BMJ Open

Protocol for a scoping review on 'surgical sabermetrics:' technology-enhanced measurement of operative non-technical skills

Emma Howie ,[1] Stephen J Wigmore,[1,2] Roger Daglius Dias,[3,4] Richard Skipworth,[1,2] Steven Yule[1,3]

To cite: Howie E, Wigmore SJ, Daglius Dias R, *et al*. Protocol for a scoping review on 'surgical sabermetrics:' technology-enhanced measurement of operative non-technical skills. *BMJ Open* 2023;**13**:e064196. doi:10.1136/bmjopen-2022-064196

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-064196).

Received 27 April 2022
Accepted 21 December 2022

[1]Division of Clinical and Surgical Sciences, School of Surgery, The University of Edinburgh, Edinburgh, UK
[2]Department of Surgery, NHS Lothian, Edinburgh, UK
[3]STRATUS Center for Medical Simulation, Brigham and Women's Hospital, Boston, Massachusetts, USA
[4]Department of Emergency Medicine, Harvard Medical School, Boston, Massachusetts, USA

Correspondence to
Miss Emma Howie;
ehowie@ed.ac.uk

ABSTRACT

Introduction Surgeons need high fidelity, high quality, objective, non-judgemental and quantitative feedback to measure their performance in order to optimise their performance and improve patient safety. This can be provided through surgical sabermetrics, defined as 'advanced analytics of digitally recorded surgical training and operative procedures to enhance insight, support professional development and optimise clinical and safety outcomes'. The aim of this scoping review is to investigate the assessment of surgeon's non-technical skills using sabermetrics principles, focusing on digital, automated measurements that do not require a human observer.

Methods and analysis To investigate the current methods of digital, automated measurements of surgeons' non-technical skills, a systematic scoping review will be conducted following Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews guidelines, using databases from medicine and other fields. Covidence software is used for screening of potential studies. A data extraction tool will be developed specifically for this study to evaluate the methods of measurement. Quality assurance will be assessed using Quality Assessment Tool for Diverse Designs. Multiple reviewers will be responsible for screening of studies and data extraction.

Ethics and dissemination This is a review study, not using primary data, and therefore, ethical approval is not required. A range of methods will be employed for dissemination of the results of this study, including publication in journals and conference presentations.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This scoping review identifies and evaluates current methods of digital, automated measurements of surgeons' non-technical skills.
⇒ This review is conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews.
⇒ Despite expected heterogeneity of study designs, a systematic search of relevant databases will be conducted using standard search terms and following inclusion and exclusion criteria.
⇒ The study is limited to review of manuscripts focusing on surgeons' non-technical skills and does not include assessments of other team members, such as anaesthetists or scrub nurses.
⇒ Data extraction and quality assessment of studies will be conducted by multiple reviewers to enhance validity and reliability of study findings.

INTRODUCTION

Surgery is a high-performing field with surgeons striving for exceptional performance, focusing on patient safety and good outcomes.[1] Surgery is a complex, dynamic sociotechnical process that requires the synchronisation of multiple cognitive and physical processes and the coordination of multiple technical and non-technical skills (NTS), all while interacting with tools and technology and responding to changes in the current situation. This dynamic field demands both physical and mental effort, with high demand on concentration and the ability to react and adapt to quick, unexpected changes.[2–6] Surgeons need high fidelity, high quality, objective and quantitative feedback to measure their performance in order to optimise their performance and improve patient safety. These data should be time-sensitive and lack bias or judgement. Real-time performance analysis can lead to adjustments in task load, preventing adverse outcomes and optimising performance.[7] Objective, automatically captured, digital data reduce bias and provide the feedback surgeons need for performance optimisation.

