Systematic review

Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload

R. D. Dias[1,4], M. C. Ngo-Howard[5,6], M. T. Boskovski[3,4], M. A. Zenati[4,6] and S. J. Yule[1,2,3,4]

[1]STRATUS Center for Medical Simulation, [2]Center for Surgery and Public Health, and [3]Department of Surgery, Brigham and Women's Hospital, [4]Harvard Medical School, and [5]Department of Otolaryngology – Head and Neck Surgery, Boston University School of Medicine, Boston, and [6]Medical Robotics and Computer Assisted Surgery Laboratory, Division of Cardiac Surgery, Veterans Affairs Boston Healthcare System, West Roxbury, Massachusetts, USA

Correspondence to: Dr S. J. Yule, STRATUS Center for Medical Simulation, Brigham and Women's Hospital, 10 Vining Street, Boston, Massachusetts 02115, USA (e-mail: syule@bwh.harvard.edu; @RogerDaglius; @BWH_STRATUS)

Background: **Surgeons in the operating theatre deal constantly with high-demand tasks that require simultaneous processing of a large amount of information. In certain situations, high cognitive load occurs, which may impact negatively on a surgeon's performance. This systematic review aims to provide a comprehensive understanding of the different methods used to assess surgeons' cognitive load, and a critique of the reliability and validity of current assessment metrics.**

Methods: **A search strategy encompassing MEDLINE, Embase, Web of Science, PsycINFO, ACM Digital Library, IEEE Xplore, PROSPERO and the Cochrane database was developed to identify peer-reviewed articles published from inception to November 2016. Quality was assessed by using the Medical Education Research Study Quality Instrument (MERSQI). A summary table was created to describe study design, setting, specialty, participants, cognitive load measures and MERSQI score.**

Results: **Of 391 articles retrieved, 84 met the inclusion criteria, totalling 2053 unique participants. Most studies were carried out in a simulated setting (59 studies, 70 per cent). Sixty studies (71 per cent) used self-reporting methods, of which the NASA Task Load Index (NASA-TLX) was the most commonly applied tool (44 studies, 52 per cent). Heart rate variability analysis was the most used real-time method (11 studies, 13 per cent).**

Conclusion: **Self-report instruments are valuable when the aim is to assess the overall cognitive load in different surgical procedures and assess learning curves within competence-based surgical education. When the aim is to assess cognitive load related to specific operative stages, real-time tools should be used, as they allow capture of cognitive load fluctuation. A combination of both subjective and objective methods might provide optimal measurement of surgeons' cognition.**

Paper accepted 17 November 2017
Published online 21 February 2018 in Wiley Online Library (www.bjs.co.uk). **DOI:** 10.1002/bjs.10795

Introduction

In the past two decades, great effort has been devoted to assess surgeons' intraoperative performance and pinpoint the many factors that may either improve, or impair surgical care[1–4]. In certain situations, high demands imposed by surgical tasks and other factors, such as teaching and flow disruption, may exceed surgeons' cognitive capacity, leading to a potentially risky cognitive overload. Many studies have demonstrated a direct relationship between surgical performance metrics and patient outcomes[5–7]. Although there are several methods for assessing cognitive performance, the majority of existing tools are administered *post hoc* and do not allow recognition or correction of a surgeon's performance in real-time[8]. One barrier to implementing these tools to support surgical performance is a lack of understanding of the evidence supporting their use.

Regardless of how competent and expert surgeons may be, they are still subject to the common cognitive limitations, frailties and fallibilities that characterize the human brain. The sensory memory system can receive and process a vast amount of visual and auditory information. Modulated by attention and situational awareness, incoming information is allocated into working memory, which organizes it to be stored efficiently in long-term memory, and retrieved when required. Although the human brain theoretically has unlimited capacity, working memory is able to process only a limited amount of information simultaneously[9–11]. Cognitive load is a contemporary scientific term that encapsulates a wide variety of nomenclatures used to describe the phenomenon of working memory use: cognitive workload, mental strain and mental effort. Cognitive load theory assumes that as surgeons engage in intraoperative tasks a certain level of load commensurate with required attention is imposed on working memory. A hallmark of surgical expertise is the ability to balance operative task demands against the available cognitive resources[12]. However, surgeons occasionally experience a state of cognitive overload; previous research has shown that this impairs performance during specific tasks, especially those involving complex and/or non-routine operative situations such as emergencies and unexpected events[13–15].

Measurement of cognitive load is also important for surgical education. More experienced surgeons and senior residents can process larger amounts of information concurrently than novices. Even when residents reach the plateau of the learning curve, operating similarly to established surgeons in terms of motion economy, task length, error rate and non-technical skills, they require a higher cognitive load to execute the same tasks[13,16,17]. Continued practice may enhance automaticity, and by using fewer cognitive resources, surgeons can dedicate sufficient working memory to deal with high-demand situations that may occur during surgery. Measuring surgeons' cognitive load in addition to the current performance metrics may be a useful strategy towards supporting and enhancing surgeons' cognitive capability.

Despite the large amount of research measuring cognitive load in healthcare and other high-risk industries[10,18], no systematic review has been performed to explore the existing cognitive load measurement methods in surgery, and evaluate the quality of evidence. Before systems can be developed to support cognitive performance in the operating theatre, an understanding of the range of available cognitive load assessment tools and evidence for their reliability and validity in surgery are needed.

Methods

The PRISMA guidelines[19,20] were used to design the present study and report the review findings.

Search strategy and data source

A literature search was conducted in December 2016, using MEDLINE (PubMed), Embase (embase.com), Web of Science (Core Collection), PsycINFO (EBSCO), ACM Digital Library (Guide to Computing Literature), IEEE Xplore (Digital Library), International Prospective Register of Systematic Reviews (PROSPERO) and Cochrane Database of Systematic Reviews (EBSCO) databases. All studies published from inception to November 2016 were considered and no restriction was imposed regarding language or study design. The Medical Subject Headings (MeSH) terms and text words from the MEDLINE search strategy (*Table S1*, supporting information) were adapted to other databases according to the specific syntax required. In addition, a hand search of references cited in the studies and reviews was conducted to ensure literature saturation.

