



Systematic review of team performance in minimally invasive abdominal surgery

W. J. van der Vliet¹ , S. M. Haenen¹, M. Solis-Velasco², C. H. C. Dejong¹, U. P. Neumann^{1,3}, A. J. Moser² and R. M. van Dam¹

¹Department of Hepatobiliary and Pancreatic Surgery, Maastricht University Medical Centre, Maastricht, the Netherlands, ²Pancreas and Liver Institute, Beth Israel Deaconess Medical Centre, Harvard Medical School, Boston, Massachusetts, USA, and ³Department of Surgery, Universitätsklinikum Aachen, Aachen, Germany

Correspondence to: Dr R. M. van Dam, ESCAM European Surgical Centre Aachen Maastricht, Maastricht UMC+, PO Box 5800, 6202 AZ, Maastricht, the Netherlands (e-mail: r.van.dam@mumc.nl)

Background: Adverse events in the operating theatre related to non-technical skills and teamwork are still an issue. The influence of minimally invasive techniques on team performance and subsequent impact on patient safety remains unclear. The aim of this review was to assess the methodology used to objectify and rate team performance in minimally invasive abdominal surgery.

Methods: A systematic literature search was conducted according to the PRISMA guidelines. Studies on assessment of surgical team performance or non-technical skills of the surgical team in the setting of minimally invasive abdominal surgery were included. Study aim, methodology, results and conclusion were extracted for qualitative synthesis.

Results: Sixteen studies involving 677 surgical procedures were included. All studies consisted of observational case series that used heterogeneous methodologies to assess team performance and were of low methodological quality. The most commonly used team performance objectification tools were 'construct'- and 'incident'-based tools. Evidence of validity for the assessed outcome was spread widely across objectification tools, ranging from low to high. Diverse and poorly defined outcomes were reported.

Conclusion: Team demands for minimally invasive approaches to abdominal procedures remain unclear. The current literature consists of studies with heterogeneous methodology and poorly defined outcomes.

Funding information

No funding

Paper accepted 26 November 2018

Published online 30 January 2019 in Wiley Online Library (www.bjsopen.com). DOI: 10.1002/bjs5.50133

Introduction

A substantial contribution to morbidity among surgical patients can be attributed to adverse events occurring in the operating theatre¹. Increasing evidence shows that a considerable portion of these adverse events cannot be attributed solely to deficient technical skills^{2–4}. Adverse events related to non-technical skills and team performance are common and estimated to be twice as frequent as errors in surgical technique⁵. Poor teamwork and lack of vigilance appear to be essential factors influencing procedural flow and increasing error rates⁶.

Surgical teams demand specific infrastructure, resources and competencies to perform effectively and maintain patient safety⁷. Effective team performance depends on physical and social interactions, including back-up

behaviour and leadership⁸. These demands, competencies and interactions encompassing effective team performance create a domain that is difficult to objectify and quantify^{7,9}.

In recent years, minimally invasive techniques have become the benchmark for a large number of abdominal surgical procedures¹⁰. These approaches introduce complex equipment, increased numbers of instrument changes and larger teams to the operative environment, resulting in increased demands in levels of coordination, anticipation, planning and communication¹¹. The impact of the variation in procedural approaches on team demands, error rates and patient safety remains unclear. In highly complex abdominal procedures, associated with learning curves for surgical technique, minimally invasive approaches could also have a significant impact on team performance and non-technical skills¹².

Recent studies have used a variety of methodologies to observe and objectify team performance. Consensus on the most efficient and methodically correct way of analysing and rating effective team performance is yet to be reached. The development of benchmarks for team observations and assessment of team performance will allow an accurate comparison of demands relative to surgical techniques. This will facilitate the development of effective, evidence-based training programmes for surgical teams, directed to increase team performance, decrease error rates and increase patient safety.

The aim of this systematic review was to assess the methodology used to objectify surgical team performance in minimally invasive abdominal surgery and explore team demands in relation to non-technical skills.

Methods

This study was performed according to the PRISMA guidelines¹³. Two researchers were involved in the search, inclusion, critical appraisal and data extraction of the articles selected for this study.

