

ORIGINAL SCIENTIFIC REPORTS

Observational Assessment of Surgical Teamwork: A Feasibility Study

Shabnam Undre, FRCS,^{1,2} Andrew N. Healey, PhD,¹ Ara Darzi, KBE, FRCS,¹
Charles A. Vincent, PhD¹

¹Clinical Safety Research Unit, Department of Surgical Oncology and Technology, Imperial College, University of London, London, UK

²Clinical Safety Research Unit, Department of Surgical Oncology and Technology, Imperial College University of London, 10th Floor, St. Mary's Hospital QEOM Building, Praed Street, W2 1NY London, UK

Abstract

Background: Teamwork is fundamental to effective surgery, yet there are currently no measures of teamwork to guide training, evaluate team interventions or assess the impact of teamwork on outcomes. We report the first steps in the development of an observational assessment of teamwork and preliminary findings.

Method: We observed 50 operations in general surgery from a single operating theater using a measure of teamwork specifically developed for use in the operating theater. The OTAS (Observational Teamwork Assessment for Surgery) comprises a procedural task checklist centered on the patient, equipment and communications tasks and ratings on team behavior constructs, namely: communication, co-operation, co-ordination, shared-leadership and monitoring.

Results: Ratings of overall team performance were reasonably high, though variable, but there was evidence that clinically significant steps were being missed which at the very least eroded safety margins. There was, for instance, a frequent failure to check both surgical and anesthetic equipment and a failure to confirm the procedure verbally, patient notes were missing in about one-eighth of the cases and delays or changes occurred in over two-thirds of the cases.

Conclusions: This study takes an initial step towards developing measures of team performance in surgery that are defined in relation to tasks and behaviors of the team. The observational method of assessment is feasible and can provide a wealth of potentially valuable research data. However, for these measures to be used for formal assessment, more research is needed to make them robust and standardized.

Research into surgical outcomes has primarily focused on the role of patient patho-physiological risk factors and on the skills of the individual surgeon. However, this approach neglects a wide range of factors that have been found to be of importance in achieving safe, high-quality performance in other high-risk environments.¹

In recent papers our group and others have argued for a much wider assessment of factors that may be relevant to surgical outcome, including such factors as equipment design, communication, team performance and factors affecting individual performance and the working environment.^{1–3} In order to carry out just such an evaluation we must have an understanding of the many factors that may influence surgical outcome, and we need to have reliable and valid measures for all the relevant ones.

Teamwork is fundamental to effective surgery, yet there are currently no measures of teamwork to help evaluate

Correspondence to: Shabnam Undre, FRCS, Clinical Safety Research Unit, Department of Surgical Oncology and Technology, Imperial College University of London, 10th Floor, St. Mary's Hospital QEOM Building, Praed Street, W2 1NY London, e-mail: s.undre@imperial.ac.uk