Surgical performance assessment improves training and provides quality assurance, benefiting surgeons and their patients.[8] There is a direct link between performance and clinical outcomes.[9] Surgical performance is affected by several factors relating to interlinking technical and NTS, including the surgeon, the surgical team, equipment and

Table 1 Demonstrating subcategories of non-technical skills for surgeons taxonomy and corresponding elements[10]

Category	Element
Situation awareness	Gathering information Understanding information Projecting and anticipating future state
Decision making	Considering options Selecting and communication option Implementing and reviewing decisions
Communication and teamwork	Exchanging information Establishing a sharing understanding Co-ordinating team activities
Leadership	Setting and maintaining standards Supporting others Coping with pressure

technology, the operating theatre (OR), the surgery and the patient themselves.[10–12] Additionally, surgeons are humans, fallible and subject to influence from external and internal forces, including: the busy healthcare environment, fatigue, burnout, noise, stress at home and the need to focus on outpatients, inpatients and the current operation concurrently.

NTSs are defined as the cognitive and social skills that characterise high performing individuals and teams. Current measurements of surgeons' NTS include the non-technical skills for surgeons (NOTSS) taxonomy (subcategories of which are shown in table 1),[10] but these measurements are flawed. These resource heavy tools either require retrospective reflection or else would disrupt the surgical process, or alternatively an expert rater presence in the OR, which most establishments would not have the resources to do. Additionally, these ratings are influenced by subject variability.[7]

A large proportion of tasks in surgery are related to cognition. A surgeon's NTSs consist of several cognitive processes, including situational awareness, decision making, teamwork and communication skills. Therefore, it is also important to evaluate a surgeon's cognitive load (CL) or mental workload. Increasing CL degrades performance. If they become overloaded then their NTS and subsequently technical skills can be negatively impacted, potentially leading to error and placing the patient in danger.[13] CL can be determined by physiological measurements acting as a proxy for CL.[14 15] Changes in physiology can indicate behavioural changes at various parts of a procedure, and could indicate mental strain.[16] New technology such as the OR Black Box also measures parts of the surgical performance via artificial intelligence and the collection of audio-visual data.[17]

The objective of this scoping review is to investigate the current technological advances of digitally measuring a surgeon's NTSs. This will provide evidence to develop the exciting field of surgical sabermetrics, the process of analysing 'in-game' data to evaluate surgeons' performance by providing advance analytics of digitally recorded surgical procedures to enhance and support personal development, training and optimise clinical, safety and financial outcomes.[18] Evidence from this scoping review will determine which methods should be further trialled in surgical studies.

As of writing, the authors are not aware of any current relevant reviews that focus on real-time, bias free, digital data. A recent scoping review by Cha and Yu[12] looked at objective measures of NTS, but did not specifically look at automated, digital measurements. Many studies included in this review required human expertise as part of the objective measurement. A systematic review by Levin *et al* reviewed all aspects of performance within the operating room, not only NTSs, and did not specifically look at surgeons.[19] A review by Dias *et al*[20] included objective measurements of CL but did not specifically look at the other NTS. Therefore, this review will provide a unique perspective on the objective, automated assessment of a surgeon's NTSs.

The aim of this study is to determine the most suitable method of measuring NTSs, focusing on technologies and physiological sensors providing data from individuals. These technologies include heart rate variability sensors and electroencephalograms,[20] which measure proxies of human cognitive and social behaviour. In addition to evaluating studies that have used these digital methods, we aim to assess the methods used for their feasibility and acceptability for use by surgeons.

METHODOLOGY

Methodology for this scoping review is structured using the framework outlined by Arksey and O'Malley,[21] itself further amended by Levac *et al*,[22] in addition to following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist[23] and Joanna Briggs Institute (JBI) guidance.[24] The framework consists of five steps with an optional sixth outlined in box 1. Searches and study selection are due to commence in August 2022 (stage 3) with data collection complete by January 2023 (stages 4 and 5).

Patient and public involvement

There was no direct patient involvement in the design or conduct of this study. Improving surgeon's NTSs improves

Box 1 Framework stages for developing a scoping review protocol[21 22]

Stage 1: identifying the research question.
Stage 2: identifying relevant studies.
Stage 3: study selection.
Stage 4: charting the data or data collection.
Stage 5: collating, summarising and reporting the results.
Stage 6 (optional): consultation.