Eligibility criteria

Studies on assessment of surgical team performance and non-technical skills of the entire surgical team (including surgeons, anaesthesia and nursing staff) in the setting of minimally invasive abdominal surgery were included. Abdominal surgery was defined as any urological, gynaecological or general surgical procedure performed intra-abdominally. Minimally invasive techniques consisted of minimal-access approaches to the abdominal cavity, including laparoscopic, video- or robot-assisted methods.

Exclusion criteria consisted of non-original research, research performed in a simulated environment or non-human subject research, and language of publication other than English.

Study selection

Two authors performed a systematic literature search using PubMed, Embase, the Cochrane Library and Google Scholar to identify articles published before 11 October 2017. Search terms were based on subject ('teamwork', 'team learning', 'team efficiency', 'non-technical skills') and setting ('minimally invasive abdominal surgery', 'laparoscopic surgery', 'robotic surgery'). After the initial search, duplicates and non-English studies were removed. Articles were screened for eligibility by title, abstract and then full text. Reference lists and citations of the

included studies were screened for missed articles. Discrepancies in study selection between the two authors were discussed with other review team members until consensus was reached.

Critical appraisal

The methodological quality of the included studies was assessed using the Oxford Centre for Evidence-Based Medicine levels of evidence, ranging from 1 (systematic review of RCTs) to 5 (expert opinion)¹⁴. Evidence of validity of the tools used to objectify and rate operative team performance was assessed using Messick's framework¹⁵, where a test should be rated for construct validity in each specific context in which the test is employed, defining five sources of evidence (content, process response, internal structure, relations to other variables, and consequences) each rated on a three-point scale¹⁶. Evidence of validity was classified as low (0–5), moderate (6–10) or high (11–15).

Data collection

The following data were extracted from the included studies: aim, design, setting, studied procedures, observational method, observer characteristics, outcomes, conclusion, and assessment of team performance and non-technical skills.

Results

The systematic search identified 2591 manuscripts, from which 69 duplicates and 193 non-English publications were removed. Based on screening of title and abstract a further 2198 articles were excluded. The remaining 138 full-text publications were reviewed, resulting in the selection of seven studies. Subsequent review of citations and references lists led to the inclusion of a further nine articles, so that a total of 16 studies were finally included in this systematic review (Fig. 1).

Study aims, designs and settings

All included studies consisted of single-centre observational series, with the exception of one dual-centre series¹⁷ of four robot-assisted procedures. Two studies^{18,19} investigated the influence of a non-technical skills training intervention on surgical team performance. Four studies^{19–22} consisted of subanalyses of results from observed cohorts published in previous work (Table S1, supporting information).

According to the Oxford Centre for Evidence-Based Medicine, all studies provided level 4 evidence (case series, poor-quality cohort studies, case–control studies). No