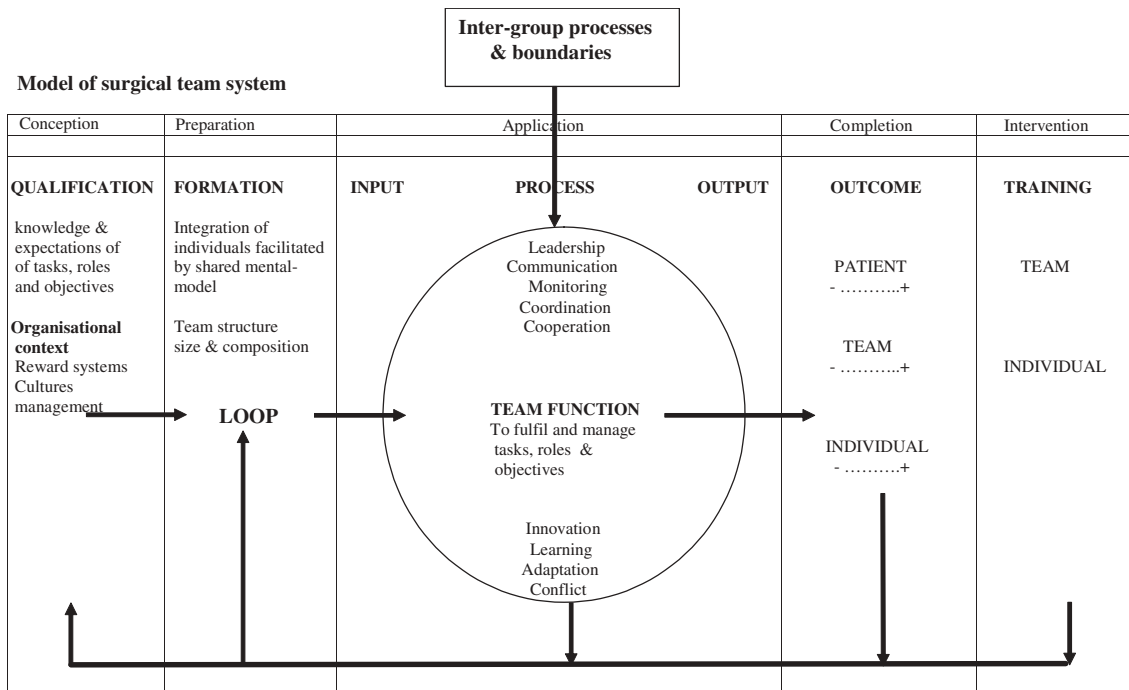


Figure 1. A model of surgical team performance. (Adapted from Healey *et al.*²³; A.N. Healey, S. Undre and C.A. Vincent, *Qual Saf Health Care*, 2004)

team interventions, guide training or assess the impact of teamwork on outcomes. Formal team training is not offered routinely in most institutions even though teams in theater are expected to function to a high standard. In practice, operating theater (OT) teams differ markedly with respect to many of the factors that may influence surgical outcome, with very little apparent standardization. To date, research that has addressed team performance in surgery has remained focused within-discipline, namely: anesthetists,⁴ nursing,⁵ surgical,⁶ or their students.⁷ The little research that has addressed interdisciplinary teamwork has tended to focus on a single behavior, most often communication, in isolation to other behaviors.⁸⁻¹¹ While it is important to describe and assess specific, individual team skills, this approach will never capture the characteristics of the whole surgical team. Furthermore, team training currently relies on informal methods of assessment and measurement, often derived from the aviation industry.¹²

The development of measures of team performance in other high-risk environments has proved to be a complex undertaking. The research in this field that has been carried out shows that to develop effective teamwork measures a framework or model of team performance needs to be constructed. Given the routine, structured nature of the surgical process, we chose to work from a basic input-process-output model of team performance

(Fig. 1). This model is established in aviation,¹³ the UK National Health Service (NHS)¹⁴ and in prominent team theory literature.¹⁵⁻¹⁸ The diagram depicted in Fig. 1 indicates that effective team function depends on input factors, such as team structure and skills, on the environment in which the team works and on the processes and guidelines underpinning teamwork.

While there are a number of methods of assessing teams, we chose to rely primarily on observation for the development of measures. Observational research has been used in many other high-risk domains effectively and more recently for assessing communication and errors in the operating theater.^{19,20} While team assessment in aviation, military and naval settings provided important guidelines, they were not directly applicable to surgery. Rather than simply adapt these measures, we sought to derive measures from guidelines of best surgical practice and combine broader dimensions of behavior with the assessment of specific tasks. Although we recognize that crises, such as severe bleeding, place particular demands on surgical teams, and hence require particular team skills, we have not addressed these in the current assessment instrument. We have, in the first instance, focused on assessing the team skills required for relatively routine surgery, while recognizing that more complex team skills may need to be incorporated at a later date.

Table 1.

The structure of OTAS is determined by critical points which mark the transition from one stage or phase of the surgical process to another

Phase ^a	Stage 1	Stage 2	Stage 3
1. PRE	Pre-operative planning and preparation	Patient sent for to anesthesia	Patient set up to operative readiness
2. OP	Opening/ access to contact of target organ	Operative-specific procedure	From prepare to close to closure complete
3 POST	Anesthetic reversal to exit	Recovery and transfer	Feedback—self-assessment

^aPRE, Pre-operative stage; OP, intra-operative stage; POST, post-operative stage.