Table 2 outlining PICo elements

Population	**Surgeons, trainee or full qualified, of all disciplines**	
Intervention/ phenomenon of interest	Digital measurement of non-technical skills:	
	Measurement	Non-technical skills
	Digital Objective Physiological	Situational awareness Decision making Communication skills Teamwork Leadership Cognitive load
Context	In situ or simulation Any environment in which the surgeon works, for example, operating room	

PICo, Population Intervention Context.

surgical performance, and ultimately, has a beneficial effect in improving patient safety.

Ethics and dissemination

This is a review study, not using primary data, and therefore, ethical approval is not required. A range of methods will be employed for dissemination of the results of this study, including publication in journals and conference presentations.

Stage 1: identifying the research question

The research questions for this study, 'What are the current digital methods of measuring a surgeon's NTSs?', was developed due to a gap in the literature summarising this information. Additionally, answering this question will guide further research by:

Table 3 Outlining inclusion and exclusion criteria for the scoping review

Inclusion criteria	Exclusion criteria
Relates to the review question	Unrelated data
Digital data measurement using technology	Solely analogue data
Automated or real-time data	Data collected requires human to collect
Involves surgeon (trainee or full qualified)	Involves solely other healthcare team members apart from surgeons
In situ or simulation study	Study conducted out of the operating room or simulated operating room
Original research only including the grey literature of thesis, dissertations, discussions or what papers	Systematic reviews
Published 2010 onwards	Before 2010
Printed in English	Not available in English

1. Identifying reliable methods of digital measurement to be used in further research projects.
2. Identifying gaps in the field to guide further research.

For the purpose of this review, a digital method of measurement consists of any technology that provides objective, real-time quantitative data without the need for a human to collect. For example, a smart watch that measures a surgeon's physiology.

In addition to the primary research question, the aim of this scoping review is to determine what is the best method from those identified. From the papers reviewed, we will analyse the methods, where possible, for:

- Ease of use and unobtrusiveness.
- Comfort.
- Interpretation of results.
- Validity and reliability.
- Cost.
- Method of data storage where applicable.
- Battery life where applicable.
- Potential problem areas, for example, use of watches with infection control.

This will allow us to reflect which method is the most effective going forward to measure NTS.

Stage 2: identifying relevant studies

The search strategy

A broad, comprehensive search strategy has been developed to identify relevant data. The search

terms were based on 'Population Intervention Context' framework[25] (see table 2). A full strategy is included in online supplemental appendix A. This review will focus on original research papers using digital technology to measure surgeon's NTSs, with the inclusion/exclusion criteria (see table 3) underpinning the study to include studies focusing on the subdivisions of NTSs and the methods of assessment. The subdivisions of NTS are defined as per the NOTSS protocol[10] (table 1) with the addition of CL due to the significant presence of cognitive processes guiding NTS. The search strategy with list of search terms can be found in online supplemental appendix A and was developed after extensive trial searches, tested against key texts and involved specialist librarian assistance. PubMed and Ovid MEDLINE used for initial trial searches. For the purpose of this review, the terms "cognitive" and "mental"; "load" and "workload" are used interchangeably.

Box 2 Outlining resources for searching

Databases
PubMed (9398).
Ovid MEDLINE (2472).
Embase (5911).
PsycINFO (409).
IEEE Xplore (228).
Web of Science (7804).
ACM digital Library (24).

Box 3 outlining sample data charting elements that are stored on an accessible spreadsheet

Article information
- Author.
- Year.
- Geographical origin of study.
- Study design.

Study design
- Population.
- Setting.

Intervention
- Which non-technical skill is being measured.
- Method of measurement.
- Evaluation of methods.

Study findings
- Study aims.
- Key findings relating to research.
- Relevance to surgical sabermetrics.

Inclusion and exclusion criteria can be seen in table 3. Surgeons in training and fully qualified surgeons (consultants/attendings), in any specialty, in simulation and real-life scenarios will be included. Studies from 2010 to present will be included to obtain the most up-to-date technology in use. Technology includes smart watches and other wearable consumers. Measurements include, but are not limited to, electroencephalography, skin conductance response and heart rate variability. Searches will be conducted in English and limited to English language only for reliability of interpretation. Technology includes the use of physiological sensors as a proxy for mental workload measurement.

