**4**:124–125.
doi:10.1136/bmjstel-2018-000330

REFERENCES

1 Younng MS, Stanton NA. Mental workload: theory, measurement and application. In: Karwowski W, ed. *International encyclopedia of ergonomics and human factors*. London: Taylor & Francis, 2001:507–9.
2 Pfeffer S, Maier T, Stricker E, *et al*. Cognitive ergonomics and informatory load in anesthesia. *Biomedical Engineering / Biomedizinische Technik* 2012;57:947–50.
3 Ahmed A, Chandra S, Herasevich V, *et al*. The effect of two different electronic health record user interfaces on intensive care provider task load, errors of cognition, and performance. *Crit Care Med* 2011;39:1626–34.
4 Wachter R. *The digital doctor – hope, hype, and harm at the dawn of medicine's computer age*. New York: McGraw Hill, 2015.
5 Webster CS. Health care technology, the human–machine interface, and patient safety during intravenous anesthesia. In: Absalom AR, Mason KP, eds. *Total intravenous anesthesia and target controlled infusions: a comprehensive global anthology*. Cham, Switzerland: Springer International, 2017.
6 Shirom A, Nirel N, Vinokur AD. Overload, autonomy, and burnout as predictors of physicians' quality of care. *J Occup Health Psychol* 2006;11:328–42.
7 Bolles RC. Occam's razor and the science of behavior. *Psychol Rep* 1957;3:321–4.
8 Almghairbi DS, Marufu TC, Moppett IK. Anaesthesia workload measurement devices: qualitative systematic review. *BMJ Stel* 2018;4:119–123.
9 Hill CS. OW! The paradox of pain. In: *Meaning, mind, and knowledge*. Oxford: Oxford University Press, 2014:155–76.
10 Dalley P, Robinson B, Weller J, *et al*. The use of high-fidelity human patient simulation and the introduction of new anesthesia delivery systems. *Anesth Analg* 2004;99:1737–41.
11 Tenner E. *Why things bite back - technology and the revenge of unintended consequences*. New York: Vintage Books, 1997.
12 Schulz CM, Schneider E, Fritz L, *et al*. Eye tracking for assessment of workload: a pilot study in an anaesthesia simulator environment. *Br J Anaesth* 2011;106:44–50.
13 Erridge S, Ashraf H, Dilley J, *et al*. Eye tracking research: seen through the patient's eyes. *BMJ Simulation and Technology Enhanced Learning* 2016;2:101–2.
14 Ross L, Williams B, Boyle M. Defibrillation safety: an examination of paramedic perceptions using eye-tracking technology. *BMJ Simulation and Technology Enhanced Learning* 2015;1:62–6.
15 Webster CS. Checklists, cognitive aids, and the future of patient safety. *Br J Anaesth* 2017;119:178–81.
16 Merry AF, Weller JM. Teamwork and minimizing error. In: Alston RP, Myles PS, Ranucci M, eds. *Oxford textbook of cardiothoracic anaesthesia*. Oxford: Oxford University Press, 2015.
17 Webster CS. Safety in unpredictable complex systems – a framework for the analysis of safety derived from the nuclear power industry. *Prometheus* 2016;34:115–32.
18 Nakarada-Kordic I, Weller JM, Webster CS, *et al*. Assessing the similarity of mental models of operating room team members and implications for patient safety: a prospective, replicated study. *BMC Med Educ* 2016;16:299.