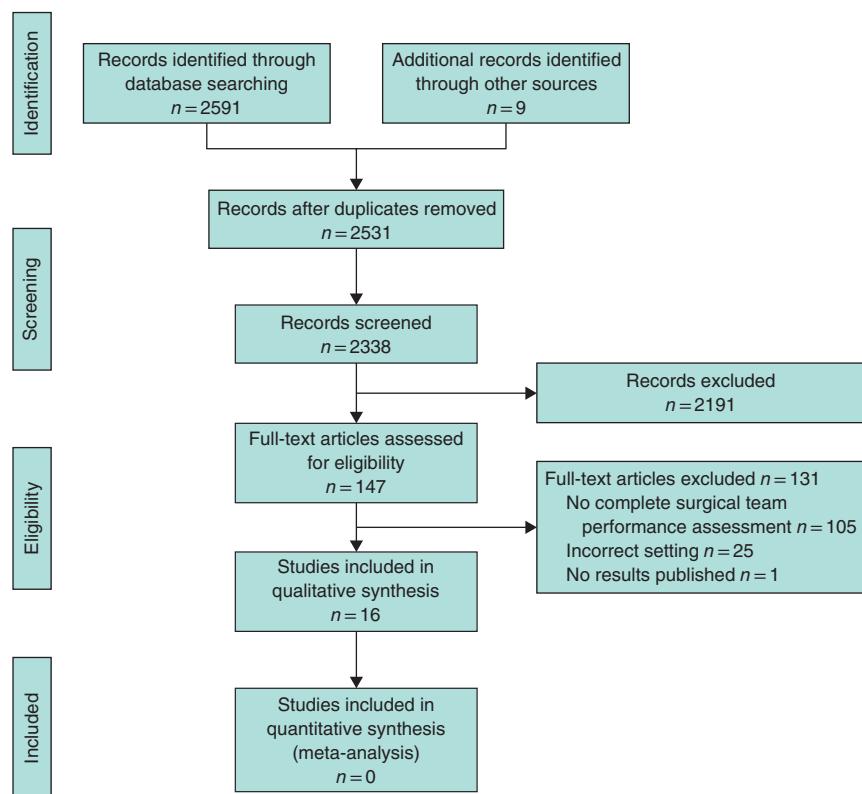


Fig. 1 Flow chart showing selection of articles for review

subgroup analysis or meta-analysis of outcomes data was attempted due to heterogeneous methodologies and outcome measurements across the included studies.

All studies observed complete surgical teams (including surgery, anaesthesia and nursing staff) with the goal of evaluating team performance through surgical workflow analysis and evaluation of disruptive events (9 studies)^{20,23–30}, the relationship between team performance and technical outcomes (3 studies)^{18,21,22}, the relationship between anticipation of surgical steps and team efficiency (1 study)³¹ or novel tools to rate team performance (3 studies)^{17,19,32}.

Of the 16 studies, 11 focused exclusively on minimally invasive procedures, whereas five included both open and minimally invasive approaches to abdominal surgical procedures. Ten studies investigated laparoscopic techniques and six a robot-assisted approach. A total of 677 procedures (281 laparoscopic, 236 robot-assisted and 160 open) were observed across the included studies, with a mean of 42.3 operations per study (*Table S1*, supporting information).

Observational methodology

The majority of studies (12) observed team performance directly, and the four remaining studies performed a

postoperative review of audiovisual recordings. Most (13) used multiple observers to evaluate team performance, with 14 reporting on methodological training of observers before the study and five including experts trained in human factor assessment or psychologists in their observing teams. Nine studies quantified interobserver reliability using a variety of methodologies; reliability was deemed good to excellent (*Table S1*, supporting information).

Most studies (14) observed team performance for the entire duration of the patient being present in the operating theatre. Six studies subdivided the procedure into preoperative, intraoperative and postoperative phases, of which four also defined a robot-docking phase.

Assessment of team performance

Seven studies used ‘construct-based’ team performance assessment tools that rated a number of behaviour constructs to create an overall score at the end of a case. Construct-based tools included: Oxford Non-Technical Skills (NOTECHS)¹⁹ (4 studies) or Observational Teamwork Assessment for Surgery (OTAS)³³ (3 studies). These tools contained moderate evidence of validity for the

Table 1 Outcome assessment tool validity according to Messick's framework of validity

Outcome assessment tool	Source of validity						Total
	Content	Response process	Internal structure	Relation to other variables	Consequence		
Construct							
NOTECHS							
Catchpole <i>et al.</i> ²¹	3	3	0	2	1	9	
Mishra <i>et al.</i> ²²	3	2	1	1	1	8	
McCulloch <i>et al.</i> ¹⁸	3	3	1	1	1	9	
Mishra <i>et al.</i> ^{19*}	3	3	1	2	1	10	
OTAS							
Mishra <i>et al.</i> ^{19*}	3	1	1	2	2	9	
Healey <i>et al.</i> ²⁵	3	2	0	1	2	8	
Undre <i>et al.</i> ²⁸	3	2	0	1	2	8	
Incident							
NOPE							
McCulloch <i>et al.</i> ¹⁸	3	0	1	1	1	6	
OR distraction assessment form							
Healey <i>et al.</i> ²⁵	2	2	0	1	1	6	
Flow disruptions							
Catchpole <i>et al.</i> ²⁴	3	3	2	2	2	12	
Catchpole <i>et al.</i> ^{20†}	3	3	2	0	2	10	
Jain <i>et al.</i> ²⁷	3	3	2	1	2	11	
Zheng <i>et al.</i> ³⁰	1	1	1	1	0	4	
Allers <i>et al.</i> ²³	1	1	1	1	0	4	
Weigl <i>et al.</i> ²⁹	3	2	0	1	0	6	
Interference assessment form							
Healey <i>et al.</i> ²⁶	2	2	1	1	1	7	
Technical							
OCHRA							
Catchpole <i>et al.</i> ²¹	3	2	0	2	1	8	
Mishra <i>et al.</i> ²²	3	2	1	1	2	9	
Mishra <i>et al.</i> ^{19*}	3	2	0	2	1	8	
OTE							
McCulloch <i>et al.</i> ¹⁸	0	1	0	1	0	2	
Workload							
NASA-TLX							
Allers <i>et al.</i> ²³	0	1	0	1	0	2	
Sexton <i>et al.</i> ³¹	3	3	3	0	1	10	
SURG-TLX							
Weigl <i>et al.</i> ²⁹	3	3	0	1	1	8	