Screening will be conducted in multiple stages as recommended by JBI.[24] Initially, analysis of title and abstracts, followed by full text, keywords and index terms, and finally the reference lists of all identified sources. The databases used for performing the searches are listed in box 2.

Stage 3: study selection

Once the search strategy is performed as stated, the records will be collated into a systematic review manager, Covidence (Veritas Health Innovation, Melbourne, Australia), aiding in the removal of duplicate records. Due to the broad search strategy required to capture key tests, initial title review will be undertaken by one reviewer. Full text and abstract screening will be conducted by at least two reviewers with any discrepancies being settled by a third expert reviewer. Reviewers will contact article authors for further information if required during screening.

A PRISMA flow diagram will be used to report the process of study selection. Information will be included to explain reasons for exclusion. Details of the full-text articles screen for inclusion will also be included as online supplemental appendix.

Stage 4: data collection

Sample data charting elements are identified in box 3. A summary table will be included in the review to accompany the narrative results. Additionally, once data have been extracted, the methods of measurement will be further investigated. The information we wish to gain about the methods is shown in table 4. Data from the scoping review will be independently collected and checked by two reviewers to ensure accuracy, with a further reviewer answering any discrepancies. Data will be stored in a spreadsheet which is maintained and accessible by all reviewers.

Stage 5: collating, summarising and reporting the results

This scoping review is providing an overview of the current research and identifying gaps in research, rather than providing quality assurance of each study. The PRISMA-ScR checklist will be used. The Quality Assessment Tool for Diverse Designs will be used for quality assessment of individual papers.[26] Each study will be reviewed in reference to Kirkpatrick's Model of Evaluation.[27] Quantitative and descriptive statistics are expected to be employed to analyse and interpret the collated data. A narrative synthesis is planned to summarise results once data collection has been completed.

Stage 6: consultation

Although considered optional by Arksey and O'Malley,[21] Levac *et al*[22] disagree and argue that consultation should be a required stage. The consultation stage involves consulting with stakeholders; in this study, this will be surgeons who we target to use this technology in further research. Surgeons will be consulted via dissemination of the results of the scoping review, online surveys

Table 4 Sample table of subgroup analysis of technology identified from scoping review with examples

Model of sensor	Measurement	NTS	Studies (references)	Cost	Battery life	Data storage	Comfort/ wearability	Potential problems
Empatica E4 wristband	SCR	Situational awareness (stressors)	Wilson *et al* [28]					
Polar H10 chest strap; INVOSTM 5100C monitor	HRV and fNIRS	Cognitive load	Kennedy-Metz *et al*[14]					

fNIRS, functional near infra-red spectroscopy; HRV, heart rate variability; NTS, non-technical skill; SCR, skin conductance response.

evaluating their attitudes and opinions towards the use of technology, and their involvement in ongoing research.

CONCLUSIONS

Surgeons need high-quality data regarding their performance. The aim of this scoping review is to identify the current technology for digital assessment of surgeons' NTSs to provide effective real-time, automatically collected, objective data that could improve surgical training, skills and outcomes; demonstrating benefits for surgeons and patients alike. This is a crucial step in furthering the innovative field of surgical sabermetrics and offers a complementary perspective to other reviews on surgical skill assessment.

Acknowledgements Acknowledgement must be given to Marshall Dozier, College Lead Librarian, College of Medicine and Veterinary Medicine, University of Edinburgh, whom without the discussions on the preparation and execution of reviews, this scoping review would not be possible.

Contributors EH constructed the main protocol in collaboration with the remaining authors (SJW/RDD/RS/SY) assisting in the planning and design of this protocol. These authors also provided a wealth of help, guidance and experience. All authors contributed to the development of the research question.