Selection process and data extraction

Only original articles published in peer-reviewed journals were considered. Studies were included that used at least one cognitive load measurement involving senior surgeons, surgical trainees or medical students performing surgical procedures. Studies that did not assess cognitive load related to the intraoperative phase and studies that measured physiological data without intent to infer relationship with cognitive load were excluded. Two authors screened title and abstract for all search results independently, and identified relevant articles based on the eligibility criteria. For these articles, the full text was read by both investigators independently, who then decided whether the study met the inclusion criteria for the systematic review. The reason for excluding articles after full-text reading was registered. Using standard forms created in the REDCap web-based platform[21], two authors extracted specific data from each study included in the systematic review independently. In case of disagreement, a third author was consulted.

Assessment of data quality and synthesis

A qualitative narrative synthesis was performed, structured around different methods used to assess cognitive load in surgeons. The methodological quality of all included studies was assessed independently by two authors using the standard Medical Education Research Study Quality Instrument (MERSQI)[22]. This is a ten-item instrument that reflects six domains of research quality: study design, sampling, data type, validity of assessments, data analysis and outcomes. Each domain scores out of 3 and the maximum total MERSQI score is 18. The mean MERSQI score of both independent assessors for each article that met inclusion criteria was reported.

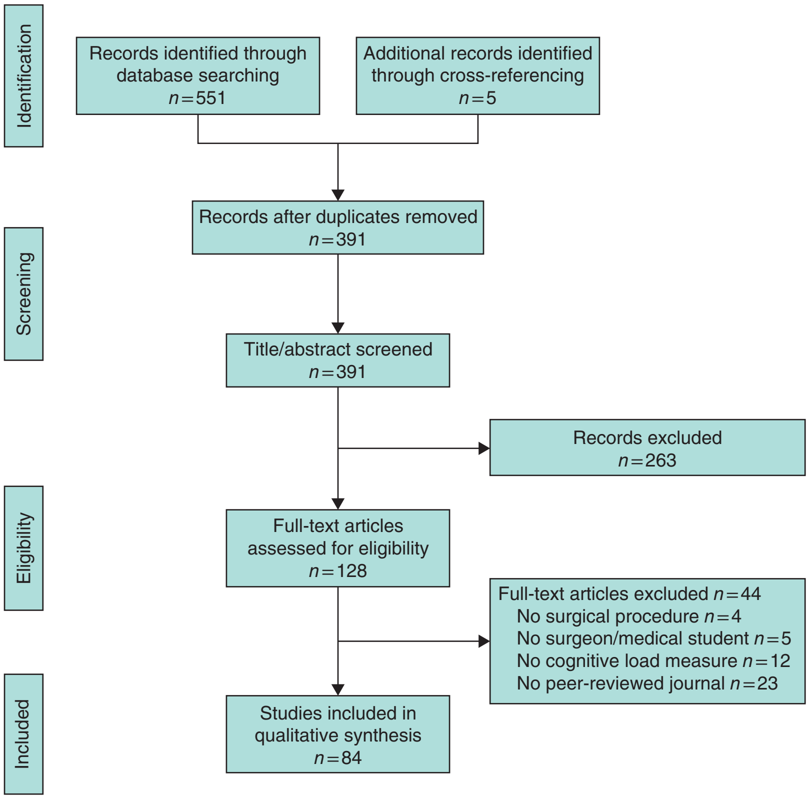


Fig. 1 PRISMA flow diagram for the review

Results

A total of 551 articles were captured in the broad search strategy, involving eight different databases, and five additional articles were identified through hand search. From these, 391 met eligibility criteria during screening and 84 met inclusion criteria after full-text review (*Fig. 1*). Pooling of data in a meta-analysis was not carried out owing to the heterogeneity regarding assessment methods and outcome measures.

Study design and setting

Fifty studies (60 per cent) were prospective cohorts, 18 (21 per cent) cross-sectional and 16 (19 per cent) RCTs. Most studies were carried out in simulated settings (59 studies, 70 per cent), 24 (29 per cent) occurred in clinical settings and only one study (1 per cent)[23] assessed surgeons' cognitive load in both settings.

Cognitive load assessment tools

The cognitive load measurement methods used in the reviewed studies are shown in *Table 1*. Most studies (58, 69 per cent) used only one tool to measure surgeons' cognitive load, and the remaining 26 (31 per cent) applied two or more methods.

Self-report tools

Sixty studies (71 per cent) used subjective (self-report) measurements of surgeons' cognitive load. The most commonly used self-report instrument was the NASA Task Load Index (NASA-TLX), used in 44 (52 per cent) of studies. This is a multidimensional assessment tool that has been used in a wide variety of domains, such as healthcare, aviation and other high-risk industries. Self-perceived cognitive load is rated from 0 to 100 points (with 5-point steps) according to six different subscales: mental demand, physical demand, temporal demand, performance, effort and frustration. The overall workload is calculated by weighting, adding or averaging each domain rating. The NASA-TLX instrument can be administered verbally, using a paper and pencil version, or by a computer-based application. Despite more than 550 studies using NASA-TLX reported in the past 20 years, few have proposed the workload redline when using this tool – a point on the scale that indicates when the workload is considered so high that it may affect human

Table 1 Different methods used to assess surgeons' cognitive load

	No. of studies
Self-reported (*post hoc*)	
NASA Task Load Index (NASA-TLX)	44
Surgery Task Load Index (SURG-TLX)	7
Subjective Mental Effort Questionnaire (SMEQ)	2
Multiple Resource Questionnaire (MRQ)	3
Pass scale	2
Borg scale	1
Subjective Workload Assessment Technique (SWAT)	1
Human Factors Evaluation Questionnaire for Computer Assisted Surgery Systems (HFEQ-CASS)	1
Other non-validated questionnaires	3
Real-time	
Physiological parameters	
Heart rate variability	11
Eye-tracking (blink rate)	5
Eye-tracking (gaze/fixation)	5
Eye-tracking (pupil dilation)	2
Electroencephalography	4
Functional near-infrared spectroscopy	3
Skin conductance response	3
Electromyography (masseter tone)	2
Heat flux (facial temperature)	1
Secondary task analysis	
Reaction time	8
Visual detection rate	4
Task precision	1
Written task	1

performance[24]. A modified version of NASA-TLX was developed and validated specifically to capture the surgical context, the Surgery Task Load Index (SURG-TLX)[25–31]. In two of the reviewed studies[13,32], the NASA-TLX questionnaire was used to capture surgeons' cognitive load over repeated training sessions, and correlated with technical performance and errors. In several other studies, self-report tools were used to compare the cognitive load imposed by different surgical procedures (open, minimally invasive and robotic surgery)[33–35], as well as different training strategies (for example simulation, cadaveric models, lectures, video-based and virtual reality)[36–38].