*Subanalysis of observational data from McCulloch *et al.*¹⁸; †subanalysis of observational data from Catchpole *et al.*²⁴. NOTECHS, Oxford Non-Technical Skills; OTAS, Observational Teamwork Assessment for Surgery; NOPE, non-operative procedural error; OCHRA, Observational Clinical Human Reliability Assessment; OTE, Operative Technical Errors; NASA-TLX, National Aeronautics and Space Administration – Task Load Index; SURG-TLX, Surgery Task Load Index.

assessed outcome according to Messick's framework, with a range of 8–10 of 15 (*Table 1*). Nine studies used an 'incident-based' team performance assessment methodology, classifying non-technical procedural errors or disruptions of surgical flow in order of causation³⁴. The evidence of validity for these outcome tools was spread widely, ranging from low to high (4–12 of 15) (*Table 1*).

No study used the same categories of surgical flow disruption. The most frequent flow disruption categories defined across the nine studies were related to equipment (8 studies), external factors (8), communication (6), supervision/training (5), environment (5) and procedure (5) (*Tables 2* and *3*).

Three studies included workload assessments in their methodology consisting of the National Aeronautics and Space Administration – Task Load Index (NASA-TLX)³⁵ (2) or the Surgery Task Load Index (SURG-TLX)³⁶ (1). These tools contained low to moderate evidence of validity (2–10 of 15) for the assessed outcome. Four studies also assessed technical performance using the Observational Clinical Human Reliability Assessment (OCHRA)³⁷ (3) and Operative Technical Errors (OTE)¹⁸ (1) tools, which both have low to moderate evidence of validity (2–9 of 15).

Team performance relative to surgical approach

Three studies^{26,28,29} compared team demands and/or performance in relation to a laparoscopic or open

Table 2 Categories of flow disruption

Category	McCulloch et al. ¹⁸	Healey et al. ²⁵	Catchpole et al. ²⁴	Catchpole et al. ^{20*}	Jain et al. ²⁷	Zheng et al. ³⁰	Allers et al. ²³	Healey et al. ²⁶	Weigl et al. ²⁹	Total
Absence	X									1
Communication	X	X	X	X	X	X				6
Case-irrelevant communication		X					X	X	X	4
Coordination	X		X	X	X					4
Supervision/training	X		X	X	X		X			5
Psychomotor error	X	X				X		X		4
Resource management	X									1
Procedural	X	X					X	X	X	5
Planning problem	X									1
Surgeon decision-making				X		X				2
Surgeon's position change						X				1
External factors	X	X	X		X	X	X	X	X	8
External staff		X				X			X	3
Environment	X		X		X			X	X	5
Duty shift of nurses						X				1
Interference of video monitors		X						X	X	3
External resource	X								X	2
Equipment	X	X	X	X	X		X	X	X	8
Instrument changes			X		X	X				3
Robot switch						X				1
Patient factors	X		X		X					3
Safety consciousness	X									1
Vigilance/awareness	X									1

*Subanalysis of observational data from Catchpole *et al.*²⁴.