Funding This research received no specific grant from a funding agency in the public, commercial of not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Emma Howie http://orcid.org/0000-0003-1256-4295

REFERENCES

1 O'Logbon J. What can surgery learn from other high-performance disciplines? *Ann Med Surg (Lond)* 2020;55:334–7.
2 Wilson GF. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology* 2002;12:3–18.
3 Dias R, Yule S, Kennedy-Metz LR, *et al*. Psychophysiological data and computer vision to assess cognitive load and team dynamics in cardiac surgery. 2019.
4 Mehta RK, Parasuraman R. Neuroergonomics: a review of applications to physical and cognitive work. *Front Hum Neurosci* 2013;7:889.
5 Mehta RK, Agnew MJ. Exertion-dependent effects of physical and mental workload on physiological outcomes and task performance. *IIE Transactions on Occupational Ergonomics and Human Factors* 2013;1:3–15.
6 Catchpole K, Bisantz A, Hallbeck MS, *et al*. Human factors in robotic assisted surgery: lessons from studies "in the wild." *Appl Ergon* 2019;78:270–6.
7 Zhou T, Cha JS, Gonzalez G, *et al*. Multimodal physiological signals for workload prediction in robot-assisted surgery. *J Hum-Robot Interact* 2020;9:1–26.
8 Nagyné Elek R, Haidegger T. Non-Technical skill assessment and mental load evaluation in robot-assisted minimally invasive surgery. *Sensors (Basel)* 2021;21:2666.
9 Birkmeyer JD, Finks JF, O'Reilly A, *et al*. Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 2013;369:1434–42.
10 Flin RH, Youngson GG, Yule S. *Enhancing surgical performance: a primer in non-technical skills*. Boca Raton, FL, 2016.
11 Aggarwal R. Intraoperative surgical performance measurement and outcomes: choose your tools carefully. *JAMA Surg* 2017;152:995–6.
12 Cha JS, Yu D. Objective measures of surgeon non-technical skills in surgery: A scoping review. *Hum Factors* 2022;64:42–73.
13 Wilson MR, Poolton JM, Malhotra N, *et al*. Development and validation of a surgical workload measure: the surgery task load index (SURG-TLX). *World J Surg* 2011;35:1961–9.
14 Kennedy-Metz LR, Dias RD, Srey R, *et al*. Sensors for continuous monitoring of surgeon's cognitive workload in the cardiac operating room. *Sensors (Basel)* 2020;20:6616.
15 Bläsing D, Bornewasser M. Influence of increasing task complexity and use of informational assistance systems on mental workload. *Brain Sci* 2021;11:102.
16 Charles RL, Nixon J. Measuring mental workload using physiological measures: a systematic review. *Appl Ergon* 2019;74:221–32.
17 Jung JJ, Jüni P, Lebovic G, *et al*. First-Year analysis of the operating room black box study. *Ann Surg* 2020;271:122–7.
18 Yule S, Janda A, Likosky DS. Surgical sabermetrics: applying athletics data science to enhance operative performance. *Ann Surg Open* 2021;2:e054.
19 Levin M, McKechnie T, Kruse CC, *et al*. Surgical data recording in the operating room: a systematic review of modalities and metrics. *Br J Surg* 2021;108:613–21.
20 Dias RD, Ngo-Howard MC, Boskovski MT, *et al*. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *Br J Surg* 2018;105:491–501.
21 Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* 2005;8:19–32.
22 Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci* 2010;5:69.
23 Tricco AC, Lillie E, Zarin W, *et al*. PRISMA extension for scoping reviews (PRISMA-scr): checklist and explanation. *Ann Intern Med* 2018;169:467–73.
24 Aromataris E, Munn Z. Joanna briggs institute manual for evidence synthesis. 2020.
25 Stern C, Jordan Z, McArthur A. Developing the review question and inclusion criteria. *Am J Nurs* 2014;114:53–6.
26 Sirriyeh R, Lawton R, Gardner P, *et al*. Reviewing studies with diverse designs: the development and evaluation of a new tool. *J Eval Clin Pract* 2012;18:746–52.
27 Kirkpatrick JD, Kirkpatrick WK. *Kirkpatrick's four levels of training evaluation*. Association for Talent Development, 2016.
28 Wilson C, Chahine S, Cristancho S, *et al*. Unusual suspects: real-time physiological evaluation of stressors during laparoscopic donor nephrectomy. *Can Urol Assoc J* 2021;15:E205–9.