Real-time tools

Thirty-eight studies (45 per cent) used objective and real-time measurements of surgeons' cognitive load. The most commonly used real-time measure was heart rate variability (HRV) analysis, used in 11 studies (13 per cent). These tools are often applied in basic and clinical research studies and can be used in an unobtrusive manner using an inexpensive, wearable device, such as a Bluetooth® (Kirkland, Washington, USA) chest strap sensor or smart watch. Recent models encompassing neurovisceral integration and polyvagal theory have proposed a unifying framework, suggesting that a common reciprocal inhibitory cortical–subcortical neural circuit serves as the structural link between emotional regulation, cognitive regulation and physiological processes, and that this circuit can be indexed with HRV[39,40]. In fact, neural network studies in humans have reported increased activity in the prefrontal cortex, linking vagally mediated HRV to a set of neural structures implicated in working memory tasks[41,42]. HRV metrics are based on the analysis of interbeat intervals (R-R intervals) and attempts to quantify the sinoatrial rhythm variability. These variability measures are divided in two broad categories: time domain and frequency domain parameters, and, in order to extract these parameters from R-R intervals, complex time series statistical methods (such as spectral analysis) are used. Two HRV parameters, low-frequency (LF) band and high-frequency (HF) band, are mostly used to reflect the balance between the sympathetic and parasympathetic autonomic nervous system. In situations imposing a high cognitive demand, there is a sympathetic predominance, increasing the LF/HF ratio. LF/HF ratio has been used as an objective and real-time measure of cognitive load[43–45].

Participants

The number of participants included in the reviewed studies varied from a minimum of one to a maximum of 192, totalling 2053 unique participants. Twenty-three studies (27 per cent) assessed the cognitive load of medical students, 18 (21 per cent) assessed senior surgeons and 15 (18 per cent) assessed surgical trainees. Twenty-eight studies (33 per cent) included a combination of these grades. Only five studies (6 per cent) assessed the cognitive load of multiple operating theatre team members besides the surgeon. Team members included registered nurses (3 studies)[30,46,47], anaesthetists (3 studies)[26,30,47], registered nurse anaesthetists (2 studies)[30,46], theatre nurses (2 studies)[26,46], scrub nurses (2 studies)[30,48], physician assistants (1 study)[48], surgical technicians (2 studies)[30,46], circulating nurses (1 study)[30] and perfusionists (1 study)[46].

Surgical specialties

In total, 14 surgical specialties/subspecialties were included (*Fig. 2*). Seventy-six studies included only one specialty, two studies included two specialties, and six studies included three or more specialties (*Fig. 3*).

Study quality

The MERSQI score varied from a minimum of 6·5 to a maximum of 16·4 points, with a mean(s.d.) score of

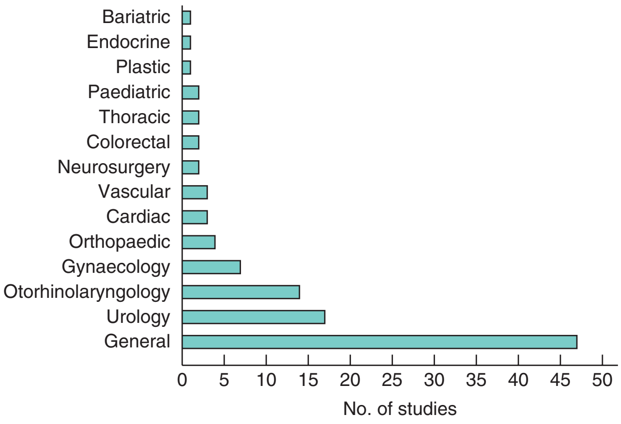


Fig. 2 Distribution of the studies according to surgical specialties

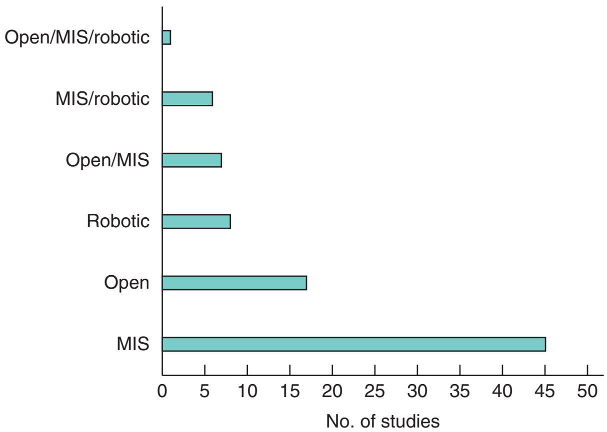


Fig. 3 Distribution of the studies according to surgical procedures. MIS, minimally invasive surgery

11·1(2·2) points. In general, studies presenting a higher MERSQI score were those that used an RCT design. The main characteristics of all included studies, as well as the MERSQI score for each are shown in *Table S2* (supporting information)[15–17,23,25–38,44–109].

Relationship with surgical outcomes

Four studies[45,89,94,98] measured actual patient outcomes, but only one[89] of these investigated the relationship between surgeons' cognitive load and patient outcomes. In that study[89], involving laparoscopic bariatric surgery, the authors reported a significant correlation between NASA-TLX score, duration of the operation and blood loss in the first 48 h after surgery, but no correlation between time to drain removal and duration of hospital stay.

Discussion

This systematic review has attempted to capture all relevant research to date that involved assessment of surgeons' cognitive load. The review shows that several methods have been developed in different subspecialties, surgical procedures, techniques, task complexities and training curricula. Tools could be categorized according to self-assessment or real-time, depending on the nature of implementation. Additionally, a subset of studies investigated the relationship between cognitive load and other factors intrinsically related to surgical care, as well as the impact of cognitive load on surgical performance.

Self-report tools commonly involve questionnaires administered after task completion to gather participants' recall of their cognitive effort during surgery. These subjective measures are easy to administer and can be used to track change over time, and evaluate interventions. Despite the wide use of psychometrically robust tools, there are limitations regarding concurrent validity of self-report tools[110]. In addition, these metrics reflect attentional and perceptual differences among individuals, memory[111], emotions and cognitive bias[112], so may not be the best for capturing fluctuations in surgeons' intraoperative cognitive load. For example, it is not known whether NASA-TLX scores reflect the actual or inferred cognitive load in each phase of a particular operation, and it is not clear how long a cognitive overload state must be present in order to cause impact[17]. Operations vary in both intensity and duration of cognitive load[43]. Interrupting an operation to administer a mental load questionnaire may interfere with patient care and the primary surgical task, so is not feasible. Self-report assessment tools must be administered after the procedure, which limits their sensitivity.