Table 3 Explanation of flow disruption categories

Category	Explanation	Example
Absence	Team member not present	Circulating nurse out of theatre when needed
Psychomotor error	Task execution error	Sterile instrument dropped on floor
Resource management	Misjudgement of team members' ability	Surgeon leaves assistant to finish without confirming ability to do so
Procedural	Events intrinsic to the case work	Arterial clamp time not recorded
Planning problem	Known difficulty not taken into account	Difficult intubation anticipated but not prepared for consequences
Surgeon decision-making	Technical procedural planning	Pause to determine next surgical step
External factors	Distraction from outside the operating theatre	Pager causing distraction
External staff	Disruption cause outside of surgical team	Medical student interference
Environment	Room conditions impacting flow	Incorrect room temperature
External resource problem	Organization outside the operating theatre	Essential instrument missing from standard set
Equipment	Equipment malfunction	Energy device not working
Robot switch	Robotic instrument change	Switch in controls on the robotic console
Safety consciousness	Failure to comply with safety protocols	Team member not wearing face mask
Vigilance/awareness	Failure to notice impending danger or difficulties	Failure to note significant drop in arterial pressure

approach to abdominal procedures, reporting different results. No study compared robot-assisted techniques with other approaches. Five studies investigated robot-assisted techniques and found that this approach to abdominal procedures increases team demands²⁴ that surgical teams were not always able to address effectively²⁰, resulting in increased operating times²⁷. The identification and analysis of flow disruptions can provide an evidence base for improving the efficiency and safety of robot-assisted procedures^{23,31} (*Table S1*, supporting information).

Discussion

The primary determinants of surgical outcome are generally perceived to be the patient's condition and the performance of the individual surgeon. Once corrected for patient risk factors, surgeons' technical skills are held accountable for variation in outcome. A number of different factors are important in achieving safe and effective surgical care, including infrastructure, equipment and surgical team performance³⁸. A minimally invasive surgical procedure is conducted in a

sophisticated environment combining patient factors, complex equipment and a large number of individuals set to do independent and team-based tasks.

This systematic review included 16 studies objectifying and rating surgical team performance during minimally invasive abdominal procedures. The studies were of low methodological quality, heterogeneous design, and utilized a number of different tools to objectify team performance.

In four studies, data were obtained via audiovisual recordings of the surgical environment. Despite apparent benefits of reviewing audiovisual recordings, the majority of studies collected their data through direct observation in the operating theatre. Benefits of data obtained through audio-visual recording include that data can be assessed by multiple, independent observers and incidents can be reviewed multiple times, increasing the validity and reliability of findings³⁹. In addition, during direct observation the focus of the observer may decrease due to fatigue, potentially resulting in failure to record important events⁴⁰. It is also possible that during direct observation findings are affected by the Hawthorne effect (change of behaviour in response to the awareness of being observed)^{41,42}. A prerequisite for accurate evaluation through review of audiovisual recordings is the quality of the recordings. Ethical considerations and potential hazards to team privacy and liability issues in case of adverse events may also be a limitation of its use⁴³.

Another source of variation across the reviewed studies was the number and type of observers collecting observational data. Interobserver reliability should be quantified to guarantee the quality of observations and objectivity of rating tools used⁴⁴.

This systematic review has demonstrated that, in the current literature, construct- and incident-based team performance objectification tools are used most commonly. Construct-based tools, including the OTAS and NOTECHS, rate a number of behavioural constructs on set Likert scales. These tools were developed for conventional approaches to surgery, providing global ratings for set constructs, and need to be validated for the identification of non-technical skills in minimally invasive surgery. Although used by a number of studies with similar aims, studies used different categories of flow disruption, with a broad range of validity. Some variation in validity can be related to the nature of Messick's framework.

According to Reason's organizational accident model, an adverse event is preceded by a chain of individually unimportant errors and/or latent threats that in sequence lead to an adverse event or breach of patient safety⁴⁵. Incident-based team performance objectification methodology can provide valuable insight into these patterns and

the interplay of complex minimally invasive surgical equipment.

None of the included studies was able to relate team performance to patient outcomes. This could be caused by insufficient power of individual studies. However, team performance as a determinant of morbidity and mortality is heavily biased by patient factors and technical performance. Surrogate markers for team performance could include operating time, intraoperative adverse events, and the number and duration of procedure flow disruptions. Larger, well designed studies are needed to display the true influence of minimally invasive techniques on team performance.

The major limitation of this review was the number and quality of available studies, providing insufficient data for a subgroup analysis or meta-analysis of outcomes. The majority of studies examined a heterogeneous group of operations, providing limited validity for the identification of unique non-technical skills related to specific procedures. Future studies should therefore analyse multiple approaches (open, laparoscopic, robot-assisted) in relation to a single procedure, use multiple trained observers to collect data, preferably from audiovisual recordings of the surgical environment, quantify interobserver reliability, objectify team performance using incident-based methodology with a predefined outcome set including causation and consequences of procedural flow disruptions, and analyse team performance in relation to direct (operating time, intraoperative adverse events) and indirect (patient morbidity and mortality) performance metrics. Such well designed studies are needed to gain insight into team performance demands unique to minimally invasive surgery in order to develop structured, evidence-based training programmes that enhance patient safety and procedural flow.