The aim of this work was to develop a practical method of assessing teamwork in the theater that is able to capture the most important behavioral dimensions of surgical teamwork and task completion. We aimed to test the feasibility and practicality of systematic observations in the OT, evaluate a framework for measuring team performance and report preliminary data using the OTAS (observational teamwork assessment for surgery) instrument.

METHOD

Design

This was an observational study of surgical team performance using specifically developed measures of surgical team performance.

Sample

Data were collected from 50 general surgery operations [29 open and 21 laparoscopic/minimal access surgery (MAS)] in a single operating theater. The patient cohort comprised 24 female patients and 26 male patients, with ages ranging from 20 to 91 years; admissions were both elective and emergency. In keeping with the objective of this study, detailed analysis of the various patient and operation types was not carried out.

Thirty-three (66%) operations were the first operation of the day, and the remaining 17 (34%) were either the second or third operation of the day. The typical mix of operations contained hernia repairs, laparoscopic cholecystectomies, colectomies, anterior resections, ileostomy reversals, hemorrhoidectomies, appendicectomies, gastrectomy, laparoscopic fundoplication, laparoscopic banding procedure and Hartmann's procedure.

The identity of the anesthetists, nurses and surgeons varied from case-to-case and sometimes within one case. However, there was a reasonable consistency of per-

sonnel in the sample. Particular nurses and OT assistants were allocated to the OT, and there was some tendency for anesthetists to work with particular surgeons, though not as a strict rule. For this sample, we limited the duration of the operation used for the purposes of data collection from 30 to 240 minutes.

Measures

OTAS has two elements, each completed in the current format by a separate observer: a task checklist, completed by a surgical observer, and an assessment of team behavior on five dimensions, completed by a post-doctoral psychologist. The general surgical process was divided into phases and stages (Table 1). Each phase consists of distinct stages. We use the abbreviations PRE, OP and POST to refer to the pre-operative, intra-operative and post-operative phases, respectively.

Task checklist

The task list was constructed for each stage and phase of the operation with the help of theater protocols, recommendations for good practice, domain knowledge and expert advice. Additional interviews were conducted regarding the appropriateness of the task list and the contributions of the items on the task list to teamwork and outcome. The results of these will be reported in another paper (A.N. Healey, S. Undre and C.A. Vincent, submitted). Tasks were placed into three categories: namely, patient, equipment and communications tasks. Patient-centered tasks comprised either actions or information associated directly with the patient, such as safe transfer to the operating table and patient notes present. Equipment-centered checks included the checking and counting of surgical instruments. Communication-centered tasks included information such as operative site laterality confirmation. The criteria for items on the checklist were marked yes or no depending on the nature of the task. For example, under the category of equipment preparation, diathermy machine preparation was scored positive

Table 2.
ANOVA of operative type and operative duration for any stage/phase of a surgical operation

Stages of a surgical operation	Type of surgical operation ^a	Mean duration (minutes)	Standard error	Minimum duration (minutes)	Maximum duration (minutes)
PRE2A	Open	29.86	3.42	10	73
MAS		27.28	2.87	8	61
PRE3	Open	11.68	0.84	4	21
MAS		9.57	1.45	2	28
OP1	Open	8.27	1.03	1	23
MAS		9.85	1.16	4	26
OP2	Open	44.62	10.89	3	240
MAS		31.47	4.43	3	82
OP3	Open	17.31	2.9	2	80
MAS		12.23	1.77	5	30
POST1	Open	9.41	1.37	2	37
MAS		10.14	1.16	3	24

^aMAS, Minimal access surgery; open, open surgical operation

behavioral exemplars/markers during each case, which provided support for the behavioral ratings given. Inter-observer reliability is currently being explored and will be reported in future studies. Preliminary data suggest a good level of agreement, with exact agreement on ratings on a seven-point scale on over 75% of occasions for all five behavioral dimensions. A full account of the development of the measures can be found in Healey *et al.*,²³ and copies of the measures themselves can be obtained at www.csru.org.uk.