Real-time assessment methods applied to surgery included HRV analysis, eye-tracking, electroencephalography (EEG) and skin conductance, together classified as physiological metrics. HRV analysis was the physiological metric used most in the reviewed literature. This is in contrast to studies of absolute heart rate parameters (minimum, maximum and average heart rate), which are insensitive to fluctuations in cognitive load, and are influenced by other factors such as physical effort and psychological stress. Despite the conceptual overlap between stress and cognitive load, research from medicine and other fields provides evidence for HRV analysis as a metric that can assess cognitive workload. HRV analysis compares favourably with the standard questionnaire used to assess mental workload (NASA-TLX), and has been shown to be accurate in detecting isolated increasing mental demand in laboratory studies[41,43,85,113,114].

Skin conductance response reflects sympathetic activity and was used in three studies[14,53,93], revealing correlation with perceived mental stress, blink rate, electro-oculogram, EEG frontalis activity and intraoperative performance. Limitations of this method are that psychological stress also activates skin conductance response, so can be a confounder[84]. Eye-tracking devices have been used to assess surgeons' cognitive states by measuring eye movement and pupil dilation, providing information regarding gaze patterns, fixations, blink frequency and sympathetic activity[23]. This has been associated with the NASA-TLX score, surgeon experience, surgical modality and task complexity[53,62,72,93,107,109]. EEG has also been correlated with NASA-TLX scores[64,65,115], as well as expertise, task complexity and poor surgical performance. Similarly, functional near-infrared spectroscopy (fNIRS) has been implemented to capture activation patterns in specific brain areas during surgical tasks in three studies[52,79,80], with resulting correlations to surgical expertise and technical performance. A limitation related to eye-tracking, EEG and fNIRS tools is that a wide variation of analytical methods is applied, making it difficult to replicate or integrate findings across studies. In addition, these are obtrusive and may interfere with a surgeon's primary tasks in the operating room[97,100]. Finally, a series of secondary task studies assessed surgeons' cognitive load associated with training modality[50], surgical technique[91], task complexity[49], expertise and dexterity metrics[15]. However, the secondary task can also be a distraction from the primary task, making it unreasonable to assess surgeons' cognitive load in the operating theatre[66,91]. This technique should be reserved for research in simulated settings.

Several limitations should be considered when interpreting the results of this systematic review. First, there is a conceptual overlap in the literature regarding the terms stress, mental strain, cognitive workload and mental demand, as reflected in the measurement metrics that have been developed in the past decade. The same measurement tools have been used to assess different cognitive states, and the present review focused on cognitive workload metrics. Second, alongside the low methodological quality of included studies, there was wide variability of application and analysis of real-time tools, which compromises the generalizability and reliability of these cognitive load metrics. Third, the methodological quality assessed by MERSQI score was low for most studies, indicating that several aspects of research in this field can be improved. Although many studies used extensively validated tools (face, content and construct validity), most of them were done in simulated settings. Few studies assessed predictive validity using patient outcomes, or RCTs to isolate the cognitive demand derived by surgical tasks alone. Future studies should include larger sample sizes, standard measurement methods and patient-centred outcomes.

In the current era of increased technology and constrained work hours for surgical residents[116], cognitive load assessment may be integrated into the competence-based medical education framework, as it is correlated to surgical expertise and can capture the impact of new technologies and training modalities on cognitive load. Future innovations can be predicted by understanding the current state of the science. For example, objective measures of surgeons' cognitive load could be integrated into the current concept of the theatre Black Box[117], to enhance performance and improve safety and surgical care. Despite the wide variety of available methods to assess cognitive load, there is no reference value or normal range that can be used as a threshold to determine when a state of cognitive overload has been reached. Most studies compared cognitive load in two or more situations (such as different surgical procedures, techniques or experience level) and, therefore, are able to demonstrate only that surgeons display a higher or lower cognitive load when comparing two or more operative conditions. To advance, this relatively new research field must build on the already validated metrics to test criterion validity. Specifically, as there are now several validated methods using objective and real-time metrics, future research should investigate the cognitive load range that is associated with surgeons' performance impairment. This can be characterized as a danger zone[118] or redline for overload, and, if associated with risk of patient harm, may be used as a reference for the development of mitigation strategies towards supporting the cognitive performance of surgeons.

Acknowledgements

This study was funded by the National Institutes of Health (NIH) (grant number R01HL126896-01A1). The NIH had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; or preparation, review or approval of the manuscript.
Disclosure: The authors declare no conflict of interest.

References

1 Tavare A. Where are we with transparency over performance of doctors and institutions? *BMJ* 2012; **345**: e4464.
2 Madani A, Vassiliou MC, Watanabe Y, Al-Halabi B, Al-Rowais MS, Deckelbaum DL *et al.* What are the principles that guide behaviors in the operating room?: creating a framework to define and measure performance. *Ann Surg* 2017; **265**: 255–267.