Disclosure

The authors declare no conflict of interest.

References

- 1 Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA *et al*. The nature of adverse events in hospitalized patients. Results of the Harvard Medical Practice Study II. *N Engl J Med* 1991; **324**: 377–384.
- 2 Anderson O, Davis R, Hanna GB, Vincent CA. Surgical adverse events: a systematic review. *Am J Surg* 2013; **206**: 253–262.
- 3 Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. *Surgery* 2003; **133**: 614–621.
- 4 Hu YY, Arriaga AF, Roth EM, Peyre SE, Corso KA, Swanson RS *et al*. Protecting patients from an unsafe system:

the etiology and recovery of intraoperative deviations in care. *Ann Surg* 2012; **256**: 203–210.

5 Healey MA, Shackford SR, Osler TM, Rogers FB, Burns E. Complications in surgical patients. *Arch Surg* 2002; **137**: 611–617.

6 Hull L, Arora S, Aggarwal R, Darzi A, Vincent C, Sevdalis N. The impact of nontechnical skills on technical performance in surgery: a systematic review. *J Am Coll Surg* 2012; **214**: 214–230.

7 Healey AN, Undre S, Sevdalis N, Koutantji M, Vincent CA. The complexity of measuring interprofessional teamwork in the operating theatre. *J Interprof Care* 2006; **20**: 485–495.

8 Greenberg CC, Regenbogen SE, Studdert DM, Lipsitz SR, Rogers SO, Zinner MJ *et al.* Patterns of communication breakdowns resulting in injury to surgical patients. *J Am Coll Surg* 2007; **204**: 533–540.

9 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.

10 Hughes-Hallett A, Mayer EK, Pratt PJ, Vale JA, Darzi AW. Quantitative analysis of technological innovation in minimally invasive surgery. *Br J Surg* 2015; **102**: e151–e157.

11 Weerakkody RA, Cheshire NJ, Riga C, Lear R, Hamady MS, Moorthy K *et al.* Surgical technology and operating-room safety failures: a systematic review of quantitative studies. *BMJ Qual Saf* 2013; **22**: 710–718.

12 Watkins AA, Kent TS, Gooding WE, Boggi U, Chalikonda S, Kendrick ML *et al.* Multicenter outcomes of robotic reconstruction during the early learning curve for minimally-invasive pancreaticoduodenectomy. *HPB (Oxford)* 2018; **20**: 155–165.

13 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009; **339**: b2700.

14 Phillips R, Ball C, Sackett D, Badenoch D, Straus S, Haynes B *et al.* *Oxford Centre for Evidence-Based Medicine – Levels of Evidence*; 2009. <http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009> [accessed 20 November 2017].

15 Messick S. Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am Psychol* 1995; **50**: 741–749.

16 Ghaderi I, Manji F, Park YS, Juul D, Ott M, Harris I *et al.* Technical skills assessment toolbox: a review using the unitary framework of validity. *Ann Surg* 2015; **261**: 251–262.

17 Cunningham S, Chellali A, Jaffré I, Classe JM, Cao CGL. Effects of experience and workplace culture in human–robot team interaction in robotic surgery: a case study. *Int J Soc Robot* 2013; **5**: 75–88.

18 McCulloch P, Mishra A, Handa A, Dale T, Hirst G, Catchpole K. The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Qual Saf Health Care* 2009; **18**: 109–115.

19 Mishra A, Catchpole K, McCulloch P. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Qual Saf Health Care* 2009; **18**: 104–108.

20 Catchpole KR, Hallett E, Curtis S, Mirchi T, Souders CP, Anger JT. Diagnosing barriers to safety and efficiency in robotic surgery. *Ergonomics* 2018; **61**: 26–39.

21 Catchpole K, Mishra A, Handa A, McCulloch P. Teamwork and error in the operating room: analysis of skills and roles. *Ann Surg* 2008; **247**: 699–706.