Clinical data

Clinical data were collected at the time of the operation, and a retrospective analysis of patient notes was carried out 6 months later to assess the immediate, peri-operative and late complications and follow-up for these patients.

Procedure

A surgeon of registrar level (observer 1) and a post-doctoral chartered psychologist (observer 2) collected data on tasks and behaviors, respectively. Other measures taken during observation included operative stage times, team composition in the theater and a record of any critical incidents.

Data Analysis

A mix of parametric and non-parametric tests was employed to analyze the data. We carried out ANOVAs to assess the differences of task completion and behavior ratings across the operative stages. In addition, we cal-

culated Spearman's rho rank order correlation coefficients (r_s) for rates of task completion and behavior rating across stages. Finally, we used chi-square (χ^2) tests for categorical analysis to explore the possible relations between behavior ratings, type and duration of the operation and post-operative outcomes.

RESULTS

Operation Duration

The overall mean duration of the operations was 136 minutes (range: 61–240 minutes). The breakdown of the operation duration into the various phases and stages is outlined in Table 2, and the means are depicted in Fig. 3. The mean duration (in minutes) for the stages were: PRE2A = 28.78 [open surgery (open) = 29.86, minimal access surgery (MAS) = 27.28], PRE3 = 10.8 (open = 11.68, MAS = 9.57), OP1 = 8.94 (open = 8.27, MAS = 9.85), OP2 = 39.1 (open = 44.62, MAS = 31.47), OP3 = 15.18 (open = 17.31, MAS = 12.23) and POST1 = 9.72 (open = 9.41, MAS = 10.14). A two-way ANOVA showed that there was no difference in operative duration between the types of operation for any stage of the procedure.

Task Completion

Table 3 summarizes task-completion, with the total number of tasks checked (n) and the mean, minimum and maximum number of tasks completed per operative phase. Task completion was high (92%) in the post-

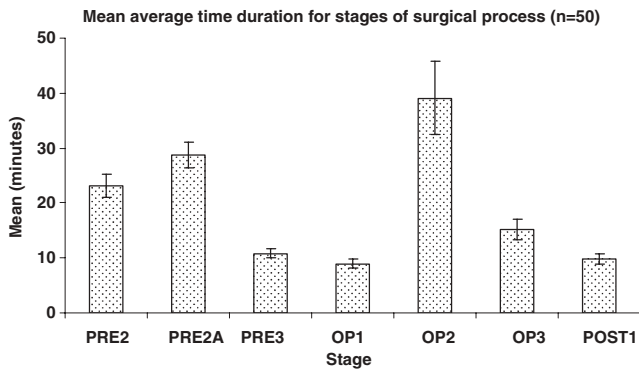


Figure 3. The average duration of the Observational Teamwork Assessment for Surgery (OTAS) operative stages. PRE2A signifies the duration of anesthesia (adapted from Healey et al. 2004²³)

operative phases, lower intra-operatively (76%) and only 69% pre-operatively. Completion of the communication tasks was the lowest [68.64 ± 1.44 (SE)], followed by equipment tasks (75.9 ± 0.656), and the completion of patient tasks was the highest (93.48 ± 0.639). Figure 4 shows that patient tasks were consistently high across phases, whereas communication remained lower in both the PRE and OP phases compared to the POST phase, while equipment task completion increased across phases. There was no significant difference between open and closed²² operations on task completion for any phase or stage.

An example of a the completed task list for PRE2 is given in Fig. 5

Team Behavior Ratings

Overall mean ratings of all team behaviors were reasonably high (>4 on a seven-point scale) and did not vary greatly across the different phases of the operation (Fig. 6), although team behaviors were rated slightly more highly in the OP phase (mean = 5.4) than in the PRE (5.2) and POST phases (5.1). Significant differences were, however, observed in the ratings on the different kinds of behavior, with communication (4.56) rated lowest, followed by leadership (5.20), shared-monitoring (5.41), co-ordination (5.48); co-operation (5.77) was rated the highest. A two-way repeated measures ANOVA, conducted on behavior⁵ and phase,³ confirmed that these behaviors differed significantly from each other overall and across phases [$F(4, 46) = 54.45$, $P < 0.000$]. Communication and co-ordination were rated higher in the OP phase than in PRE and POST phases, whereas leadership, co-operation and shared-monitoring were comparatively more consistent across phases. As with task-completion, there was no significant difference

between open or closed operations with respect to behaviors.