3 Landon BE, Normand SL, Blumenthal D, Daley J. Physician clinical performance assessment: prospects and barriers. *JAMA* 2003; **290**: 1183–1189.
4 Rogers D, Lingard L, Boehler ML, Espin S, Klingensmith M, Mellinger JD *et al.* Teaching operating room conflict management to surgeons: clarifying the optimal approach. *Med Educ* 2011; **45**: 939–945.
5 Hyder JA, Roy N, Wakeam E, Hernandez R, Kim SP, Bader AM *et al.* Performance measurement in surgery through the National Quality Forum. *J Am Coll Surg* 2014; **219**: 1037–1046.
6 Christian CK, Gustafson ML, Roth EM, Sheridan TB, Gandhi TK, Dwyer K *et al.* A prospective study of patient safety in the operating room. *Surgery* 2006; **139**: 159–173.
7 Modi HN, Singh H, Orihuela-Espina F, Athanasiou T, Fiorentino F, Yang GZ *et al.* Temporal stress in the operating room: brain engagement promotes 'Coping' and disengagement prompts 'Choking'. *Ann Surg* 2017; [Epub ahead of print].
8 Maruthappu M, El-Harasis MA, Nagendran M, Orgill DP, McCulloch P, Duclos A *et al.* Systematic review of methodological quality of individual performance measurement in surgery. *Br J Surg* 2014; **101**: 1491–1498.
9 Young JQ, Van Merrienboer J, Durning S, Ten Cate O. Cognitive load theory: implications for medical education: AMEE guide no. 86. *Med Teach* 2014; **36**: 371–384.
10 Van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ* 2010; **44**: 85–93.
11 Miller GA. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol Rev* 1956; **63**: 81–97.
12 Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci* 1988; **12**: 257–285.
13 Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D. Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simul Healthc* 2010; **5**: 267–271.
14 Schuetz M, Gockel I, Beardi J, Hakman P, Dunschede F, Moenk S *et al.* Three different types of surgeon-specific stress reactions identified by laparoscopic simulation in a virtual scenario. *Surg Endosc* 2008; **22**: 1263–1267.
15 Haji FA, Rojas D, Childs R, de Ribaupierre S, Dubrowski A. Measuring cognitive load: performance, mental effort and simulation task complexity. *Med Educ* 2015; **49**: 815–827.
16 Eversbusch A, Grantcharov TP. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. *Surg Endosc* 2004; **18**: 1514–1518.
17 Carswell CM, Clarke D, Seales WB. Assessing mental workload during laparoscopic surgery. *Surg Innov* 2005; **12**: 80–90.
18 Naismith LM, Cavalcanti RB. Validity of cognitive load measures in simulation-based training: a systematic review. *Acad Med* 2015; **90**(Suppl): S24–S35.
19 Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009; **339**: b2535.
20 Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015; **349**: g7647.
21 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42**: 377–381.
22 Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA* 2007; **298**: 1002–1009.
23 Richstone L, Schwartz MJ, Seideman C, Cadeddu J, Marshall S, Kavoussi LR. Eye metrics as an objective assessment of surgical skill. *Ann Surg* 2010; **252**: 177–182.
24 Hart SG. NASA-Task Load Index (NASA-TLX); 20 years later. *Proc Hum Factors Ergon Soc Annu Meet* 2006; **50**: 904–908.
25 Wilson MR, Poolton JM, Malhotra N, Ngo K, Bright E, Masters RS. Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). *World J Surg* 2011; **35**: 1961–1969.
26 Weigl M, Antoniadis S, Chiapponi C, Bruns C, Sevdalis N. The impact of intra-operative interruptions on surgeons' perceived workload: an observational study in elective general and orthopedic surgery. *Surg Endosc* 2015; **29**: 145–153.
27 Berg RJ, Inaba K, Sullivan M, Okoye O, Siboni S, Minneti M *et al.* The impact of heat stress on operative performance and cognitive function during simulated laparoscopic operative tasks. *Surgery* 2015; **157**: 87–95.
28 Weigl M, Stefan P, Abhari K, Wucherer P, Fallavollita P, Lazarovici M *et al.* Intra-operative disruptions, surgeon's mental workload, and technical performance in a full-scale simulated procedure. *Surg Endosc* 2016; **30**: 559–566.
29 Yu D, Dural C, Morrow MM, Yang L, Collins JW, Hallbeck S *et al.* Intraoperative workload in robotic surgery assessed by wearable motion tracking sensors and questionnaires. *Surg Endosc* 2017; **31**: 877–886.
30 Yu D, Lowndes B, Thiels C, Bingener J, Abdelrahman A, Lyons R *et al.* Quantifying intraoperative workloads across the surgical team roles: room for better balance? *World J Surg* 2016; **40**: 1565–1574.
31 Wucherer P, Stefan P, Abhari K, Fallavollita P, Weigl M, Lazarovici M *et al.* Vertebroplasty performance on simulator for 19 surgeons using hierarchical task analysis. *IEEE Trans Med Imaging* 2015; **34**: 1730–1737.