22 Mishra A, Catchpole K, Dale T, McCulloch P. The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surg Endosc* 2008; **22**: 68–73.

23 Allers JC, Hussein AA, Ahmad N, Cavuoto L, Wing JF, Hayes RM *et al.* Evaluation and impact of workflow interruptions during robot-assisted surgery. *Urology* 2016; **92**: 33–37.

24 Catchpole K, Perkins C, Bresee C, Solnik MJ, Sherman B, Fritch J *et al.* Safety, efficiency and learning curves in robotic surgery: a human factors analysis. *Surg Endosc* 2016; **30**: 3749–3761.

25 Healey AN, Olsen S, Davis R, Vincent CA. A method for measuring work interference in surgical teams. *Cogn Technol Work* 2008; **10**: 305–312.

26 Healey AN, Sevdalis N, Vincent CA. Measuring intra-operative interference from distraction and interruption observed in the operating theatre. *Ergonomics* 2006; **49**: 589–604.

27 Jain M, Fry BT, Hess LW, Anger JT, Gewertz BL, Catchpole K. Barriers to efficiency in robotic surgery: the resident effect. *J Surg Res* 2016; **205**: 296–304.

28 Undre S, Healey AN, Darzi A, Vincent CA. Observational assessment of surgical teamwork: a feasibility study. *World J Surg* 2006; **30**: 1774–1783.

29 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.

30 Zheng B, Martinec DV, Cassera MA, Swanström LL. A quantitative study of disruption in the operating room during laparoscopic antireflux surgery. *Surg Endosc* 2008; **22**: 2171–2177.

31 Sexton K, Johnson A, Gotsch A, Hussein AA, Cavuoto L, Guru KA. Anticipation, teamwork and cognitive load: chasing efficiency during robot-assisted surgery. *BMJ Qual Saf* 2018; **27**: 148–154.

32 Guerlain S, Adams RB, Turrentine FB, Shin T, Guo H, Collins SR *et al.* Assessing team performance in the operating room: development and use of a 'black-box' recorder and other tools for the intraoperative environment. *J Am Coll Surg* 2005; **200**: 29–37.

33 Hull L, Arora S, Kassab E, Kneebone R, Sevdalis N. Observational teamwork assessment for surgery: content

validation and tool refinement. *J Am Coll Surg* 2011; **212**: 234–243.e1–e5, 243.e5.

34 Wiegmann DA, ElBardissi AW, Dearani JA, Daly RC, Sundt TM III. Disruptions in surgical flow and their relationship to surgical errors: an exploratory investigation. *Surgery* 2007; **142**: 658–665.

35 Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In *Advances in Psychology, Volume 52: Human Mental Workload*, Hancock PA, Meshkati N (eds). North-Holland: Amsterdam, 1988; 139–183.

36 Wilson MR, Pootlon 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.

37 Tang B, Hanna GB, Joice P, Cuschieri A. Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. *Arch Surg* 2004; **139**: 1215–1220.

38 Agha RA, Fowler AJ, Sevdalis N. The role of non-technical skills in surgery. *Ann Med Surg (Lond)* 2015; **4**: 422–427.

39 Xiao Y, Schimpff S, Mackenzie C, Merrell R, Entin E, Voigt R *et al*. Video technology to advance safety in the operating room and perioperative environment. *Surg Innov* 2007; **14**: 52–61.

40 Seelandt JC, Tschan F, Keller S, Beldi G, Jenni N, Kurmann A *et al*. Assessing distractors and teamwork during surgery: developing an event-based method for direct observation. *BMJ Qual Saf* 2014; **23**: 918–929.

41 McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne effect: a randomised, controlled trial. *BMC Med Res Methodol* 2007; **7**: 30.

42 Sedgwick P, Greenwood N. Understanding the Hawthorne effect. *BMJ* 2015; **351**: h4672.

43 Millat B, Fingerhut A, Cuschieri A. Live surgery and video presentations: seeing is believing ... but no more: a plea for structured rigor and ethical considerations. *Surg Endosc* 2006; **20**: 845–847.

44 Carthey J. The role of structured observational research in health care. *Qual Saf Health Care* 2003; **12**(Suppl 2): ii13–ii16.

45 Helmreich RL. On error management: lessons from aviation. *BMJ* 2000; **320**: 781–785.

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the article.