Relations Between Behavior and Task Completion

After aggregating tasks into mean percentage scores, we tested whether any or all of the behavior rated 's' correlated with overall task completion. In the PRE and POST phases ratings of communication rating correlated with overall task completion ($r_s = 0.468$, $P < 0.000$ and $r_s = 0.345$, $P = 0.007$, respectively), but this was not the case in the OP phase. We also tested whether there was any correlation between completion of the separate task-types and ratings of separate behaviors. In the OP phase there was no correlation between behavior ratings and tasks. However, pre-operatively there was a highly significant positive correlation between communication tasks and rating on communication behavior ($r_s = 0.415$, $P = 0.001$) and a marginally significant positive correlation between communication tasks and rating of leadership ($r_s = 0.233$, $P = 0.05$). Post-operatively, communication again was positively correlated with communication task ($r_s = 0.308$, $P = 0.01$), and co-ordination was positively correlated with equipment task completion ($r_s = 0.321$, $P = 0.01$). These results suggest overall that there is some relation between broad-based team behaviors and task completion but that, fundamentally, they are addressing different aspects of team performance.

Post-operative Complications

Of the 19 cases with complications, only four were in the MAS operation category, whereas 15 were in the open operation category. Principal complications included pain, pyrexia, wound infection, urinary retention, splenic-injury, bladder injury and two post-operative deaths (one cardiac arrest and one post-operative sepsis with multi-organ failure leading to death). We found no significant relationships between rates of task-completion and the occurrence of complications. We did find a relationship between team behaviors and complication occurrence. However, while the results show that there may be an association between teamwork and post-operative complications, this analysis cannot be interpreted at face value at this time. A full risk-stratified analysis and other factors, such as technical skills, will have to be taken into account before any such association can be made; this area will be explored further in future studies.

Table 3.
Summary of task completion per phase and task type

Task type	PRE			OP			POST		
	Mean (SE)	Min% (max%)	Range (n tasks)	Mean (SE)	Min% (max%)	Range (n tasks)	Mean (SE)	Min% (max%)	Range (n tasks)
Equipment	56.4 (61.38)	42.11 (80.95)	38.84 (22)	81.94 (1.02)	52.17 (100)	47.83 (26)	89.29 (0.66)	75 (100)	25 (10)
Communication	60.72 (1.99)	12.5 (86.67)	74.17 (18)	54.84 (2.65)	22.2 (100)	77.78 (10)	90.34 (1.44)	61.54 (100)	38.46 (14)
Patient	89.60 (0.95)	68.75 (100)	31.25 (18)	93.49 (1.56)	64.29 (100)	35.71 (14)	97.36 (0.68)	73.33 (100)	26.67 (16)
Overall mean	68.93			76.76			92.33		
Overall N task			58			50			40

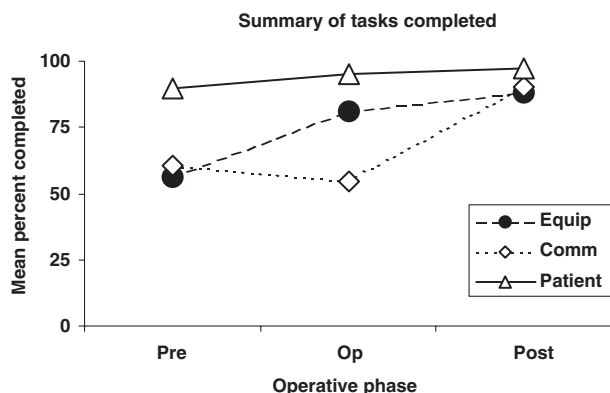


Figure 4. Average task completion per phase of the operation.