32 Zheng B, Fu B, Al-Tayeb TA, Hao YF, Qayumi AK. Mastering instruments before operating on a patient: the role of simulation training in tool use skills. *Surg Innov* 2014; **21**: 637–642.
33 Bertolaccini L, Viti A, Terzi A. Ergon-trial: ergonomic evaluation of single-port access *versus* three-port access video-assisted thoracic surgery. *Surg Endosc* 2015; **29**: 2934–2940.
34 Dixon BJ, Daly MJ, Chan H, Vescan A, Witterick IJ, Irish JC. Augmented real-time navigation with critical structure proximity alerts for endoscopic skull base surgery. *Laryngoscope* 2014; **124**: 853–859.
35 Foo JL, Martinez-Escobar M, Juhnke B, Cassidy K, Hisley K, Lobe T *et al.* Evaluating mental workload of two-dimensional and three-dimensional visualization for anatomical structure localization. *J Laparoendosc Adv Surg Tech A* 2013; **23**: 65–70.
36 Britt RC, Scerbo MW, Montano M, Kennedy RA, Prytz E, Stefanidis D. Intracorporeal suturing: transfer from Fundamentals of Laparoscopic Surgery to cadavers results in substantial increase in mental workload. *Surgery* 2015; **158**: 1428–1433.
37 Chowriappa A, Raza SJ, Fazili A, Field E, Malito C, Samarasekera D *et al.* Augmented-reality-based skills training for robot-assisted urethrovesical anastomosis: a multi-institutional randomised controlled trial. *BJU Int* 2015; **115**: 336–345.
38 Gardner AK, Clanton J, Jabbour II, Scott L, Scott DJ, Russo MA. Impact of seductive details on the acquisition and transfer of laparoscopic suturing skills: emotionally interesting or cognitively taxing? *Surgery* 2016; **160**: 580–585.
39 Porges SW. The polyvagal perspective. *Biol Psychol* 2007; **74**: 116–143.
40 Thayer JF, Lane RD. Claude Bernard and the heart–brain connection: further elaboration of a model of neurovisceral integration. *Neurosci Biobehav Rev* 2009; **33**: 81–88.
41 Gianaros PJ, Van Der Veen FM, Jennings JR. Regional cerebral blood flow correlates with heart period and high-frequency heart period variability during working-memory tasks: implications for the cortical and subcortical regulation of cardiac autonomic activity. *Psychophysiology* 2004; **41**: 521–530.
42 Thayer JF, Hansen AL, Saus-Rose E, Johnsen BH. Heart rate variability, prefrontal neural function, and cognitive performance: the neurovisceral integration perspective on self-regulation, adaptation, and health. *Ann Behav Med* 2009; **37**: 141–153.
43 Rieger A, Stoll R, Kreuzfeld S, Behrens K, Weippert M. Heart rate and heart rate variability as indirect markers of surgeons' intraoperative stress. *Int Arch Occup Environ Health* 2014; **87**: 165–174.
44 Rieger A, Fenger S, Neubert S, Weippert M, Kreuzfeld S, Stoll R. Psychophysical workload in the operating room: primary surgeon *versus* assistant. *Surg Endosc* 2015; **29**: 1990–1998.
45 Böhm B, Rötting N, Schwenk W, Grebe S, Mansmann U. A prospective randomized trial on heart rate variability of the surgical team during laparoscopic and conventional sigmoid resection. *Arch Surg* 2001; **136**: 305–310.
46 Wadhera RK, Parker SH, Burkhart HM, Greason KL, Neal JR, Levenick KM *et al.* Is the 'sterile cockpit' concept applicable to cardiovascular surgery critical intervals or critical events? The impact of protocol-driven communication during cardiopulmonary bypass. *J Thorac Cardiovasc Surg* 2010; **139**: 312–319.
47 Wheelock A, Suliman A, Wharton R, Babu ED, Hull L, Vincent C *et al.* The impact of operating room distractions on stress, workload, and teamwork. *Ann Surg* 2015; **261**: 1079–1084.
48 Tiferes J, Hussein AA, Bisantz A, Kozlowski JD, Sharif MA, Winder NM *et al.* The loud surgeon behind the console: understanding team activities during robot-assisted surgery. *J Surg Educ* 2016; **73**: 504–512.
49 Andersen SA, Konge L, Cayé-Thomasen P, Sørensen MS. Retention of mastoidectomy skills after virtual reality simulation training. *JAMA Otolaryngol Head Neck Surg* 2016; **142**: 635–640.
50 Andersen SA, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen MS. Cognitive load in mastoidectomy skills training: virtual reality simulation and traditional dissection compared. *J Surg Educ* 2016; **73**: 45–50.
51 Andersen SA, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen MS. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *Laryngoscope* 2016; **126**: E74–E79.
52 Andreu-Perez J, Leff DR, Shetty K, Darzi A, Yang GZ. Disparity in frontal lobe connectivity on a complex bimanual motor task aids in classification of operator skill level. *Brain Connect* 2016; **6**: 375–388.
53 Berguer R, Smith WD, Chung YH. Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surg Endosc* 2001; **15**: 1204–1207.
54 Bharathan R, Vali S, Setchell T, Miskry T, Darzi A, Aggarwal R. Psychomotor skills and cognitive load training on a virtual reality laparoscopic simulator for tubal surgery is effective. *Eur J Obstet Gynecol Reprod Biol* 2013; **169**: 347–352.
55 Bjerrum F, Sorensen JL, Konge L, Lindschou J, Rosthøj S, Ottesen B *et al.* Procedural specificity in laparoscopic simulator training: protocol for a randomised educational superiority trial. *BMC Med Educ* 2014; **14**: 215.
56 Carswell CM, Lio CH, Grant R, Klein MI, Clarke D, Seales WB *et al.* Hands-free administration of subjective workload scales: acceptability in a surgical training environment. *Appl Ergon* 2010; **42**: 138–145.
57 Czyzewska E, Kiczka K, Czarnecki A, Pokinko P. The surgeon's mental load during decision making at various stages of operations. *Eur J Appl Physiol Occup Physiol* 1983; **51**: 441–446.