DISCUSSION

We have described the first full trial of an observational method of team assessment in surgery. The assessment covers both tasks and behaviors in the PRE, OP and POST phases of surgery. Team performance was measured against a structured protocol of tasks and teamwork behaviors which were assessed on a set of ordinal scales. Ratings of behaviors were global and based on the observation of effective and ineffective behaviors in the OT. This framework of measurements therefore covers both specific tasks carried out by the team and also the team's overall performance. This distinction is important because different teams may complete a similar number of routine tasks, but vary in the quality of their communication and co-ordination.

Our findings suggest that observational assessment in operating theaters is feasible, purposeful and informative. The ratings of overall team performance were reasonably high, though variable, but the completion of operative tasks was some way below best-practice guidelines and certainly below the standard of performance expected of

high-reliability teams. Clinically, the lapses in task completion did not appear to affect outcome, but there was evidence that clinically significant steps were being missed, which at the very least eroded safety margins. There was, for instance, a frequent failure to check both surgical and anesthetic equipment and a failure to confirm the procedure verbally, patient notes were absent in about one-eighth of the cases and in over a third of the cases there was no verbal confirmation of readiness to start the operation. There were also incomplete notes, lack of equipment, lack of blood results and patients not being starved on wards.

Considerable variation in stage duration was also found, which reflected, in part, variations in procedures and patient-specific problems. However, the variation pre-operatively was also attributed in many cases to delays. Delays and changes to the case-lists occurred in over 70% of the cases. The reasons for delays were numerous, including the patient journey to theater, busy ward staff or porters and bed allocation problems. The time that elapsed in the anesthetic room once the patient had arrived varied for various reasons: the patient's condition and the absence of the surgeon or anesthetist. Other delaying factors included the staff being unfamiliar with stock locations, coupled with a lack of compensatory supervision. The potential risk of an error and/or accident is the highest when these delays and associated deviations from best practice mount up and are compounded by the additional pressures of workload.²⁴

Communication was rated lower than other behaviors, particularly in the PRE and POST phases. This was due in part to the fact that inter-disciplinary communication is less formalized and more distributed before and after the actual operation. There was a positive correlation between communication rating and overall task completion pre-operatively and post-operatively, but not intra-opera-

Pre2	Opt	Act	Effect on team function or source of deviation
Pt sent for (PS) - to Anaesthesia (AS)	?		
Sent for time 8.00	?		
Arrival time 8.30	?		
Anaesthesia start 8.55	?		
Observer 1			
Team Tasks			
Patient			
patient sent for	1	1	
correct patient verified	1	0	wrong patient arrived
surgical site & laterality verified	1	Na	
surgical procedure verified	1	0	
notes & x-rays present for patient	1	1	
patient details entered to pc	1	0	
booked operation time	1	1	1.36
patient condition monitored by Anaesthetist	1	1	
equipment & provisions			
Anae. Equip. checked and working	1	1	
Surg. Instruments checked and working	1	1	
Surg. Instruments covered till op	1	0	
op-specific equip. checked working	1	0	
gowns & gloves prepared	1	0	
Anae. Drugs prepared	1	1	
Communication			
surgeon briefs team on procedure	1	0	
Anae-ODA discuss pt requirements	1	0	
Sn & Cn confirm instruments check	1	1	
correct patient confirmed verbally	1	0	
procedure confirmed verbally	1	0	
surgical site laterality verbally confirmed	1	na	
Condition of patient			
Anaesthetised	1	1	
temperature within range	1	na	
urine output within range	1	na	
cardiac output in range	1	1	
Anae-ODA (pt spec requirements)	1	1	
Critical incidents			
critical incident	0	1	
critical incidents reported	1	0	
hazards to pt	0	1	potentially

Figure 5. Example of task list from PRE2.

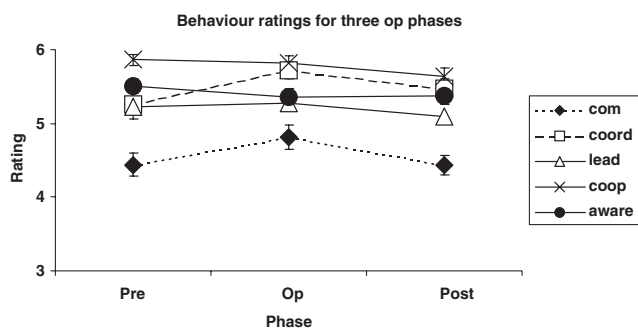


Figure 6. Average behavior ratings per phase of the operation.

tively. There was also a positive correlation between communication tasks and the rating on communication behavior and leadership. These results suggest overall that there is some relation between broad-based team behaviors and task completion. However, importantly, the results also address different aspects of team performance in providing both information about protocol and deviations from established protocol and evidence-based information on *ad hoc* variations in team behavior.