58 Demirtas Y, Tulmac M, Yavuzer R, Yalcin R, Ayhan S, Latifoglu O *et al.* Plastic surgeon's life: marvelous for mind, exhausting for body. *Plast Reconstr Surg* 2004; **114**: 923–931.
59 Di Stasi LL, Diaz-Piedra C, Rieiro H, Sánchez Carrión JM, Martin Berrido M, Olivares G *et al.* Gaze entropy reflects surgical task load. *Surg Endosc* 2016; **30**: 5034–5043.
60 Dixon BJ, Daly MJ, Chan H, Vescan A, Witterick IJ, Irish JC. Augmented image guidance improves skull base navigation and reduces task workload in trainees: a preclinical trial. *Laryngoscope* 2011; **121**: 2060–2064.
61 Dixon BJ, Chan H, Daly MJ, Vescan AD, Witterick IJ, Irish JC. The effect of augmented real-time image guidance on task workload during endoscopic sinus surgery. *Int Forum Allergy Rhinol* 2012; **2**: 405–410.
62 Erol Barkana D, Açık A, Duru DG, Duru AD. Improvement of design of a surgical interface using an eye tracking device. *Theor Biol Med Model* 2014; **11**(Suppl 1): S4.
63 Gómez-Gómez E, Carrasco-Valiente J, Valero-Rosa J, Campos-Hernández JP, Anglada-Curado FJ, Carazo-Carazo JL *et al.* Impact of 3D vision on mental workload and laparoscopic performance in inexperienced subjects. *Actas Urol Esp* 2015; **39**: 229–235.
64 Guru KA, Esfahani ET, Raza SJ, Bhat R, Wang K, Hammond Y *et al.* Cognitive skills assessment during robot-assisted surgery: separating the wheat from the chaff. *BJU Int* 2015; **115**: 166–174.
65 Guru KA, Shafiei SB, Khan A, Hussein AA, Sharif M, Esfahani ET. Understanding cognitive performance during robot-assisted surgery. *Urology* 2015; **86**: 751–757.
66 Haji FA, Khan R, Regehr G, Drake J, de Ribaupierre S, Dubrowski A. Measuring cognitive load during simulation-based psychomotor skills training: sensitivity of secondary-task performance and subjective ratings. *Adv Health Sci Educ Theory Pract* 2015; **20**: 1237–1253.
67 Hedman L, Klingberg T, Enochsson L, Kjellin A, Felländer-Tsai L. Visual working memory influences the performance in virtual image-guided surgical intervention. *Surg Endosc* 2007; **21**: 2044–2050.
68 Heemskerk J, Zandbergen HR, Keet SW, Martijnse I, van Montfort G, Peters RJ, *et al.* Relax, it's just laparoscopy! A prospective randomized trial on heart rate variability of the surgeon in robot-assisted *versus* conventional laparoscopic cholecystectomy. *Dig Surg* 2014; **31**: 225–232.
69 Horner RD, Szaflarski JP, Jacobson CJ, Elder N, Bolon S, Matthews G *et al.* Clinical work intensity among physician specialties: how might we assess it? What do we find? *Med Care* 2011; **49**: 108–113.
70 Hu JS, Lu J, Tan WB, Lomanto D. Training improves laparoscopic tasks performance and decreases operator workload. *Surg Endosc* 2016; **30**: 1742–1746.
71 Hussein AA, Shafiei SB, Sharif M, Esfahani E, Ahmad B, Kozlowski JD *et al.* Technical mentorship during robot-assisted surgery: a cognitive analysis. *BJU Int* 2016; **118**: 429–436.
72 Jiang X, Zheng B, Tien G, Atkins MS. Pupil response to precision in surgical task execution. *Stud Health Technol Inform* 2013; **184**: 210–214.
73 Joseph B, Parvaneh S, Swartz T, Haider AA, Hassan A, Kulvatunyou N *et al.* Stress among surgical attending physicians and trainees: a quantitative assessment during trauma activation and emergency surgeries. *J Trauma Acute Care Surg* 2016; **81**: 723–728.
74 Klein MI, Warm JS, Riley MA, Matthews G, Doarn C, Donovan JF *et al.* Mental workload and stress perceived by novice operators in the laparoscopic and robotic minimally invasive surgical interfaces. *J Endourol* 2012; **26**: 1089–1094.
75 Koca D, Yıldız S, Soyupek F, Günyeli İ, Erdemoglu E, Soyupek S. Physical and mental workload in single-incision laparoscopic surgery and conventional laparoscopy. *Surg Innov* 2015; **22**: 294–302.
76 Kopac DS, Chen J, Tang R, Sawka A, Vaghadia H. Comparison of a novel real-time SonixGPS needle-tracking ultrasound technique with traditional ultrasound for vascular access in a phantom gel model. *J Vasc Surg* 2013; **58**: 735–741.
77 Kurahashi AM, Harvey A, MacRae H, Moulton CA, Dubrowski A. Technical skill training improves the ability to learn. *Surgery* 2011; **149**: 1–6.
78 Lee GI, Lee MR, Clanton T, Sutton E, Park AE, Marohn MR. Comparative assessment of physical and cognitive ergonomics associated with robotic and traditional laparoscopic surgeries. *Surg Endosc* 2014; **28**: 456–465.
79 Leff DR, Elwell CE, Orihuela-Espina F, Atallah L, Delpy DT, Darzi AW *et al.* Changes in prefrontal cortical behaviour depend upon familiarity on a bimanual co-ordination task: an fNIRS study. *Neuroimage* 2008; **39**: 805–813.
80 Leff DR, James DR, Orihuela-Espina F, Kwok KW, Sun LW, Mylonas G *et al.* The impact of expert visual guidance on trainee visual search strategy, visual attention and motor skills. *Front Hum Neurosci* 2015; **9**: 526.
81 Li L, Yang J, Chu Y, Wu W, Xue J, Liang P *et al.* A novel augmented reality navigation system for endoscopic sinus and skull base surgery: a feasibility study. *PLoS One* 2016; **11**: e0146996.
82 Mouraviev V, Klein M, Schommer E, Thiel DD, Samavedi S, Kumar A *et al.* Urology residents experience comparable workload profiles when performing live porcine nephrectomies and robotic surgery virtual reality training modules. *J Robot Surg* 2016; **10**: 49–56.
83 Muresan C, Lee TH, Seagull J, Park AE. Transfer of training in the development of intracorporeal suturing skill in medical student novices: a prospective randomized trial. *Am J Surg* 2010; **200**: 537–541.
84 Pluyter JR, Rutkowski AF, Jakimowicz JJ. Immersive training: breaking the bubble and measuring the heat. *Surg Endosc* 2014; **28**: 1545–1554.

85 Prichard RS, O'Neill CJ, Oucharek JJ, Holmes CY, Colinda YH, Delbridge LW *et al.* A prospective study of heart rate variability in endocrine surgery: surgical training increases consultant's mental strain. *J Surg Educ* 2012; **69**: 453–458.
86 Rasmussen SR, Konge L, Mikkelsen PT, Sørensen MS, Andersen SA. Notes from the field: secondary task precision for cognitive load estimation during virtual reality surgical simulation training. *Eval Health Prof* 2016; **39**: 114–120.
87 Rieder E, Martinec DV, Cassera MA, Goers TA, Dunst CM, Swanstrom LL. A triangulating operating platform enhances bimanual performance and reduces surgical workload in single-incision laparoscopy. *J Am Coll Surg* 2011; **212**: 378–384.
88 Rojas D, Haji F, Shewaga R, Kapralos B, Dubrowski A. The impact of secondary-task type on the sensitivity of reaction-time based measurement of cognitive load for novices learning surgical skills using simulation. *Stud Health Technol Inform* 2014; **196**: 353–359.
89 Ruiz-Rabelo JF, Navarro-Rodriguez E, Di-Stasi LL, Diaz-Jimenez N, Cabrera-Bermon J, Diaz-Iglesias C *et al.* Validation of the NASA-TLX score in ongoing assessment of mental workload during a laparoscopic learning curve in bariatric surgery. *Obes Surg* 2015; **25**: 2451–2456.
90 Sørensen SM, Mahmood O, Konge L, Thinggaard E, Bjerrum F. Laser visual guidance *versus* two-dimensional vision in laparoscopy: a randomized trial. *Surg Endosc* 2017; **31**: 112–118.
91 Scerbo MW, Britt RC, Stefanidis D. Differences in mental workload between traditional and single-incision laparoscopic procedures measured with a secondary task. *Am J Surg* 2017; **213**: 244–248.
92 Sethi AS, Peine WJ, Mohammadi Y, Sundaram CP. Validation of a novel virtual reality robotic simulator. *J Endourol* 2009; **23**: 503–508.
93 Smith WD, Chung YH, Berguer R. A virtual instrument ergonomics workstation for measuring the mental workload of performing video-endoscopic surgery. *Stud Health Technol Inform* 2000; **70**: 309–315.
94 Song MH, Tokuda Y, Nakayama T, Sato M, Hattori K. Intraoperative heart rate variability of a cardiac surgeon himself in coronary artery bypass grafting surgery. *Interact Cardiovasc Thorac Surg* 2009; **8**: 639–641.
95 Stefanidis D, Wang F, Korndorffer JR, Dunne JB, Scott DJ. Robotic assistance improves intracorporeal suturing performance and safety in the operating room while decreasing operator workload. *Surg Endosc* 2010; **24**: 377–382.
96 Stelter K, Ertl-Wagner B, Luz M, Muller S, Ledderose G, Siedek V *et al.* Evaluation of an image-guided navigation system in the training of functional endoscopic sinus surgeons. A prospective, randomised clinical study. *Rhinology* 2011; **49**: 429–437.
97 Stelter K, Theodoraki MN, Becker S, Tsekmistrenko V, Olzowy B, Ledderose G. Specific stressors in endonasal skull base surgery with and without navigation. *Eur Arch Otorhinolaryngol* 2015; **272**: 631–638.
98 Strauss G, Koulechov K, Röttger S, Bahner J, Trantakis C, Hofer M *et al.* Evaluation of a navigation system for ENT with surgical efficiency criteria. *Laryngoscope* 2006; **116**: 564–572.
99 Tarr ME, Brancato SJ, Cunkelman JA, Polcari A, Nutter B, Kenton K. Comparison of postural ergonomics between laparoscopic and robotic sacrocolpopexy: a pilot study. *J Minim Invasive Gynecol* 2015; **22**: 234–238.
100 Theodoraki MN, Ledderose GJ, Becker S, Leunig A, Arpe S, Luz M *et al.* Mental distress and effort to engage an image-guided navigation system in the surgical training of endoscopic sinus surgery: a prospective, randomised clinical trial. *Eur Arch Otorhinolaryngol* 2015; **272**: 905–913.
101 Tomasko JM, Pauli EM, Kunselman AR, Haluck RS. Sleep deprivation increases cognitive workload during simulated surgical tasks. *Am J Surg* 2012; **203**: 37–43.
102 Wilhelm D, Reiser S, Kohn N, Witte M, Leiner U, Mühlbach L *et al.* Comparative evaluation of HD 2D/3D laparoscopic monitors and benchmarking to a theoretically ideal 3D pseudodisplay: even well-experienced laparoscopists perform better with 3D. *Surg Endosc* 2014; **28**: 2387–2397.
103 Yamashita J, Morikawa O, Kumagai T, Yokoyama K, Tomoda K. A distance education system for manual skills training in endoscopic sinus surgery. *Oto-Rhino-Laryngology Tokyo* 2009; **51**: 385–389.
104 Youssef Y, Lee G, Godinez C, Sutton E, Klein RV, George IM *et al.* Laparoscopic cholecystectomy poses physical injury risk to surgeons: analysis of hand technique and standing position. *Surg Endosc* 2011; **25**: 2168–2174.
105 Zheng B, Cassera MA, Martinec DV, Spaun GO, Swanström LL. Measuring mental workload during the performance of advanced laparoscopic tasks. *Surg Endosc* 2010; **24**: 45–50.
106 Zheng B, Janmohamed Z, MacKenzie CL. Reaction times and the decision-making process in endoscopic surgery. *Surg Endosc* 2003; **17**: 1475–1480.
107 Zheng B, Jiang X, Tien G, Meneghetti A, Panton ON, Atkins MS. Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surg Endosc* 2012; **26**: 2746–2750.
108 Zheng B, Rieder E, Cassera MA, Martinec DV, Lee G, Panton ON *et al.* Quantifying mental workloads of surgeons performing natural orifice transluminal endoscopic surgery (NOTES) procedures. *Surg Endosc* 2012; **26**: 1352–1358.
109 Zheng B, Tien G, Atkins SM, Swindells C, Tanin H, Meneghetti A *et al.* Surgeon's vigilance in the operating room. *Am J Surg* 2011; **201**: 673–677.
110 Naismith LM, Cheung JJ, Ringsted C, Cavalcanti RB. Limitations of subjective cognitive load measures in simulation-based procedural training. *Med Educ* 2015; **49**: 805–814.

111 Ericsson KA, Simon HA. Verbal reports as data. *Psychol Rev* 1980; **87**: 215–251.
112 Fraser K, Huffman J, Ma I, Sobczak M, McIlwrick J, Wright B *et al.* The emotional and cognitive impact of unexpected simulated patient death: a randomized controlled trial. *Chest* 2014; **145**: 958–963.
113 Wanyan X, Zhuang D, Zhang H. Improving pilot mental workload evaluation with combined measures. *Biomed Mater Eng* 2014; **24**: 2283–2290.
114 Fallahi M, Motamedzade M, Heidarimoghadam R, Soltanian AR, Miyake S. Effects of mental workload on physiological and subjective responses during traffic density monitoring: a field study. *Appl Ergon* 2016; **52**: 95–103.
115 Berka C, Levendowski DJ, Lumicao MN, Yau A, Davis G, Zivkovic VT *et al.* EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviat Space Environ Med* 2007; **78**(Suppl): B231–B244.
116 Klingensmith ME. The future of general surgery residency education. *JAMA Surg* 2016; **151**: 207–208.
117 Goldenberg MG, Jung J, Grantcharov TP. Using data to enhance performance and improve quality and safety in surgery. *JAMA Surg* 2017; **152**: 972–973.
118 Colle HA, Reid GB. Estimating a mental workload redline in a simulated air-to-ground combat mission. *Int J Aviat Psychol* 2005; **15**: 303–319.

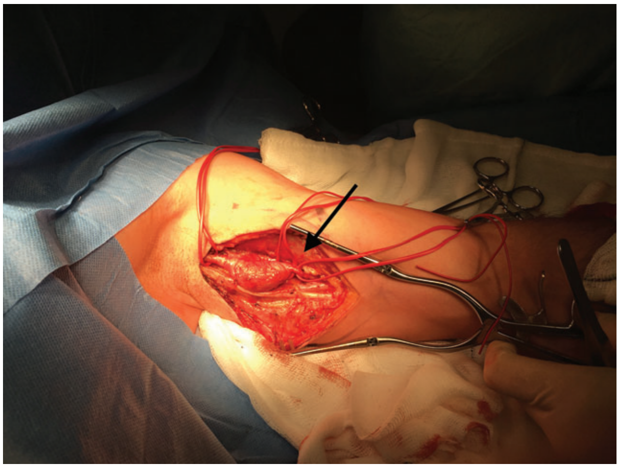
Supporting information

Additional supporting information can be found online in the supporting information tab for this article.

Snapshot quiz

Snapshot quiz 18/4

Question: What is the abnormality depicted in this brachial artery?



The answer to the above question is found on p. 519 of this issue of *BJS*.

Chan YM, Wilson A: Department of Vascular Surgery, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, AB25 2ZN, UK (e-mail: sharonchanym@gmail.com)

Snapshots in Surgery: to view submission guidelines, submit your snapshot and view the archive, please visit www.bjs.co.uk