A crucial issue to be explored in future studies is whether team performance can be shown to affect outcome. In our study, we found no significant relationships between task-completion and complication. However, behavior ratings were associated with the occurrence of complications. For instance, we found that verbal communication confirming antibiotics was observed in only 53% of cases, which may have influenced infection outcome. On further investigation only four patients had a post-operative infection, and of these only two had a lack of verbal confirmation of antibiotics being given. However, lack of confirmation verbally does not necessarily mean that the antibiotics were not written up already or that the patient did not already receive them with the induction of anaesthesia. Outcome data have to be interpreted with caution as appropriate risk stratification must be applied for a proper analysis and various other factors, such as technical skills, have to be taken into consideration. This study was a feasibility study to assess teamwork at this stage and not to relate teamwork to outcome, although this may be possible in the future once other factors have been accounted for. The main point of this study is simply to illustrate that different forms of team performance data may be used in the general modeling of the system and its relationship to outcome. It will be the aim of future studies to test which form of measurement and data are the most effective for analyzing team performance in general.

Another sensitive subject to be taken into account while undertaking research of this nature is the fear of blame and disciplinary action. As pointed out by Vincent *et al.*¹ in their paper, fears may be expressed by members of surgical teams that observation may be used for 'surveillance', checking up and, possibly, as a basis for disciplinary action. We have stressed to the team that our data will only be used as a research tool. Most importantly, it is necessary to emphasize that the purpose of such observations is not to study individuals, but processes, procedures and team performance in general. The aim is to observe common patterns over a series of operations to help improve teamwork and efficiency, not to examine individual deficiencies.

We are currently in the process of refining the observational measures and testing in a further sample of operations in a different theater setting. A particularly important development is to provide ratings of the three sub-teams (nursing, anaesthetic and surgical) as we believe this will give a more accurate reflection of theater performance than the overall team ratings used in this study. We also intend to develop a short version suitable for use in training and simulation and, crucially, for direct comparison of the team performance during training with that actually observed in the OT. Further assessment of reliability and validity is also required and is being addressed in on-going studies. We need to pay particular attention to delineating the process of observation of behaviors and specifying how these should be rated so that the measures may be more widely used. A clearer specification of scoring may provide the necessary detail to show why certain teams perform at certain levels on each behavioral dimension and how and why those performance elements affect outcome. A closer analysis of tasks, particularly communication, may provide some indication of those relations. Moreover, a consideration of team composition and its relationship to team co-ordination is also important. Indeed, we see the potential to construct specific behavior ratings that are more closely related to the activity of particular sections of the team at particular stages.

Downloads

The measures used in this study will be freely available from The Clinical Safety Research website: www.csru.org.uk

ACKNOWLEDGEMENTS

We thank the BUPA foundation and the Department of Health: Patient Safety Research Programme for funding this work. We are grateful to Dr. Nick Sevdalis for his contribution to the revision of this manuscript. We would also like to thank our Surgical, Anaesthetic and Nursing Colleagues for their support and co-operation in this study.

REFERENCES

1. Vincent C, Moorthy K, Sarker SK, Chang A, Darzi AW. Systems approaches to surgical quality and safety: from concept to measurement. *Ann Surg* 2004;239:475–482.

2. Calland JF, Guerlain S, Adams RB, Tribble CG, Foley E, Chekan EG. A systems approach to surgical safety. *Surg Endosc* 2002;16:1005–1014.
3. Undre S, Sevdalis N, Healey AN, Darzi A, Vincent C. Teamwork in the operating theatre: cohesion or confusion? *J Eval Clin Pract* 2006 12:182–189. .
4. Fletcher GC, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. *Br J Anaesth* 2002;88:418–429.
5. Millward LJ, Jeffries N. The team survey: a tool for health care team development. [Miscellaneous Article]. *J Adv Nurs* 2001;35:276–287.
6. Baldwin PJ, Paisley AM, Brown SP. Consultant surgeons' opinion of the skills required of basic surgical trainees. *Br J Surg* 1999;86:1078–1082.
7. Lang NP, Rowland-Morin PA, Coe NP. Identification of communication apprehension in medical students starting a surgery rotation. *Am J Surg* 1998;176:41–45.
8. Hawryluck LA, Espin SL, Garwood KC, Evans CA, Lingard LA. Pulling together and pushing apart: tides of tension in the ICU team. *Acad Med* 2002;77:(Suppl 10) S73–S76.
9. Lingard L, Reznick R, DeVito I, Espin S. Forming professional identities on the health care team: discursive constructions of the 'other' in the operating room. *Med Educ* 2002;36:728–734.
10. Thomas EJ, Sexton JB, Helmreich RL. Discrepant attitudes about teamwork among critical care nurses and physicians. *Crit Care Med* 2003;31:956–959.
11. Grommes P. Contributing to coherence: an empirical study of OR team communication. In: Minnick-Fox M, Williams A, Kaser E, editors. *Proceedings of the 24th Penn Linguistics Colloquium*. Univ Penn Working Papers Linguistics 2000;7:1, 87–98.
12. Morey JC, Simon R, Jay GD, Wears RL, Salisbury M, Dukes KA, *et al*. Error reduction and performance improvement in the emergency department through formal teamwork training: evaluation results of the MedTeams project. *Health Serv Res* 2002;37:1553–1581.
13. Helmreich RL, Foushee HC. (1993) Why crew resource management? Empirical and theoretical basis of human factors training in aviation. In: Weiner EL, Kanki BG, Helmreich RL editors. *Cockpit Resource Management Academic*, New York, pp 3–45.
14. West M, Borrill C, Unsworth K. Team effectiveness in organisations. In: Cooper CL, Robertson IT, editors. *International Review of Industrial Organisational Psychology*, vol 13. Chichester, Wiley, 1998:1–48.
15. Cohen S, Bailey D. What makes teams work: group effectiveness research from the shop floor to the executive suite. *J Manage* 1997;23:239–290.
16. Gladstein D. Groups in context: a model of group task effectiveness. *Adm Sci Q* 1984;29:499–517.
17. Guzzo RA, Shea GP. (1992) Group performance and intergroup relations in organisations. In: Dunnette MD, Hough LM editors. *Handbook of Industrial and Organisational*

- Psychology. Palo Alto, Consultant Psychological Press, CA, pp 269–313.
18. Stewart GL, Barrick MR. Team structure and performance: assessing the mediating role of intrateam process and the moderating role of task type. *Acad Manage J* 2000; 43:135–148..
 19. Carthey J, de Leval MR, Reason JT. The human factor in cardiac surgery: errors and near misses in a high technology medical domain. *Ann Thorac Surg* 2001;72:300–305.
 20. Lingard L, Espin S, Whyte S, Regehr G, Baker GR, Reznick R, *et al.* Communication failures in the operating room: an observational classification of recurrent types and effects. *Qual Saf Health Care* 2004;13:330–334.
 21. Dickinson T L, McIntyre RM.. A conceptual framework for teamwork measurement. In: Brannick MT, Salas E, Prince C editors, *Team performance assessment and measurement, theory, methods, and applications* . Mahwah: Lawrence Erlbaum Associates 1997:19-43..
 22. Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' non-technical skills (ANTS): evaluation of a behavioural marker system. *Br J Anaesth* 2003;90:580–588.
 23. Healey AN, Undre S, Vincent CA. Developing observational measures of performance in surgical teams. *Qual Saf Health Care* 2004;13:(Suppl 1) i33–i40.
 24. de Leval MR, Carthey J, Wright DJ, Farewell VT, Reason JT. Human factors and cardiac surgery: a multicenter study. *J Thorac Cardiovasc Surg* 2000;119:661–672.